NUMERICALLY INTENSIVE ECONOMIC POLICY ANALYSIS V

Queen's University at Kingston

May 12-13, 2006

INDEX

[Speakers and Participants] [Program]

SPEAKERS

Victor Aguirregabiria	Boston University	*A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments*
Samuel Danthine	University of Quebec at Montreal	*Bargaining Frictions and Hours Worked*
Suqin Ge	University of Minnesota	*Women's College Choice: How Much Does Marriage Matter?*
Samuel Kortum	University of Minnesota	*An Anatomy of International Trade: Evidence from French Firms.*
Carrie Lee	Queen's University	*Wealth Shocks, Birth Cycles and Wage Trends: A Dynamic Equilibrium Model of Fertility, Wage and Labor Supply in the 20th Century*
Rasmus Lentz	University of Wisconsin	*An Empirical Model of Growth through Product Innovation*
Elena Pastorino	University of Iowa	*Career Dynamics Under Uncertainty: Estimating the Value of Firm Experimentation*
Carlos J. Serrano	University of Toronto	*The Market for Intellectual Property:*
Matthew Shum	Johns Hopkins University	*Pairwise-Difference Estimation of a Dynamic Optimization Model*

E-POSTER SESSION

Ahu Gemici	University of Pennsylvania	*Understanding Family Migration and Employment Decisions*
Jean-Francois Houde	Queen's University	*Identification and 2-step Estimation of DDC Models with Unobserved Heterogeneity*
Kazuko Kano	University of British Columbia	*Menu Costs, Strategic Interactions, and Retail Price Movements*
Natalia Mishagina	Queen's University	*Career Dynamics of Science and Engineering Doctorates*
Travis Ng	University of Toronto	*Outsourcing by risk-pooling*
Joel Rodrigue	Queen's University	*Imported Productivity Gains and Domestic Firms*
Mari Sakudo	University of Pennsylvania	*Co-residence and intergenerational transfers*
Katsumi Shimotsu	Queen's University	*Nested Pseudo-likelihood Estimation and Bootstrap-based Inference for Structural Discrete Markov Decision Models*

Jun Zhang Queen's University *Nonparametric Identification and Estimation of Contests*

PARTICIPANTS

PROGRAM

The presentations are scheduled on the QED Workshop Calendar.

Thursday Night (May 11) Informal Welcome: after 8pm, Grizzly Grill

Friday: Sessions Workshop Dinner (Pan Chancho 6:30pm)

Saturday: Sessions; Traditional Therapy Session (takeout dinner at Chris's house)

Presentations are in Dunning Hall Room 27. The e-poster session is in the Dunning Hall Conference Room.

Presenters have 60 minutes with no discussant

An expense claim form is available here

MAP OF NIEPA LOCATIONS (updated)

A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments

Victor Aguirregabiria
Boston University

Chun-Yu Ho
Boston University

This version: April, 2006
PRELIMINARY AND INCOMPLETE

Abstract

This paper proposes and estimates a dynamic oligopoly model of the US airline industry. We use the model to study the sources of market power in this industry and to evaluate counterfactual policy experiments such as changes in the fees that airports charge to airlines. In our model, airlines decide in which local markets (i.e., airport-pairs) to operate as well as their fares in each of these markets. An airline's network has two types of effects on its decision to operate in a local market. First, the scale of operation of an airline in the origin and destination airports of a given market reflects services which are valued by consumers (e.g., more convenient flight schedule and landing facilities). And second, the entry cost and costs of operation (both marginal and fixed) may also depend on the airline's network through economies of density or scale. The Markov perfect equilibria in the different local markets are interconnected through these network effects such that a temporary shock in a local market can have long-lasting effects on all the markets in the industry. One of the objectives of this paper is to quantify these different network effects. We use a panel data set from the Airline Origin and Destination Survey with information on quantities, prices, and entry and exit decisions for every airline company over more than two thousand city-pair markets and several years. We first estimate variable profits using information on quantities and prices. The structure of the model provides instruments for the estimation of demand parameters. In a second step, we estimate fixed operating costs and sunk costs in the dynamic game of entry and exit. Our estimates show very large heterogeneity across markets in entry costs and fixed operating costs. Ignoring this heterogeneity induces serious biases in the estimation of these costs.

Keywords: Airline industry; Networks; Sunk costs; Industry dynamics; Estimation of dynamic games.

JEL codes: L11, L13.

1 Introduction

The market structure of the US airline industry has undergone important changes since the deregulation in 1978 which removed entry and exit restrictions and allowed carriers to set airfares.[1] Entry is argued to be one of the driving forces for achieving competitive outcome which result in improving the welfare of consumers. Sunk cost is one of the determinants to govern entry and exit process and therefore the knowledge of sunk cost is crucial for evaluating the market structure of airline industry. In this paper, we construct and estimate a dynamic oligopoly model of the US airline industry *which allow us to recover the information of sunk cost and fixed operating cost*. We use the model to study the sources of market power in this industry and to evaluate counterfactual policy experiments such as changes in the fees that airports charge to airlines.

In our model, airlines decide in which local markets (i.e., airport-pairs) to operate as well as their fares in each of these markets. An airline's network has two types of effects on its decision to operate in a local market. First, the scale of operation of an airline in the origin and destination airports of a given market reflects services which are valued by consumers. For instance, a larger hub operation implies more convenient flight schedules and landing facilities. And second, the entry cost and costs of operation (both marginal and fixed) may also depend on the airline's network through economies of density or scale. The Markov perfect equilibria in the different local markets are interconnected through these network effects such that a temporary shock in a local market can have long-lasting effects on all the markets in the industry. One of the objectives of this paper is to quantify these different network effects.

We use a panel data set from the Airline Origin and Destination Survey with information on quantities, prices, and entry and exit decisions for every airline company over more than two thousand city-pair markets and several years. We first estimate variable profits using information on quantities and prices. The structure of the model provides instruments for the estimation of demand parameters. In a second step, we estimate fixed operating costs and sunk costs in the dynamic game of entry and exit. We use the method proposed by Aguirregabiria and Mira (2006) for estimating the dynamic game to recover the sunk cost and fixed operating cost for the US airline industry.[2]

Both reduced form and structural econometric models of entry have been used to study the market structure of the US airline industry. For reduced form approaches, Sinclair (1995)

[1] Borenstein (1992) provides an excellent overview on the airline industry. Early policy discussion includes Bailey et al (1985), Morrison and Winston (1986) and Levine (1987). Recent discussion of evaluating the deregulation can be found in Tranportation Research Board (1999), Kahn (2001) and Morrison and Winston (2001).

[2] Bajari, Benkard, and Levin (2005), Pakes, Ostrovsky, and Berry (2005), and Pesendorfer and Schmidt-Dengler (2004) also suggest alternative methods to estimate dynamic games.

examines the effect of sunk cost and network on entry and exit decision. Recently, Bogaslaski et al. (2004) look into the entry strategy of Southwest airline in 1990s and its impact on other traditional airlines. Januszewski and Lederman (2003) investigate the factors that affect the entry pattern of a set of low cost carriers. More structural approaches have been considered by Reiss and Spiller (1989), Berry (1992) and Ciliberto and Tamer (2006). Reiss and Spiller (1989) model the demand, cost and entry decisions simultaneously for the small markets. Berry (1992) extends the model of Bresnahan and Reiss (1990, 1991) to allow unobserved heterogeneity among players to estimate the decision of entry in the 50 largest US cities in year 1980. Ciliberto and Tamer (2006) extends Tamer (2003) to allow for multiple equilibria in the entry game explicitly.[3] These structural studies consider a static framework. In this paper, we extend the structural analysis of the US airline industry by considering a dynamic game of entry and exit.

In this regard, the dynamic oligopoly model proposed by Pakes and McGuire (1994) and Ericson and Pakes (1995) provides a useful framework for addressing the dynamics of the US airline industry. Our model is in the spirit of these previous models, but it incorporates private information shocks. Therefore, we consider a dynamic game of incomplete information.

In a related study, Bresnahan and Reiss (1993) show that the difference between entry and exit thresholds provide information on sunk cost which is important to determine the market structure and industry dynamics. They argue that firms are forward-looking. The current entry decision also takes the market structure in the future and exit policy into account, hence econometrician will not able to correctly infer the sunk cost from entry and exit behavior if one only use information from current period. They illustrate the intuition by a highly stylized two period model of monopolist without incorporating the industry dynamics, but they do not go further to show the idea will still hold under more realistic model. Our analysis extends Bresnahan and Reiss (1993) approach.

The rest of this incomplete version of the paper is organized as follows. Sections 2 presents the model and our basic assumptions. The data set and our working sample are described in section 3. Section 4 discusses the estimation procedure and presents our preliminary results. Section 5 concludes.

2 Model

The industry is configured by N airline companies and M local markets. A local market is a particular origin-destination-origin city-pair. For instance, the route Chicago(Midway)-Detroit-Chicago(Midway) is a local market and the route Detroit-Chicago(Midway)-Detroit is

[3]Hendricks et al. (2003) show that multiple equlibria is plausible in the entry game with hub-and-spoke network if the small carriers have lower cost than the traditional airlines.

a different market. We index markets by m and airlines by i. Time is discrete and indexed by t. Our model provides a separate (Markov perfect) equilibrium for each individual local market. However, the M local market equilibria are interconnected and this interconnection provides a joint dynamics for the whole US airline industry. There are two factors behind the interrelationship between local markets. First, consumers value the network of an airline. And second, entry costs and fixed operating costs also depend on an airline's network (i.e., economies of density). Our model distinguishes between these two sources of network effects.

At period t there are $\bar{n}_{mt}$ airlines operating in market m. We call these firms *incumbents.* The rest of firms, i.e., $N - \bar{n}_{mt}$, are potential entrants. Each incumbent firm has two characteristics that affect demand and costs: the type of airline, i.e., network carrier, low cost carrier, or other; and the "quality" of the airline with its group. We denote an airline type by $g_i \in \{1, 2, 3\}$. This type is invariant over markets and over time. The quality of airline i in market m at period t is α_{imt}. This quality depends, among other things, on the airline's network. The set of possible qualities is Ω and it is finite and discrete. Therefore, we can represent the distribution of qualities in the market using the vector:

$$n_{mt} = \{\ n^g_{mt}(\alpha) : g \in \{1, 2, 3\},\ \alpha \in \Omega\ \} \tag{1}$$

where $n^g_{mt}(\alpha)$ is the number of incumbent firms of type g with quality α in market m at period t.

Every period t the set of incumbent firms know n_{mt}, the demand system and variable costs and they compete in prices to determine variable profits. The evolution of n_{mt} is endogenous in this model and follows a first order Markov process.

2.1 Consumer demand

Our specification and estimation of the demand system follows Berry (1994) and Berry, Levinshon and Pakes (1995) and it is based on the aggregation of heterogeneous consumers' discrete choices. This type of demand system has been applied to the US airline industry by Gayle (2004), Lederman (2004) and Berry, Carndall and Spiller (2006).

Let H_m be the number of consumers in market (city-pair) m. Every period t, consumers decide whether to purchase or not a ticket for this city-pair flight and which airline to patronize. The indirect utility function of a consumer who purchases an air ticket from airline i in this market is:

$$u_{imt} \;=\; \tilde{\alpha}^g_{imt} - p_{imt} + v_{imt} \tag{2}$$

where p_{imt} is the air ticket price of airline i and $\tilde{\alpha}^g_{imt}$ is the *quality* of airline i that is the same for all consumers. The variable v_{imt} is consumer specific and it captures consumer heterogeneity in preferences for difference airlines. The objective quality $\tilde{\alpha}^g_{imt}$ depends on different exogenous characteristics of the airline and of the airport-pair. More importantly for this paper, this quality also depends on the hub operation of an airline in origin and destination airports. In order to emphasize this dependence, we call $\tilde{\alpha}^g_{imt}$ the *hub index* of airline i at market m. In section 4 (i.e., equation (22)), we provide an specification of $\tilde{\alpha}^g_{imt}$ in terms of measures of the airline's operation in the origin and destination airports. For the moment, we treat $\tilde{\alpha}^g_{imt}$ as a state variable. The choice of not purchasing any air ticket for this city-pair flight is called the *outside alternative*. The index of the outside alternative is $i = 0$. The value for the outside alternative may vary across markets and over time and $u_{0mt} = \tilde{\alpha}_{0mt} + v_{0mt}$.

A consumer chooses alternative i if and only if u_{imt} is greater than the utility associated with any other choice alternative. These conditions describe the unit demands of individual consumers. To obtain aggregate demands we have to integrate individual demands over the idiosyncratic v variables. The form of the aggregate demands depends on our assumption on the probability distribution of consumer heterogeneity. We use the nested logit framework. More specifically, we follow Cardell (1997) and assume that:

$$v_{imt} = (1 - \sigma_v)\ v^{(1)}_{imt} + \sigma_v\ v^{(2)}_{gmt} \tag{3}$$

where $v^{(1)}_{imt}$ and $v^{(2)}_{gmt}$ are independent random variables and σ_v is a parameter that can vary across markets. The variable $v^{(1)}_{imt}$ is Type I extreme value distributed and it varies within groups. The variable $v^{(2)}_{gmt}$ varies only between groups and it has a distribution such that v_{imt} is also a Type I extreme value random variable. The parameter σ_v measures the relative importance of between-groups horizontal product differentiation. If σ_v is close to zero, then our grouping of airlines is not significant from the point of view of the demand. Given this distribution, the demand for airline i (that belongs to group g) in market m at period t is:

$$q_{imt} = \frac{H_m\ a_{imt}\ \exp\left\{\dfrac{\tilde{\alpha}_{imt} - \tilde{\alpha}_{0mt} - p_{imt}}{\sigma_v}\right\}}{\sum_{j \in B_g} a_{jmt}\ \exp\left\{\dfrac{\tilde{\alpha}_{jmt} - \tilde{\alpha}_{0mt} - p_{jmt}}{\sigma_v}\right\}} \frac{\left(\sum_{j \in B_g} a_{jmt}\ \exp\left\{\dfrac{\tilde{\alpha}_{jmt} - \tilde{\alpha}_{0mt} - p_{jmt}}{\sigma_v}\right\}\right)^{\sigma_v}}{1 + \sum_{g'=1}^{3}\left(\sum_{j \in B_{g'}} a_{jmt}\ \exp\left\{\dfrac{\tilde{\alpha}_{jmt} - \tilde{\alpha}_{0mt} - p_{jmt}}{\sigma_v}\right\}\right)^{\sigma_v}} \tag{4}$$

$a_{imt} \in \{0, 1\}$ is the indicator of the event "airline i is active in market m at period t".

For the characterization of the Bertrand equilibrium that we present below, it will be convenient to represent the previous demand system in terms of price-cost margins and *hub index*. Let c_{imt} be the unit variable cost of airline i at period t. This cost may depend on hub index $\tilde{\alpha}_{imt}$. Define the price-cost margin $\tau_{imt} \equiv (p_{imt} - c_{imt})/\sigma_v$ and firm i's *cost-adjusted hub index*:

$$\alpha_{imt} \equiv \frac{\tilde{\alpha}_{imt} - \tilde{\alpha}_{0mt} - c_{imt}}{\sigma_v} \tag{5}$$

Using these definitions, we can re-write the demand system as:

$$q_{imt} = \frac{H_m \; a_{imt} \;\; \exp\{\alpha_{imt} - \tau_{imt}\}}{\sum_{j \in B_g} a_{jmt} \;\; \exp\{\alpha_{jmt} - \tau_{jmt}\}} \; \frac{\left(\sum_{j \in B_g} a_{jmt} \;\; \exp\{\alpha_{jmt} - \tau_{jmt}\}\right)^{\sigma_v}}{1 + \sum_{g'=1}^{3} \left(\sum_{j \in B_{g'}} a_{jmt} \;\; \exp\{\alpha_{jmt} - \tau_{jmt}\}\right)^{\sigma_v}} \tag{6}$$

The variable profit of airline i is $(p_{imt} - c_{imt}) * q_{imt} = (\sigma_v \; \tau_{imt}) * (\; q_{imt} \; H_m)$.

2.2 Price competition and Bertrand equilibrium

Consumers demand is static and there are not price adjustment costs. Therefore, price competition in this model is static. Given the hub index $\{\alpha_{imt}\}$, which are common knowledge, each incumbent firm chooses a price-cost margin τ_{imt} to maximize his current variable profit.

The Nash-Bertrand equilibrium is characterized by the system of best response equations: $\tau_{imt} = -q_{imt} \left(\partial q_{imt} / \partial \tau_{imt}\right)^{-1}$. Equilibrium margins depend only on the own cost-adjusted hub index of the firm and on the cost-adjusted hub index of competitors. Then, the equilibrium price-cost margin for airline depends only on the firm's own cost-adjusted hub index and on the vector n_t. Let $\tau(\alpha_{imt}, n_{mt})$ be the equilibrium margin of airline i. Therefore, the equilibrium variable profit is $R(\alpha_{imt}, n_{mt})$ where:

$$R(\alpha_{imt}, n_{mt}) \;\; = \;\; \frac{\sigma_v \; H_m \; \tau(\alpha_{imt}, n_{mt}) \;\; \exp\{\alpha_{it} - \tau(\alpha_{imt}, n_{mt})\}}{\sum_{\alpha \in \Omega} n_{mt}^{g}(\alpha) \;\; \exp\{\alpha - \tau(\alpha, n_{mt})\}} \frac{\left(\sum_{\alpha \in \Omega} n_{mt}^{g}(\alpha) \;\; \exp\{\alpha - \tau(\alpha, n_{mt})\}\right)^{\sigma_v}}{1 + \sum_{g'=1}^{3} \left(\sum_{\alpha \in \Omega} n_{mt}^{g'}(\alpha) \;\; \exp\{\alpha - \tau(\alpha, n_{mt})\}\right)^{\sigma_v}} \tag{7}$$

Given that the set of possible cost-adjusted hub index is discrete, we can compute (in finite time) the Nash-Bertrand equilibrium for each possible value of the vector n_{mt}.

2.3 Profit function

Every period t, incumbent firms decide their respective prices, p_{imt}, and whether to remain in the market or exit at next period, $a_{im,t+1}$. Potential entrants choose whether to enter or not in the market. Current profits of airline i in market m at period t are:

$$\Pi_{imt} = \begin{cases} 0 & if \quad \{a_{imt} = 0\} \text{ and } \{a_{im,t+1} = 0\} \\ -EC(\alpha_{imt}, n_{mt}) - \sigma_E \; \varepsilon_{E,imt} & if \quad \{a_{imt} = 0\} \text{ and } \{a_{im,t+1} = 1\} \\ R(\alpha_{imt}, n_{mt}) - FC(\alpha_{imt}, n_{mt}) & if \quad \{a_{imt} = 1\} \text{ and } \{a_{im,t+1} = 1\} \\ R(\alpha_{imt}, n_{mt}) - FC(\alpha_{imt}, n_{mt}) + EV(\alpha_{imt}, n_{mt}) - \sigma_X \; \varepsilon_{X,imt} & if \quad \{a_{imt} = 1\} \text{ and } \{a_{im,t+1} = 0\} \end{cases} \tag{8}$$

The first regime, with $\Pi_{imt} = 0$, represents the case of a potential entrant that decides to stay out of the market. The second regime, with $\Pi_{it} = -EC(\alpha_{imt}, n_{mt}) - \sigma_E \; \varepsilon_{E,imt}$, is the profit of a potential entrant who decides to be active in the next period. Current profits in this case are equal to entry costs. Note that entry costs are paid at period t, but the firm starts to operate at period $t+1$. Entry costs have two components. The first component, $EC(\alpha_{imt}, n_{mt})$, is common knowledge to all firms in the market. The term $\sigma_E \; \varepsilon_{E,imt}$ represents a component of the entry cost that is private information of the firm. σ_E is a parameter, and $\varepsilon_{E,it}$ is a standard normal random variable that is independently and identically distributed over firms and over time. The parameter σ_E is common knowledge but $\varepsilon_{E,imt}$ is private information of the firm.

The third regime of the profit function, with $\Pi_{imt} = R(\alpha_{imt}, n_{mt}) - FC(\alpha_{imt}, n_{mt})$, corresponds to an incumbent airline that stays in the market. The first term is the variable profit, and $FC(\alpha_{imt}, n_{mt})$ is the fixed operating cost, that depends on the own hub index and on the hub index of competitors. Finally, the fourth regime, with $\Pi_{it} = R(\alpha_{imt}, n_{mt}) - FC(\alpha_{imt}, n_{mt}) + EV(\alpha_{imt}, n_{mt}) - \sigma_X \; \varepsilon_{X,imt}$, is the profit of an incumbent firm that decides to exit from the market at the end of period t. This exiting firm is operative during period t, it obtains variable profits pays fixed costs. It also receives an exit value $EV(\alpha_{imt}, n_{mt}) - \sigma_X \; \varepsilon_{X,imt}$. The component $EV(\alpha_{imt}, n_{mt})$ of the exit value is common knowledge. The variable $\varepsilon_{X,imt}$ is private information and it has the same statistical properties as $\varepsilon_{E,imt}$ We use the vector $\varepsilon_{imt} \equiv (\varepsilon_{E,imt}, \varepsilon_{X,imt})$ to represent the set of private information shocks.

2.4 Transitions of state variables

We have define above the vector n_{mt} that contains the number of incumbent firms for each possible hub index level α and for each group. Let e_{mt} be the same type of vector but for the potential entrants in market m at period t.

$$e_{mt} = \{ \; e^g_{mt}(\alpha) : g \in \{1, 2, 3\}, \;\; \alpha \in \Omega \; \} \tag{9}$$

where $e^g_{mt}(\alpha)$ is the number of potential entrants of type g with cost-adjusted hub index α in market m at period t. The vector e_{mt} is a payoff-relevant state vector because the decisions of potential entrants, and therefore an airline's expected future profits, depend on this vector.

The vector of payoff relevant state variables is $(a_{imt}, \alpha_{imt}, n_{mt}, e_{mt}, \varepsilon_{imt})$. As described above, the private information variables in ε_{imt} are i.i.d. standard normal random variables. Hub index α_{imt} follows an exogenous Markov process with transition probability function $F_\alpha(\alpha_{im,t+1}|\alpha_{imt})$. The incumbent status a_{imt} and the vectors with the distribution of hub index of incumbent firms and potential entrants, n_{mt} and e_{mt}, follow endogenous stochastic processes which are determined in the equilibrium of the entry-exit game. For the moment, assume that (n_{mt}, e_{mt}) follows a first order Markov process with transition probability function $Q(n_{m,t+1}, e_{m,t+1}|n_{mt}, e_{mt})$.

2.5 Markov perfect equilibrium

We assume that an airline's strategy in market m depends only on its payoff relevant state variables in that market: $(a_{imt}, \alpha_{imt}, n_{mt}, e_{mt}, \varepsilon_{imt})$. For notational simplicity, we use the vector x_{imt} to describe all the common knowledge state variables associated with the decision of airline i: $x_{imt} \equiv (a_{imt}, \alpha_{imt}, n_{mt}, e_{mt})$.

Let $\sigma \equiv \{\sigma_i(x_{imt}, \varepsilon_{imt}) : i \in I\}$ be a set of strategy functions, one for each airline, such that $\alpha_i : \{0,1\} \times |\Omega|^N \times \mathbb{R}^2 \to \{0,1\}$. A Markov Perfect Equilibrium (MPE) in this game is a set of strategy functions such that each airline's strategy maximizes the value of the airline (in the local market) for each possible state $(x_{imt}, \varepsilon_{imt})$ and taking other firms' strategies as given.

In order to describe the equilibrium mapping that characterizes a MPE in this game, it is useful to define *conditional choice value functions*. Suppose that all the airlines other than i behave now and in the future according to their strategies in σ, and airline i chooses a_i at period t and then follows his strategy σ_i in the future. The value of airline i under this behavior is:

$$\pi_i(a_i, x_{imt}, \varepsilon_{imt}) + v_i^\sigma(a_i, x_{imt}) \tag{10}$$

$\pi_i(a_i, x_{imt}, \varepsilon_{imt})$ is the current profit and $v_i^\sigma(a_i, x_{imt})$ is the present value of the stream of future profits:

$$v_i^\sigma(a_i, x_{imt}) \quad \equiv \quad \textstyle\sum_{\tau=t+1}^{\infty} \beta^{\tau-t} \; E\left(\; \pi_i(\sigma_i\left(x_{im\tau}^{\sigma,a_i}, \varepsilon_{im\tau}\right), \; x_{im\tau}^{\sigma,a_i}, \; \varepsilon_{im\tau}) \mid a_i, x_{imt}\right) \tag{11}$$

where the expectation is taken over all the possible future paths of $\{x_{im\tau}, \varepsilon_{im\tau}\}$ conditional on .(a_i, x_{imt}) and on the strategies in σ We use the notation $x_{im\tau}^{\sigma,a_i}$ to emphasize that the stochastic process of the state variables in x depends on the strategies in σ and on the initial choice a_i. Define also the *differential value function* $\tilde{v}_i^\sigma(x_{imt})$ as:

$$\tilde{v}_i^\sigma(x_{imt}) \equiv v_i^\sigma(1, x_{imt}) - v_i^\sigma(0, x_{imt}) \tag{12}$$

Let $\psi_i(x_{imt}, \varepsilon_{imt}; \sigma)$ be airline i's best response function. Given the definition of the differential value function $\tilde{v}_i^\sigma$, we can describe this best response function as follows. First, consider the case where airline i is an incumbent in market m. Then, his best response is:

$$\psi_i(x_{imt}, \varepsilon_{imt}; \sigma) \quad = \quad I\left\{ \; \tilde{v}_i^\sigma(x_{imt}) - EV(\alpha_{imt}, n_{mt}) - \sigma_X \; \varepsilon_{X,imt} \geq 0 \; \right\} \tag{13}$$

where $I\{.\}$ is the true operator. Similarly, if airline i is not active in market m, his best response is:

$$\psi_i(x_{imt}, \varepsilon_{imt}; \sigma) \quad = \quad I\left\{ \; \tilde{v}_i^\sigma(x_{imt}) - EC(\alpha_{imt}, n_{mt}) - \sigma_E \; \varepsilon_{E,imt} \geq 0 \; \right\} \tag{14}$$

Note that this function is a best response to other firms' strategies but also to the own firm strategy α_i. That is, this best response function incorporates a *'policy iteration'* in the firm's dynamic programming problem. The *Representation Lemma* in Aguirregabiria and Mira (2006) shows that we can use this type of best response functions to characterize every MPE in the model. That is, a set of strategy functions is a MPE in this model if and only if these strategies are a fixed point of this best response function.

DEFINITION: A set of strategy function $\sigma^* \equiv \{\sigma_i^*(x_{imt}, \varepsilon_{imt}) : i \in I\}$ is a MPE in this model if and only if for any firm i and any state $(x_{imt}, \varepsilon_{imt})$ we have that:

$$\sigma_i^*(x_{imt}, \varepsilon_{imt}) = \psi_i(x_{imt}, \varepsilon_{imt}; \sigma^*) \tag{15}$$

Equations (13) and (14) show that the only way in which the set of strategies σ enters in the best response function ψ_i is through the differential value function $\tilde{v}_i^\sigma$. Therefore, we can write the best response function as $\psi_i(x_{imt}, \varepsilon_{imt}; \tilde{v}^\sigma)$. This representation is useful to characterize a MPE in this model in terms of the value functions $\{\tilde{v}_i^\sigma : i \in I\}$. And that characterization is useful to compute an equilibrium. The following Proposition establishes this result.

PROPOSITION 1: Let $\sigma \equiv \{\sigma_i(x_{it}, \varepsilon_{it}) : i \in I\}$ be a set of strategy functions. And let $\tilde{v}^\sigma \equiv \{\tilde{v}_i^\sigma : i \in I\}$ be the differential value functions associated with σ as we have defined them in equation (12). Then, σ is a MPE if and only if the vector of value functions $\tilde{v}^\sigma$ is a solution to the fixed-point problem $v = \Gamma(v)$ where $\Gamma(v) = \{\Gamma_i(v) : i \in I\}$ and:

$$\begin{aligned}\Gamma_i(v)(x_{it}) \;&=\; \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} \quad \Big[E\left(\pi_i\left(\psi_i(x_{i\tau}^{\psi(v),1}, \varepsilon_{i\tau}; v),\; x_{i\tau}^{\psi(v),1},\; \varepsilon_{i\tau}\right) \mid 1, \mathbf{x}_{it}\right) \\ &- \qquad E\left(\pi_i\left(\psi_i(x_{i\tau}^{\psi(v),0}, \varepsilon_{i\tau}; v),\; x_{i\tau}^{\psi(v),0},\; \varepsilon_{i\tau}\right) \mid 0, \mathbf{x}_{it}\right)\Big]\end{aligned} \tag{16}$$

where the expectation is taken over all the possible future paths of x and ε. The vector $x_{i\tau}^{\psi(v),a_i}$ represents a realization of the vector of state variables $x_{i\tau}$ when firms behave using strategies $\{\psi_i(v) : i \in I\}$ and firm i's initial choice is a_i.

Proof: TBW.

Let $\tilde{v}^* = \{\tilde{v}_i^* : i \in I\}$ be the set of differential value functions in a MPE of this game. Taking into account the expressions for the best response function in equations (13) and (14) we have that the equilibrium probability of exit is such that:

$$P_X(x_{imt}) = \Phi\left(\frac{EV(\alpha_{imt}, n_{mt}) - \tilde{v}(x_{imt})}{\sigma_X}\right) \tag{17}$$

And the equilibrium probability of entry is:

$$P_E(x_{imt}) = \Phi\left(\frac{\tilde{v}(x_{imt}) - EC(\alpha_{imt}, n_{mt})}{\sigma_E}\right) \tag{18}$$

As pointed out by Bresnahan and Reiss (1993), the difference between the probabilities of entry and staying in the market provides information about the magnitude and the structure of sunk costs. More precisely, note that:

$$EC(\alpha_{imt}, n_{mt}) - EV(\alpha_{imt}, n_{mt}) = -\sigma_X \ \Phi^{-1}\left(P_X(x_{imt})\right) - \sigma_E \ \Phi^{-1}\left(P_E(x_{imt})\right) \tag{19}$$

$EC(\alpha_{imt}, n_{mt}) - EV(\alpha_{imt}, n_{mt})$ is the sunk component of the entry cost and $\Phi^{-1}(.)$ is the inverse of the CDF of the standard normal.

Given these entry and exit equilibrium probability functions and the exogenous stochastic process of α_{imt}, we can obtain the Markov process $Q(n_{m,t+1}, e_{m,t+1}|n_{mt}, e_{mt})$.

2.6 A simulator of the equilibrium mapping

For the estimation of the model, we approximate the equilibrium mapping $\Gamma_i(v)$ using 'forward simulation' as suggested by Hotz *et. al.* (1994) and Bajari *et. al.* (2005). This involves simulating paths for the error structure and for the exogenously evolving state variables, and then computing optimal behavior along these paths according to $\Gamma(.)$ and $\psi(.)$. We then average the simulated values over several (many) simulated paths.

More precisely, we consider the following simulator of the equilibrium mapping:

$$\tilde{\Gamma}_i^R(v)(x_{it}) \quad = \quad \frac{1}{R}\sum_{r=1}^{R}\left(v_{it,r}^{\psi(v),1} - v_{it,r}^{\psi(v),0}\right) \tag{20}$$

where

$$v_{it,r}^{\psi(v),a_i} \quad = \quad \sum_{\tau=t+1}^{t+T} \beta^{\tau-t} \ \pi_i\left(\psi_i(x_{i\tau,r}^{\psi(v),a_i}, \varepsilon_{i\tau,r}; v), \ x_{i\tau,r}^{\psi(v),a_i}, \ \varepsilon_{i\tau,r}\right) \tag{21}$$

$\{x_{i\tau,r}^{\psi(v),a_i}, \varepsilon_{i\tau,r}\}$ are the simulated paths of the state variables for the r-th simulation, R is the number of simulated paths, and T is the number of periods ahead of the simulated paths.

3 Data and descriptive statistics

The data set that we use in this paper is the domestic Origin and Destination Survey (DB1B) from the U.S. Bureau of Transportation Statistics (BTS). This survey is based on a 10% sample of the airline tickets in the U.S. airline market. The frequency of the data is quarterly. Each observation in the raw DB1B data consists of a unique airline itinerary, including the starting and ending airports of each flight coupon, the market fare of the ticket, the reporting carrier, the nonstop distance between starting and ending airport, and the quarter. Our working sample covers the first and third quarters of years 2003 and 2004. The airline industry has experienced significant changes just before our sample period. Some of the most important recent shocks affecting this industry are: the September 11th terrorist attacks in 2001; US Airways and United

Airlines filed for Chapter 11 bankruptcy protection in August and December 2002, respectively;[4] and low-cost carriers like Southwest, jetBlue and AirTran have continued to expand and remain profitable.

In the construction of our working sample, we follow the same criteria as those in previous studies using this data set, such as Borenstein (1989), Berry (1992), and, very particularly, Ciliberto and Tamer (2006). A market is defined as a directional airport pair. That is, passengers traveling from Boston to Chicago are in a different market than passengers traveling between Chicago to Boston. We use the 2004 population data from population division of U.S. Census Bureau to find out the 50 largest U.S. cities and then we restrict the sample to these cities. Intermediate transfers do not alter the market definition which is only concerned with the starting and ending airports. Following Ciliberto and Tamer (2006), but unlike Berry (1992), the market is defined at the airport level but not at the city level. As argued by Ciliberto and Tamer, defining the market at city level ignores some segments of the consumer demand can have strong preference on airport in the same city. Our data consists of 2135 markets and 34 carriers. The maximum number of airlines serving a market is 20.

We aggregate the itinerary information using reporting carrier, thus the observation is indexed by airline-market-year. For the demand estimation, we use the sample of Q1-2004 and Q1-2004 since the firm use the information in Q1 to make the entry and exit decision in Q3. For the entry and exit estimation, we use all the airlines that appear once in the whole sample to be global player. The existing airlines are incumbents and the rest are potential entrants.

Table 1 presents statistics on entry and exit in local markets for the 25 largest carriers between the first and the third quarters of 2004. This table includes the number of markets served by each airline in Q1-2004, the airline-specific entry rates between Q1-2004 and Q3-2004, the airline-specific exit rates between Q1-2004 and Q3-2004, and the average market presence in Q1-2004. Entry rates are calculated as the number of new markets that the airline entered divided by the total number of markets that the airline operates in Q3-2004. Exit rates are equal to the number of markets that the airline has quit divided by the total number of markets that the airline operates in Q3-2004. Market presence is constructed by using market share of each carrier in each market. Heterogeneity across airlines can be observed in the number of market served and the market presence in each airport pair. The number of markets served by a carrier varies from more than 1800 to less than 100. The gross flow of entry and exit is more stable for large carriers than for the other types of airlines. Entry rates are typically higher than exit rates because the US airline industry is under recovery from the crunch in year 2001.

[4] US Airways emerged from Chapter 11 after fewer than eight months but with smaller capacity and 36% fewer employees, 30% fewer available seats and 25% fewer flights than it did before the Sept. 11 attacks, and has cut its annual costs by $1.9 billion. Its debt load is down to about $8 billion from $10.65 billion.

There is significant cross airlines heterogeneity in entry and exit rates.

Table 2 provides information on market characteristics. The summary statistics are reported for the all markets and for each quartile in terms of total population of the origin-destination city-pair. We present total population at the origin and destination cities, the non-stop distance of the markets and an indicator for "tourist market" for airports in California and Florida. The last quartile composed of markets with much larger size, it is on average three times larger than that in other quartiles. The average non-stop distance are similar for each group of markets. The average yield and total revenue generated by large markets are higher than the other. However, very interestingly, we do not observe that larger cities support more incumbents. Markets in group 1 with a population of one million consumers have on average 9.5 incumbent firms, and markets in group 4, with more than four million consumers, have on average 10.9 incumbents. This evidence is consistent with the existence of endogenous sunk costs. As we mentioned before, the airline market in the sample year experience recovery, here we observe the phenomenon does not skew towards a particular type of market since the average entry rates in all types of markets are higher than the exit rate..

4 Estimation of the structural model

In this section, we describe our approach to estimate the structural model. First, we estimate the demand system using information on prices, quantities and airline characteristics which related to quality. The estimation of the demand follows Berry (1994), Berry, Levinshon and Pakes (1995) and Nevo (2000). Second, we use the best response functions in the Bertrand game to estimate variable costs. Given these estimates we construct the *cost-adjusted qualities* α_{imt}. For the estimation of the dynamic game of entry and exit, we discretize these qualities and apply the method in Aguirregabiria and Mira (2005). The estimation of the dynamic game provides estimates of the structure of entry costs, fixed costs and exit values.

4.1 Estimation of the demand system

We consider the following specification of the airline-market-specific qualities $\tilde{\alpha}_{imt}$ in terms of observable and unobservable variables for the econometrician.

$$\tilde{\alpha}_{imt} = \gamma_1 \; HUB_{imt}^{O} + \gamma_2 \; HUB_{imt}^{D} + \gamma_3 \; DIST_m + \xi_i^{(1)} + \xi_{Omt}^{(2)} + \xi_{Dmt}^{(3)} + \xi_{imt}^{(4)} \tag{22}$$

γ_1, γ_2 and γ_3 are parameters. HUB_{imt}^{O} and HUB_{imt}^{D} are indexes that represent the scale of operation of airline i in the origin and destination airports of market m, respectively. $DIST_m$ is the nonstop distance between the origin and destination cities. We include this variable as a proxy of the value of air transportation relative to the outside alternative (i.e., relative to other transportation modes). Air transportation is a more attractive transportation mode when distance is relatively large. $\xi_i^{(1)}$ is an airline fixed-effect that captures differences between airlines' qualities which are constant over time and across markets. $\xi_{Omt}^{(3)}$ represents the interaction of origin-city dummies and time dummies, and $\xi_{Dmt}^{(4)}$ captures the interaction of destination-city dummies and time dummies. These two terms account for aggregate shocks, such as seasonal effects, which can vary across cities and over time. $\xi_{imt}^{(4)}$ is an airline-market-time specific demand shock.

The variables HUB_{imt}^O and HUB_{imt}^D deserve more explanation. These indexes are calculated excluding operation in market m. That is,

$$HUB_{imt}^O = \sum_{c \neq m} I_{imct}^O \left(\frac{Q_{mct}^O}{\sum_{c'} Q_{mc't}^O} \right) \tag{23}$$

where I_{imct}^O is the indicator of the event "airline i provides service between the origin city of market m and city c". This index is weighted by the importance of market (m, c), that is measured by $Q_{mct}^O / \sum_{c'} Q_{mc't}^O$ where Q_{mct}^O is the total number of passengers between the origin city of market m and city c. Essentially, this index adds up all the routes that airline i flies out from the origin airport, others than the route in market m. Given that some routes are more popular than others (i.e., they are more valued by consumers), we weight routes by the proportion of passengers travelling in this route over all passenger flying out from the origin airport. It measures the scale of operation of an airline out of the origin airport and it reflects services provided by the airline to consumers. A larger hub operation implies more convenient flight schedules and landing facilities. The hub variable for destination is constructed analogously. There are several studies that have considered this type of measures for airlines' quality. Berry (1990, 2006) suggests that airport presence can be used to capture the firm's ability to differentiate itself from the other to increase the demand and charge at a higher price. Borenstein (1991) show that airline with hub advantage charge higher premium than the other airlines after controlling cost variables. Januszeski (2004) shows that price drops due to flight delays (i.e. poor service quality) are less important if the competition of the city pair is less intense. Lederman (2004) estimates a differentiated demand system and finds that frequent flyer plan is useful to increase airline demand when the airline dominate the airports of end points.

Define the vector of regressors:

$$\begin{aligned} x_{imt} \equiv \; & (\; HUB_{imt}^O, \; HUB_{imt}^D, \; DIST_{mt}, \text{ Airline Dummies}, \\ & \text{Origin City Dummies} \times \text{ Time Dummies, Destination City Dummies} \times \text{ Time Dummies }) \end{aligned} \tag{24}$$

And let γ the vector of parameters associated with these regressors. Therefore, we can write airlines' qualities as $\alpha_{imt} = x_{imt} \, \gamma + \xi_{imt}^{(4)}$. Given the nested logit specification of consumers' heterogeneity, the demand system can be written as:

$$\ln(s_{imt}) - \ln(s_{0mt}) \quad = \quad x_{imt}\,\gamma - \gamma_p\, p_{imt} + \sigma_v \;\ln\left(s_{i|gmt}\right) + \xi^{(4)}_{imt} \tag{25}$$

where s_{imt} is the market share of airline i in the city-pair m at period t, i.e., $s_{imt} \equiv q_{imt}/H_{mt}$, with q_{imt} being the number of tickets sold by airline i; $s_{0mt} \equiv 1 - \sum_{i=1}^{N} s_{imt}$ is the share of the outside alternative; $s_{i|gmt}$ is the market share of the airlines i in the market m; and $\sigma_v \in [0,1]$ is the parameter govern the cross-elasticity.

The main econometric issue in the estimation of the demand system in equation (25) is the endogeneity of prices and market shares $\ln\left(s_{i|gmt}\right)$. Equilibrium prices depend on the characteristics (observable and unobservable) of all products, and therefore the regressor p_{imt} is correlated with the unobservable $\xi^{(4)}_{imt}$. This correlation is positive and therefore the OLS estimator of γ_p is downward biased, i.e., it underestimates the own-price demand elasticities. Similarly, the regressor $\ln\left(s_{i|gmt}\right)$ depends on unobserved characteristics and it is endogenous. Our approach to deal with this endogeneity problem combines the *control function* and the *instrumental variables* approaches. First, airline dummies, and the interaction of city dummies and time dummies capture part of the unobserved heterogeneity (i.e., control function approach). And second, to control for the endogeneity associated with the unobservable $\xi^{(4)}_{imt}$ we use instruments.

The specification of the stochastic process of $\xi^{(4)}_{imt}$ is particularly important to determine which instruments are valid in the estimation of demand parameters. We consider the following assumption:

$$\text{For any } i \neq j \text{ and } m \neq m' : \quad \xi^{(4)}_{imt} \perp \xi^{(4)}_{jm't} \tag{26}$$

The idiosyncratic demand shocks of two different airlines at two different markets are independently distributed. There may be cross market correlation in the demand shocks of an airline, and there may be correlation between the shocks of two airlines in the same market. However, once we have accounted for $\xi^{(2)}_{Omt}$ and $\xi^{(3)}_{Dmt}$, there is independence between the shocks of two airlines at different markets. Note that the variables HUB^{O}_{imt} and HUB^{D}_{imt} depend on the entry decisions of airline i in markets different than m. Therefore, these variables depend on the demand $\xi^{(4)}_{im't}$ in markets m' different than m. The variables HUB^{O}_{jmt} and HUB^{D}_{jmt} for an airline $j \neq i$ are correlated with $\xi^{(4)}_{jm't}$ but, under our assumption in (26), they can be independent of $\xi^{(4)}_{imt}$. We explicitly assume that:

$$E\left(\xi^{(4)}_{imt} \mid HUB^{O}_{jmt},\, HUB^{D}_{jmt}\right) = 0 \tag{27}$$

Furthermore, by the equilibrium condition, prices depend on the characteristics HUB^{O}_{jmt} and HUB^{D}_{jmt} of every active firm in the market. Therefore, we can use HUB^{O}_{jmt} and HUB^{D}_{jmt} as instruments for the price p_{imt} and the market share $\ln\left(s_{i|gmt}\right)$. Indeed, any function of the variables $\{HUB^{O}_{jmt}, HUB^{D}_{jmt} : j \neq i\}$ can be used as an instrument.

As usual, to avoid the small sample bias of IV estimation, we want to use the smallest number of instruments with the largest explanatory power. We construct the following six instruments. For $g = 1, 2, 3$ and $h = O, D$, we define the instrumental variables:

$$IV^{h,g}_{imt} = \frac{1}{N_g} \sum\nolimits_{j \neq i, j \in g} HUB^{h}_{jmt} \tag{28}$$

The variable $IV^{h,g}_{imt}$ represents the average hub index at the origin ($h = O$) or destination ($h = D$) airports for airlines other than i that belong to group g. This variable correlates with the price and within group market share since the hub operation in other market improve the competitiveness of these carriers in the market m, but it is not endogenous since the decision in those market of competing carriers is due to $\xi^{(4)}_{jm't}$ which is independent of $\xi^{(4)}_{imt}$ by the identifying assumption (26).

Tables 3 presents estimates of the demand system. To illustrate the endogeneity problem, we first report the estimation results from OLS and fixed effects. The price coefficient implies an own price elasticity at the average fare of 0.14, which is too small whereas the estimate for σ_v is close to 1 which indicates the choice between flying through plane and outside alternative dominate in the decision making. In the second column, the market fixed effects cannot alleviate the problem. This indicates that the endogeneity cannot be solved only using time-invariant market effect. Thus, we proceed to instrumental variable estimation and test the validity of different set of instruments by using a Sargan-Hansen test of over-identifying restrictions. The results of the IV estimation are shown in last column in table 3. The magnitude of price coefficients is much larger than that from OLS and fixed effect estimation with the price elasticity at the average fare is about 9.8. The cross elasticity with the outside good is higher than before with the estimate of σ_v at 0.35. The chi-square statistic of Sargan-Hansen test of overidentifying restrictions cannot reject the validity of the instruments. The estimated effects of the hub indexes are alo plausible. Expanding the scale of hub operation in origin and destination airports increase the demand where the effect from origin airport is stronger than that from the destination airport. The result is consistent with hub effect obtained in the literature such as Berry (1990). Finally, longer nonstop distance makes consumer more inclined to use airplane transportation than other transportation modes.

4.2 Estimation of variable costs and cost-adjusted hub index

With the estimates of the demand system, we use the supply side to recover the marginal cost for each airline and market. The best response functions of the Bertrand pricing game are:

$$p_{imt} = c_{imt} - \frac{s_{imt}}{\partial s_{imt}/\partial p_{imt}} \tag{29}$$

Given our nested logit specification, we have that:

$$\frac{\partial s_{imt}}{\partial p_{imt}} = \frac{-\gamma_p}{1-\sigma_v} \; s_{imt} \; \left(1 - \sigma_v \; s_{i|gmt} - (1-\sigma_v) s_{imt}\right) \tag{30}$$

Therefore, given our estimates of the parameters γ_p and σ_v and our information on prices and market shares, we can construct consistent estimates of the variables costs as:

$$c_{imt} = p_{imt} - \frac{1-\sigma_v}{\gamma_p \left(1 - \sigma_v \; s_{i|gmt} - (1-\sigma_v) s_{imt}\right)} \tag{31}$$

Finally, we compute estimates of hub index as:

$$\alpha_{imt} \equiv \frac{x_{imt} \; \gamma + \xi^{(4)}_{imt} - c_{imt}}{\sigma_v} \tag{32}$$

where $\xi^{(4)}_{imt}$ is obtained as the residual from the estimated demand. Table 4 presents average estimates of marginal costs, price-cost margins and hub index for selected airlines. The air fare is lower for low cost carrier, but it is not universal. The most prominent case of LCC is Southwest Airline that charges the fare at a much lower price and still achieves to earn a high price-cost margin. However, other low cost carriers do not share the same success as Southwest. Actually, the price-cost margin is higher for those regional airlines which do not belong to network carriers or low cost carriers. In terms of quality of services, the network carriers are higher on average mainly due to the services provided by the network carriers through their large scale operation in hub airports.

4.3 Estimation of the dynamic game

Unobserved market heterogeneity is an important issue for estimating the entry and exit model since it creates endogeneity problem for the regressor related to number of incumbents in the market. If the unobserved market fixed effect is strong, it creates upward bias in the coefficient of number of incumbents and produces a misleading conclusion that strategic interaction is not significant or even the competitors are strategic complement rather than substitute. The airline market is no exception to this problem. To illustrate this problem in our data, we perform the reduced form estimation of the following specification for firm i entering into the market m at time t if

$$Entry_{imt} \quad = \quad I\{\beta_X \ X_{mt} + \beta_N \log(1+\bar{n}_{mt}) + \beta_R R_{mt} + \beta + \varepsilon_{imt} \geq 0\} \tag{33}$$

where X includes constant, log(distance), log(population) and tourist and $\bar{n}_m$ is the number of incumbent in the beginning of the period. The specification also include a residual from the first stage regression

$$\log(revenue_{mt}) \quad = \quad \gamma_0 + \gamma_X X_{mt} + R_{mt} \tag{34}$$

which try to capture the demand side unobserved factor into the entry model. The results of column "Basic" in table 5 show that without controlling the unobserved market effect, the coefficient of number of incumbent is positive which is not economically plausible. As shown in column "Resid", the residual from the revenue equation can capture part of the endogeneity between the number of incumbents and unobserved market characteristics, but it is not enough to deliver a correct sign for the coefficient of number of incumbents. It indicates that the unobserved market characteristics is not only due to demand side factor but also the factors from cost side where the sunk cost and fixed operating cost that is not observed from researchers can be a reason of endogeneity. Contrary to the demand estimation, the endogeneity in the entry/exit model is due to the time-invariant market effect but not the unobserved demand shock which indicates sunk cost and fixed operating cost specific to each local market is an important determinant for entry and exit decision.

In order to capture the time-invariant market unobserved heterogeneity in both demand and cost sides, we use the panel structure of our data and create the variables, market type, by using a two-step procedure. First, we estimate the probability of active in market nonparametrically by counting the number of active firm and divide it by the total number of firm, N_t for each market and year as follow

$$\widehat{P}_{imt} \quad = \quad \frac{\sum_{i=1}^{N} a_{imt}}{N_t} \tag{35}$$

where a_{im1} is the indicator of being active in the market in Q3. Then, we estimate Chamberlain's conditional logit for the choice of being active in the market on a set of variables using the following specification

$$1(Active_{imt}) \quad = \quad \alpha_m + \alpha_P \log(Pop_{mt}) + \alpha_N \log(1+N_{mt}) + \alpha_{PE} a_{imt-1} + \varepsilon_{imt} \tag{36}$$

We recover the market fixed effects with the population heterogeneity by

$$\widehat{\alpha}_m \quad = \quad \ln\left(\frac{\widehat{P}_{imt}}{1-\widehat{P}_{imt}}\right) - \widehat{\alpha}_N \log(1+N_{mt}) - \widehat{\alpha}_{PE} a_{imt-1} \tag{37}$$

We interpret it as an index of profitability of the market. However, the estimate for the fixed effect parameters are not consistent. Using it as an explanatory variable will biased all estimates in the entry/exit regression. Alternatively, we rank the fixed effect parameter in an ascending order, categorize them into $G+1$ groups according to the profitability and create an indicator for each group, thus the measurement error is reduced since we do not use the exact value of the fixed effect parameter. For example, if we divide all the market into 10 groups, the markets with $\widehat{\alpha}_m$ ranked 1 to 200 are in group 1, 201 to 400 are in group 2 and so on. Since we have 2135 markets, the last 135 market with largest profitability index are always the last group. In this specification, we assume the average profitability is the same for all markets within the group but can be varied across groups.

In the column "G10" in table 5, it shows that with only $10+1$ groups the unobserved heterogeneity can well be captured and the sign of coefficient of number of incumbent is correct. The results do not change much when we partition the group into finer level and there is no more noticeable change if we divide the markets into more than 80 groups. Since the fixed effect include heterogeneity in population, the coefficient of population is still positive but the magnitude is much smaller than that in the previous specification. The entry probability is higher when the non-stop distance is longer which can due to the variable cost is more important in the cost structure. The impact of fixed portion of the cost structure is smaller in these markets and hence the entry barrier is lower. The consumer in the tourist market are more price sensitive which make the price competition in these markets are tougher, thus the probability of active in these market are smaller. The incumbent airlines are more incline to operate in the market than potential entrant which show sunk entry cost is an important concern in the entry decision. In the table, we show the coefficient of the group located at the rank 600, 1200, 1800. For instance, if we divided the group into 10, then $g600$ is the group 3 with the $\widehat{\alpha}_m$ ranked 601 to 800. Similarly, if we divide the group into 40, then $g600$ is the group 12 with the $\widehat{\alpha}_m$ ranked 601 to 650. For the markets in a more profitable group, the probability of operating in the market is higher for all firms and the difference between market groups are significant.

Regression with interaction term with incumbent status is shown in last two columns in table 5 which indicates significant difference between entry decision between potential entrant and incumbent. As the market size increases, the incumbents and potential entrants are more like to exit and enter respectively. It shows that large market is characterized by more entry and exit. Contrary to potential entrants, the distance do not affect much on the incentive of the incumbent to participate in a market since they have already paid the sunk cost and the variable cost are similar for market in all distances. Since the price competition in the tourist market can be tougher due to the more price sensitive consumers, it makes the entry and exit flow is low in these markets and the incumbents are more likely to stay. Finally, the impact of increasing the number of incumbents is stronger on incumbents since the exit barrier is lower than the entry barrier.

Before exploiting the whole structure of the model, we estimate a dynamic model with number of incumbent, incumbent status and market types as explanatory variables. We assume that the profitability of market is known to researcher. The variable profit is specified as $R(N_m) = \theta_M - \theta_N \log(1 + N_m)$ which is a decreasing function of the number of incumbent in the market. If there is no other firm in the market, the active firm will earn monopoly profit, θ_M. Therefore, the profit function of firm i active in a market m in group g is

$$\Pi_{imt}(g, N_{mt}) \quad = \quad \theta_{M,g} - \theta_N \log(1 + N_{mt}) - \theta_{FC,g} - \theta_{SC}(1 - a_{it-1}) + \varepsilon_{imt} \tag{38}$$

In this specification, we cannot separately identify the monopoly profit and fixed cost but the net difference between these two amount, i.e. monopoly's net variable profit, in each profitability group. The coefficient θ_N represent the strategic interaction between airlines in the entry and exit game whereas the sunk cost is estimated by θ_{SC}. The dynamic game of entry and exit is estimated by Aguirregabiria and Mira (2005). It is initiated by the estimated conditional choice by logit with the following specification

$$P(a_{imt}|g, N_{mt}) \quad = \quad \frac{\exp(\theta_{M,g} - \theta_N \log(1 + N_{mt}) - \theta_{FC,g} - \theta_{SC}(1 - a_{imt-1}))}{1 + \exp(\theta_{M,g} - \theta_N \log(1 + N_{mt}) - \theta_{FC,g} - \theta_{SC}(1 - a_{imt-1}))} \tag{39}$$

Table 6 presents the results from nested pseudo likelihood (NPL) estimation of dynamic game. The estimation converge before 20 iteration and the results presented under the column "20th NPL" is equivalent to the asymtotically efficient estimates by maximum likelihood estimation where the standard error is shown to be smaller than the estimates from logit and NPL in the 1st iteration. Thus, we focus on the results from the 20th iteration of NPL. The strategic interaction is positive significantly which show more incumbents in the market reduce the probability of entry. The sunk cost is positive and significant. In terms of economic significance, it is about 85% (=4.382/5.183*100%) of the net variable profit that a monopoly can earn in a market with average profitability. There is heterogeneity in the monopoly's net variable profit from different types of markets. A market in group 8 with the rank from 1601 to 1800 is 47% (=6.472/4.412*100%) more profitable than a market in group 2 with rank 201 to 400. Moreover, we are interested in obtaining evidences of endogenous sunk cost as advocated in Sutton (1991) in the US airline industry. Therefore, we extend the above specification with the interaction term between number of incumbent and the incumbent status and obtain the information of sunk cost as a function of number of incumbent as $\theta_{SC}(1 - \frac{\theta_{NSC}}{\theta_{SC}} \log(1 + N_{mt}))$ where θ_{NSC} is the coefficient for the interaction term. The sunk cost can be approximated by $5.8(1 - 0.1 \log(1 + N_{mt}))$. In the sample, the maximum number of firm is 34. Even if all the potential entrant enter into the market, the sunk cost is still about 64% of that encounter by the first entrant which is equivalent to 51% of the monopoly net variable profit in the median group. It shows that the entry barrier in airline industry cannot be eliminated even with such a large number of airlines. The widespread phenomenon of large hub operation of airline can be a form of endogenous sunk cost which hinder entry and exit process to achieve the competitive outcome.

5 Conclusion

TBW

References

[1] Aguirregabiria, V., and P. Mira (2006): “Sequential estimation of dynamic discrete games,” *Econometrica,* forthcoming.

[2] Bailey, E., D. Graham, and D. Kaplan (1985): *Deregulating the Airlines*. Cambridge, Mass.: The MIT Press.

[3] Bajari, P., L. Benkard and J. Levin (2005): “Estimating dynamic models of imperfect competition,” *Econometrica,* forthcoming.

[4] Berry, S. (1992): “Estimation of a model of entry in the airline industry”, *Econometrica,* 60(4), 889-917.

[5] Berry, S. (1994): “Estimating Discrete-choice Models of Product Differentiation.” *RAND Journal of Economics*, 25, 242–262.

[6] Berry, S., M. Carnall, and P. Spiller (2006): “Airline Hubbing, Costs and Demand,” in *Advances in Airline Economics*, Vol. 1: Competition Policy and Antitrust, D. Lee, ed. Elsevier Press.

[7] Berry, S, J. Levinsohn, and A. Pakes (1995). “Automobile Prices in Market Equilibrium.” *Econometrica*, 63, 841–890.

[8] Boguslaski, C., H. Ito, and D. Lee (2004), "Entry Patterns in the Southwest Airlines Route System", *Review of Industrial Organization* 25(3), 317-350.

[9] Borenstein, S. (1989): “Hubs and high fares: Dominance and market power in the US airline industry”, *Rand Journal of Economics,* 20(3), 344-365.

[10] Borenstein, S. (1991), "The Dominant Firm-Advantage in Multiproduct Industries: Evidence from the U.S. Airlines", *Quarterly Journal of Economics* 106(4), 1237-1266.

[11] Borenstein, S. "The Evolution of U.S. Airline Competition." *Journal of Economic Perspectives*, Vol. 6 (1992).

[12] Bresnahan, T., and P. Reiss (1990): “Entry in monopoly markets”, *Review of Economic Studies*, 57, 531-553.

[13] Bresnahan, T., and P. Reiss (1991): “Entry and competition in concentrated markets”, *Journal of Political Economy,* 99, 977-1009.

[14] Bresnahan, T., and P. Reiss (1993): “Measuring the importance of sunk costs”, *Annales D'Economie Et De Statistique*, 31, 181-217.

[15] Carndell, S. (1997): "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity ," *Econometric Theory*, 13, 185-213.

[16] Ciliberto, F. and E. Tamer (2006): "Market structure and multiple equilibria in airline markets," manuscript, University of Virginia.

[17] Ericson, R., and A. Pakes (1995): “Markov-Perfect industry dynamics: A framework for empirical work”, *Review of Economic Studies,* 62, 53-82.

[18] Gayle, K. (2004): "Does Price Matter? Price or Non-Price Competition in the Airline Industry ," manuscript, Department of Economics, Kansas State University.

[19] Hendricks, K., M. Piccione, and G. Tan (1997): “Entry and exit in hub-spoke networks,” *Rand Journal of Economics*, 28(2), 291–303.

[20] Januszewski, J (2004): "The Effect of Air Traffic Delays on Airline Prices," manuscript, Department of Economics, UC San Diego.

[21] Januszewski, J. and M. Lederman (2003): "Entry Patterns of Low-Cost Airlines ," manuscript, Rotman School of Management, University of Toronto.

[22] Lederman, M. (2004): "Do Enhancements to Loyalty Programs Affect Demand? The Impact of International Frequent Flyer Partnerships on Domestic Airline Demand ," manuscript, Rotman School of Management, University of Toronto.

[23] Kahn, A. (2004): *Lessons from Deregulation: Telecommunications and Airlines After the Crunch*, Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

[24] Pakes, A. and P. McGuire (1994): “Computing Markov-Perfect Nash equilibria: Numerical implications of a dynamic differentiated product model, *Rand Journal of Economics*, 555-589.

[25] Pakes, A., M. Ostrowsky, and S. Berry (2005): “Simple estimators for the parameters of discrete dynamic games (with entry/exit examples),” manuscript, Harvard University.

[26] Pesendorfer, M. and Schmidt-Dengler (2004): “Identification and estimation of dynamic games,” London School of Economics Working Paper.

[27] Morrison, S. A. and C. Winston, (1986): *The Economic Effects of Airline Deregulation*, Brookings, Washington, D. C.

[28] Morrison, S. A. and C. Winston, (1989): "Enhancing the Performance of the Deregulated Air Transportation System," *Brookings Papers on Microeconomics*.

[29] Morrison, S. A. and C. Winston, (1995): *The Evolution of the Airline Industry*, Brookings, Washington, D.C.

[30] Morrison, S. A., and C. Winston. (2000): "The remaining role for government policy in the deregulated airline industry," in *Deregulation of network industries: What's next?*, edited by Sam Peltzman and Clifford Winston. Washington, DC: AEI-Brookings Joint Center for Regulatory Studies.

[31] Reiss, P.C. and P.T. Spiller (1989), "Competition and Entry in Small Airline Markets", *Journal of Law and Economics* 32, S179-S202.

[32] Sinclair, R.A. (1995), "An Empirical Model of Entry and Exit in Airline Markets", *Review of Industrial Organization* 10, 541-557.

[33] Sutton, J., (1991): *Sunk Cost and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT Press, Cambridge.

[34] Tamer, E (2003): "Incomplete Bivariate Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–167.

[35] Transportation Research Board (1999): "Entry and Competition in the U.S. Airline Industry: Issues and Opportunities," National Research Council, Special Report 255, Washington, D.C.

Table 1
Market Presence, and Entry and Exit Rates for the Largest 25 Carriers in 2004

Carrier Code	Carrier	# Markets Served Q3-2004	Market Presence Q1-2004	Entry Rate between Q1-2004 and Q3-2004	Exit Rate between Q1-2004 and Q3-2004
AA	American	1896	0.085	0.028	0.025
DL	Delta	1882	0.082	0.024	0.027
UA	United	1809	0.078	0.032	0.034
NW	Northwest	1689	0.071	0.043	0.027
YV	Mesa	1670	0.062	0.198	0.093
CO	Continental	1637	0.067	0.086	0.042
RU	Expressjet	1581	0.062	0.107	0.066
MQ	American Eagle	1524	0.059	0.134	0.073
OH	Comair	1471	0.057	0.116	0.128
EV	Atlantic Southeast	1363	0.049	0.193	0.150
ZW	Air Wisconsin	1262	0.039	0.311	0.132
HP	America West	1088	0.050	0.063	0.049
US	US Air	1071	0.040	0.134	0.071
XJ	Mesaba	866	0.028	0.278	0.176
WN	Southwest	846	0.054	0.074	0.011
AX	Trans States	785	0.026	0.247	0.178
F9	Frontier	597	0.022	0.224	0.077
QX	Horizon	461	0.015	0.299	0.082
TZ	ATA	443	0.018	0.199	0.183
F8	Freedom	428	0.023	0.086	0.521
16	PSA	394	0.000	1.000	0.000
AS	Alaska	389	0.010	0.401	0.077
YX	Midwest Express	324	0.013	0.259	0.133
FL	AirTran Airways	252	0.026	0.028	0.016
HA	Hawaii	81	0.004	0.086	0.185

Source: DB1B.

Table 2
Market Characteristics

Variable	All Markets	Group 1	Group 2	Group 3	Group 4
Population	1,940,395	943,795	1,151,261	1,572,750	4,091,908
Distance	1268	1428	1210	1258	1175
Tourist	0.24	0.17	0.16	0.33	0.28
Revenue	123,476	84,573	90,231	112,024	207,005
Yield	0.19	0.16	0.20	0.20	0.22
Incumbents	10.47	9.54	10.65	10.80	10.91
Entry	1.75	1.47	1.76	1.72	2.05
Exit	1.02	0.97	1.04	1.10	0.98
N	2135	533	534	534	534

Source: DB1B; Total population is the mean of the sum of population of the cities that origin and destination airport are located; Distance is the mean non-stop distance between airport pair; incumbent is the mean number of airline active in the market in year 2004; entry is the mean number of entry of all market in the group; exit is the mean number of exit of all market in the group; N is the number of market in the group; Revenue = Total ticket fare obtain from all iternary in the market; Quantity = Total number of pessenge obatined from all iternary in the market; Price = Weighted avergae price of the ticket computed by Revenue/Quantity; Yield = Median yield (Ticket fare/Mile flown) from all iternary in the market.

Table 3
Demand estimation

	OLS	**FE**	**IV**
fare	-0.037	-0.004	-4.888
	(0.003)	(0.001)	(0.990)
log(S_j/S_g)	0.895	0.995	0.354
	(0.002)	(0.000)	(0.159)
hub_orig	0.495	0.032	2.941
	(0.016)	(0.003)	(0.438)
hub_dest	0.476	0.029	1.187
	(0.016)	(0.003)	(0.699)
distance	-0.276		2.139
	(0.006)		(0.337)
N	41453	41453	41453
Res R^2			0.0009
N*R^2			1.510(2)
p-value			0.47

Note: We use the data from Q1 in year 2003 & 4. Average fare = 2.035 (100 dollars), average hub at origin = 0.703, average hub at origin = 0.700 and nonstop mile = 1.343 (1000 miles). OLS estimation uses airline, origin-year, destination-year and year dummies; Fixed effect estimation with market and year as index uses airlines dummies; IV estimation uses airline, origin-year, destination-year and year dummies. The number in the bracket is the number of identifying restrictions.

Table 4

Marginal Cost & Markup for Largest 25 Airlines except PSA

Carrier Code	Carrier	Type	Fare	MC	$\frac{\text{Fare}-\text{MC}}{\text{Fare}}$	Hub
AA	American	NC	2.045	1.856	0.137	-10.972
DL	Delta	NC	2.014	1.830	0.112	-12.343
UA	United	NC	1.989	1.804	0.131	-11.884
NW	Northwest	NC	2.006	1.821	0.113	-12.740
YV	Mesa	RC	1.932	1.753	0.182	-17.263
CO	Continental	NC	2.338	2.156	0.097	-10.896
RU	Expressjet	RC	2.134	1.955	0.102	-15.529
MQ	American Eagle	RC	2.153	1.973	0.176	-14.994
OH	Comair	RC	1.907	1.728	0.180	-18.265
EV	Atlantic Southeast	RC	2.000	1.818	0.186	-17.388
ZW	Air Wisconsin	RC	1.950	1.771	0.198	-17.517
HP	America West	RC	2.132	1.949	0.114	-11.046
US	US Airway	NC	2.181	1.996	0.252	-11.441
XJ	Mesaba	RC	1.955	1.777	0.228	-18.878
WN	Southwest	LCC	1.497	1.299	0.174	-13.902
AX	Trans States	RC	1.952	1.774	0.307	-19.377
F9	Frontier	LCC	1.850	1.671	0.120	-15.747
QX	Horizon	RC	1.713	1.534	0.299	-18.713
TZ	ATA	LCC	2.082	1.903	0.102	-13.647
F8	Freedom	LCC*	2.088	1.910	0.127	-16.780
AS	Alaska	NC	2.773	2.590	0.214	-6.752
YX	Midwest Express	LCC	2.013	1.829	0.272	-15.559
FL	AirTran Airways	LCC	1.570	1.390	0.121	-14.760
HA	Hawaii	RC*	3.169	2.984	0.158	-3.086

Note: We use the data from Q1 in year 2003 & 4. PSA is dropped because it starts providing services after 2004Q1. The average is calculated across market and year. The asterisks in the type column denote the type is assigned by the authors according to their business model.

Table 5
Estimates from Reduced Form Active Model, 2003&4

Variable	Basic	Resid	G10	G40	G80	G10	G40	G80
Constant	-6.980	-7.070						
	(0.267)	(0.269)						
log(Pop)	0.134	0.150	0.069	0.047	0.042	0.136	0.124	0.119
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.021)	(0.021)	(0.021)
log(Distance)	0.234	0.312	0.104	0.064	0.053	0.203	0.150	0.139
	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.021)	(0.021)	(0.022)
Tourist	-0.219	-0.291	-0.111	-0.052	-0.047	-0.738	-0.266	-0.198
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.138)	(0.147)	(0.151)
log(1+#Incum)	0.454	0.157	-2.171	-3.691	-3.774	-2.414	-3.771	-3.831
	(0.031)	(0.038)	(0.070)	(0.090)	(0.091)	(0.077)	(0.093)	(0.095)
Incumbent	4.623	4.637	4.741	4.812	4.824	11.467	11.004	10.943
	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.557)	(0.558)	(0.559)
log(Pop)_i						-0.259	-0.249	-0.246
						(0.035)	(0.036)	(0.036)
log(Distance)_i						-0.257	-0.235	-0.234
						(0.033)	(0.033)	(0.033)
Tourist_i						0.270	0.093	0.065
						(0.058)	(0.062)	(0.063)
log(1+#Incum)_i						-0.507	-0.127	-0.401
						(0.070)	(0.068)	(0.069)
Residual		0.163						
		(0.012)						
Year	-0.041	-0.010	0.032	0.529	0.541	0.361	0.553	0.562
	(0.021)	(0.020)	(0.022)	(0.023)	(0.024)	(0.022)	(0.024)	(0.024)
g600			0.107	4.019	4.317	-1.081	2.458	2.691
			(0.322)	(0.359)	(0.371)	(0.374)	(0.406)	(0.418)
g1200			0.974	5.195	5.459	-0.113	3.689	3.889
			(0.332)	(0.374)	(0.387)	(0.383)	(0.419)	(0.432)
g1800			1.465	5.998	6.348	0.448	4.526	4.810
			(0.339)	(0.384)	(0.397)	(0.391)	(0.430)	(0.442)
N	136671	136671	136543	136543	136543	136543	136543	136543

The first stage regression is log(mktrevenue) $= \gamma_1$logpop $+ \gamma_2$logdist $+ \varepsilon$, where mktrevenue is the total revenue of the market computed by adding up all the ticket fare during 2003Q1. It has R-sqaure at 0.19. The second stage results are similar for first order expansion and second order expansion of the first stage residual. The following variables are computed using data in Q1 (for the market without any player in Q1, we use data from Q3)

Pop = Total population of the city pair where the airport pair located. Data: US Census.

Distance = Non-stop distance of airport pair. Data: DB1B

Table 6
NPL Estimates from Active Model, 2003&4

Variable	Logit	1st NPL	20th NPL	Logit	1st NPL	20th NPL
log(1+#Incum)	1.735	1.754	1.424	2.282	2.296	2.073
	(0.064)	(0.064)	(0.041)	(0.091)	(0.091)	(0.075)
log(1+#Incum)_i				0.561	0.555	0.617
				(0.064)	(0.064)	(0.060)
Incumbent	4.720	4.721	4.382	6.070	6.072	5.818
	(0.021)	(0.021)	(0.028)	(0.156)	(0.158)	0.145
g2	5.097	5.187	4.412	6.330	6.425	5.763
	(0.133)	(0.135)	(0.086)	(0.197)	(0.200)	(0.161)
g4	6.051	6.147	5.354	7.348	7.449	6.854
	(0.153)	(0.156)	(0.104)	(0.217)	(0.220)	(0.181)
g6	6.503	6.602	5.783	7.827	7.931	7.330
	(0.162)	(0.165)	(0.109)	(0.225)	(0.228)	(0.188)
g8	6.760	6.861	6.022	8.104	8.210	7.598
	(0.169)	(0.172)	(0.114)	(0.232)	(0.235)	(0.193)
g10	7.268	7.371	6.472	8.628	8.736	8.067
	(0.177)	(0.180)	(0.116)	(0.239)	(0.242)	(0.199)
N	136543	136543	136543	136543	136543	136543

Note: The coefficients of all groups from 1 to 11 are estimated, but only the result of selected groups are reported for brevity.

DISCUSSION PAPER SERIES

IZA

IZA DP No. 1722

Bargaining Frictions and Hours Worked

Stéphane Auray
Samuel Danthine

August 2005

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Bargaining Frictions and Hours Worked

Stéphane Auray
Université Charles-de-Gaulle Lille 3,
GREMARS and CIRPÉE

Samuel Danthine
Université du Québec à Montréal,
CIRPÉE and IZA Bonn

Discussion Paper No. 1722
August 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1722
August 2005

ABSTRACT

Bargaining Frictions and Hours Worked*

A matching model with labor/leisure choice and bargaining frictions is used to explain (i) differences in GDP per hour and GDP per capita, (ii) differences in employment, (iii) differences in the proportion of part-time work across countries. The model predicts that the higher the level of rigidity in wages and hours the lower are GDP per capita, employment, part-time work and hours worked, but the higher is GDP per hours worked. In addition, it predicts that a country with a high level of rigidity in wages and hours and a high level of income taxation has higher GDP per hour and lower GDP per capita than a country with less rigidity and a lower level of taxation. This is due mostly to a lower level of employment. In contrast, a country with low levels of rigidity in hour and in wage setting but with a higher level of income taxation has a lower GDP per capita and a higher GDP per hour than the economy with low rigidity and low taxation, because while the level of employment is similar in both economies, the share of part-time work is larger.

JEL Classification: E24, J22, J30, J41, J50, J64

Keywords: models of search and matching, bargaining frictions, economic performance, labor market institutions, part-time jobs, labor market rigidities

Corresponding author:

Samuel Danthine
Université du Québec à Montréal
Département des Sciences Économiques
Case postale 8888, succursale Centre-Ville
Montréal (QC) H3C 3P8
Canada
Email: danthine.samuel@uqam.ca

* We would like to thank Mark Bils, Jean-Pierre Danthine, Jeremy Greenwood, Lance Lochner, Miguel Molico, Javier Ortega, Peter Rupert, Etienne Wasmer, participants at the WEGMaNS conference in Rochester (2003), at the macro-workshop at the University of Western Ontario (2003) and at the SED and ESEM summer meetings (2004) for helpful comments. The traditional disclaimer applies.

1 Introduction

This paper belongs to a strand of the literature that explores the impact on economic performance of labor market institutions. It goes beyond the usual Europe versus US comparison by introducing a wider range of measures of economic performance and a mix of labor market institutions and labor income taxation. Specifically, it is argued that, within a two-sided search framework, ex-ante heterogeneity on both sides of the labor market, labor/leisure choices, bargaining frictions on wages and hours worked, and labor income taxation are crucial in explaining economic performance. Table 1 provides a summary evaluation of labor market arrangements in France, the Netherlands and the US, as well as four measures of economic performance.

Table 1: Economic Performance and Labor Market Characteristics

	France	Netherlands	US
GDP per capita	77	82	100
GDP per hour	103	106	100
Employment rate (%)	62	73	72
Part time (%)	13.7	33.9	13.4
Flexibility: wages	−	+	++
Flexibility: hours	−	+	++
Labor income taxation	+	+	−

Notes: All data from the OECD data base for 2002. GDP per capita and GDP per hour are expressed relative to the US.

In terms of economic performance, the focus is on the following facts. First, GDP per capita is higher in the US than it is in Europe. Second, GDP per hour is higher in France and the Netherlands than it is in the US. Third, employment is much higher in the US and in the Netherlands than it is in France. Fourth, a large proportion of jobs in the Netherlands is part-time, which is not the case in either of the other countries.

In terms of labor market institutions and taxation, the US is undoubtedly the country with the most flexible labor market. The share of workers covered by wage bargaining is very low and the level of coordination between unions and employers is low. There is no legal maximum number of hours worked and the level of income taxation is low. The Netherlands have a more flexible labor market than France. The share of workers covered by wage bargaining is high in both countries, and a legal maximum number of hours worked is imposed by law. However, while in France there is a low level of coordination between the unions and the employers, there is a high level of coordination in the Netherlands. As argued by Nickell and van Ours (2000), this high level of coordination in the Netherlands leads to a higher degree of flexibility of the labor market. Furthermore, wage bargaining takes place at smaller intervals in the Netherlands than in France. In addition, agreements between the unions, the employers and the government in the Netherlands in the early 1980's have led to more flexibility in the choice of hours worked in that country as the union gave up their resistance to part-time jobs (see Nickell and van Ours (2000) for a discussion). Finally, labor income taxation in both France and the Netherlands is high. To summarize, the US and France represent two extremes in terms of labor market flexibility and in terms of labor income taxation. The Netherlands is an intermediate case with a flexible labor market but with a high level of taxation.

A quantitative two-sided search model with the following four characteristics is considered. First, there is ex-ante heterogeneity in both worker and firm types, and they are affected by idiosyncratic shocks. Employment in the model can be viewed as a match between a firm and a worker. Because of ex-ante heterogeneity, matches may be of different quality. This results in a situation in which high levels of employment can translate in more or less production per hour depending on the quality of sorting in the economy. In particular, an increase in the level of unemployment has two opposite effects on production. The fall in employment has a negative effect on production. The improvement in sorting due to the destruction of low quality matches has a positive one. Second, it is assumed that firms and workers may bargain over both hourly wages and hours worked. Labor/leisure choice introduces the possibility to work part-time when a pair matches. Third, the bargaining process is subject to frictions: firms and workers engaged in a match cannot renegotiate every period, but they know the probability with which they will be allowed to bargain in the future. Given the idiosyncratic shocks they face, firms and workers may want to readjust the number of hours they work and the corresponding hourly wage. This is not always possible, however, because of the bargaining frictions. These frictions thus create a distortion in both the choice to work or not to work and in the choice of the length of the working day.[1] Fourth, differences in

[1]The assumption of bargaining frictions is also useful to link the model with a recent trend in the search and matching literature. A number of recent papers have shown that wage rigidities are necessary for a search model to explain the business cycles facts (see Shimer (2004, 2005) and Hall (2003, 2005)). If wage rigidities are necessary to explain short-run fluctuations, it is preferable that it not be detrimental in explaining long-run facts.

labor income taxation are introduced. Taxes distort the value of employment for workers. For similar levels of rigidities, an increase in the labor income tax induces some workers to switch from full-time to part-time employment, others to abandon their full-time jobs, and still others to quit their part-time jobs.

The model predicts that the higher the level of rigidity in wages and hours, the lower are GDP per capita, employment, part-time work and hours worked, but the higher is GDP per hours worked. This replicates the differences between France and the Netherlands. The model also predicts that a country with a high level of rigidity in wages and hours and a high level of income taxation has a higher GDP per hour and a lower GDP per capita than a country with less rigidity and a lower level of taxation. This is due mostly to a lower level of employment and better sorting (and not to a higher degree of part-time work). The model can thus replicate the differences between France and the US. In contrast, a country with low levels of rigidity in hour and in wage setting but with a higher level of income taxation has a lower GDP per capita and a higher GDP per hour than the economy with low rigidity and low taxation. The reason is that while the level of employment is similar, part-time work is more prevalent and sorting is increased. This replicates the differences between the Netherlands and the US.

In substance, using the model to filter observations, the US is a country with a very flexible labor market and a low level of income taxation, resulting in a high level of employment. This implies that matches of high quality cohabit with matches of lesser quality (i.e. there is little sorting) in this economy. Hence, while GDP per capita is high, GDP per hour is relatively low. In contrast, France is a country with high levels of rigidity and income taxation resulting in a lower level of employment but better sorting. In terms of economic performance, GDP per capita is lower, but GDP per hour is higher, than in the US. Finally, the Netherlands are characterized by a flexible labor market and a high level of income taxation. Employment is high in that country because of the flexibility of the labor market. High taxes, however, imply that a large share of the jobs are part-time. They also force a high level of sorting, as some prospective low quality matches refuse to engage in production due to the tax distortion. In terms of economic performance, GDP per capita is low because much of the employment is part-time. GDP per hour is high, however, both because of the high share of part-time jobs and the higher degree of sorting.

The paper is organized as follows. In Section 2, data on economic performance and labor market institutions are presented for France, the Netherlands, and the US. The model is described in Section 3. The economy is parameterized, and the effects of changes in the probability of recontracting and in the rate of taxation are presented and analyzed. Finally, the relative importance of the rigidity in wages and the rigidity in hours choices to the results are presented. A final section concludes.

2 Economic Performance and Labor Market Institutions

In this section, more details about economic performance and labor market institutions for the US, the Netherlands, and France are provided.

2.1 Economic Performance

Hours worked, GDP per capita, employment and labor force participation for the period 1970 to 2000 for the US and France, for the period 1985 to 2000 for the Netherlands are traced in Figure 1.[2] Notice that, while France and the US had a similar total number of hours worked in 1970, hours have decreased steadily ever since in France while they have only decreased partially in the US. In the same time, the employment rate remained relatively stationary in France, but increased in the US. All this translates in an increase in GDP per hour in France relative to the US. Finally, labor force participation increases in both countries, but much less in France than in the US. The US has the highest level of per capita output, but France and the Netherlands are more efficient when one looks at production per hour.[3] From the mid-eighties to today, hours worked in the Netherlands follow the French trend, but employment shoots up to US levels. This results in a high level of GDP per hour coupled with a high level of employment. In addition, labor force participation is similar to participation in the US.

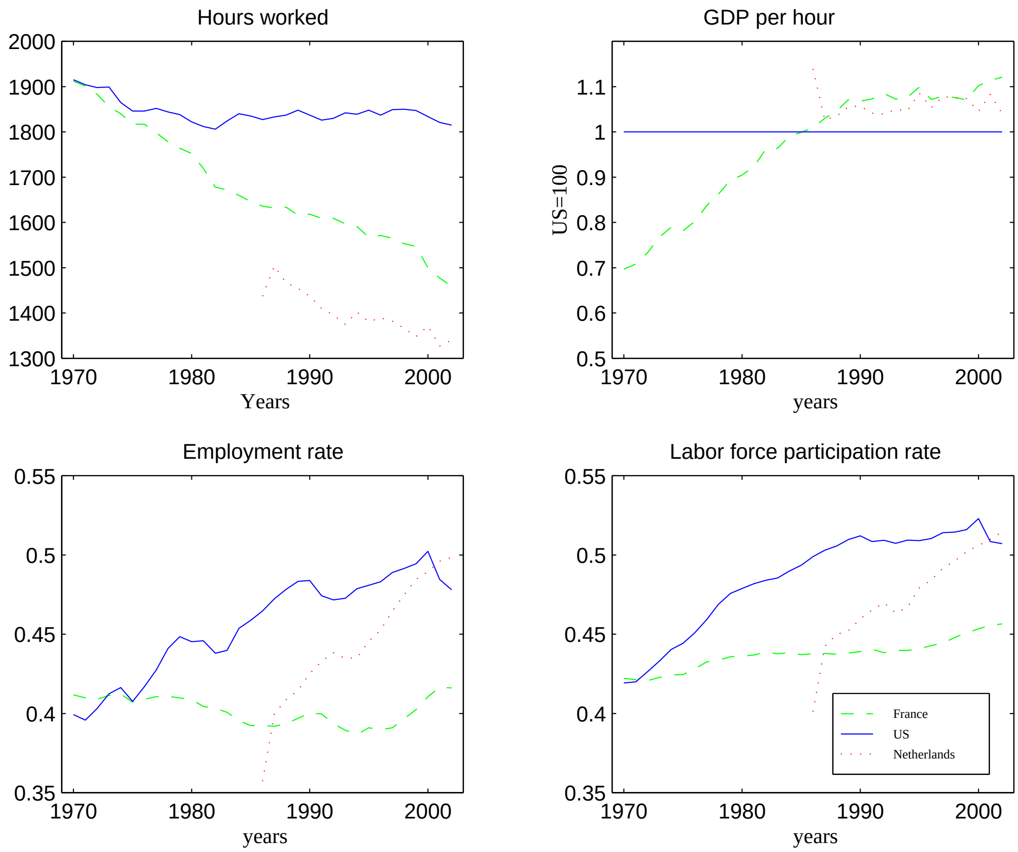
These features of the data for the Netherlands can mostly be attributed to an increased flexibility regarding part-time work in the Netherlands.[4] Data on part-time jobs as a proportion of all jobs in 2002 can be read in Table 2. In addition to the numbers for the whole population, data for three categories of age is given. The Netherlands have the highest pro-

[2] All data used here is from the OECD statistical database. Some variables are not available before 1985 for the Netherlands.

[3] GDP per hour and GDP per capita, relative to the US, for all three countries from 1950 to today is represented in Figure 2.

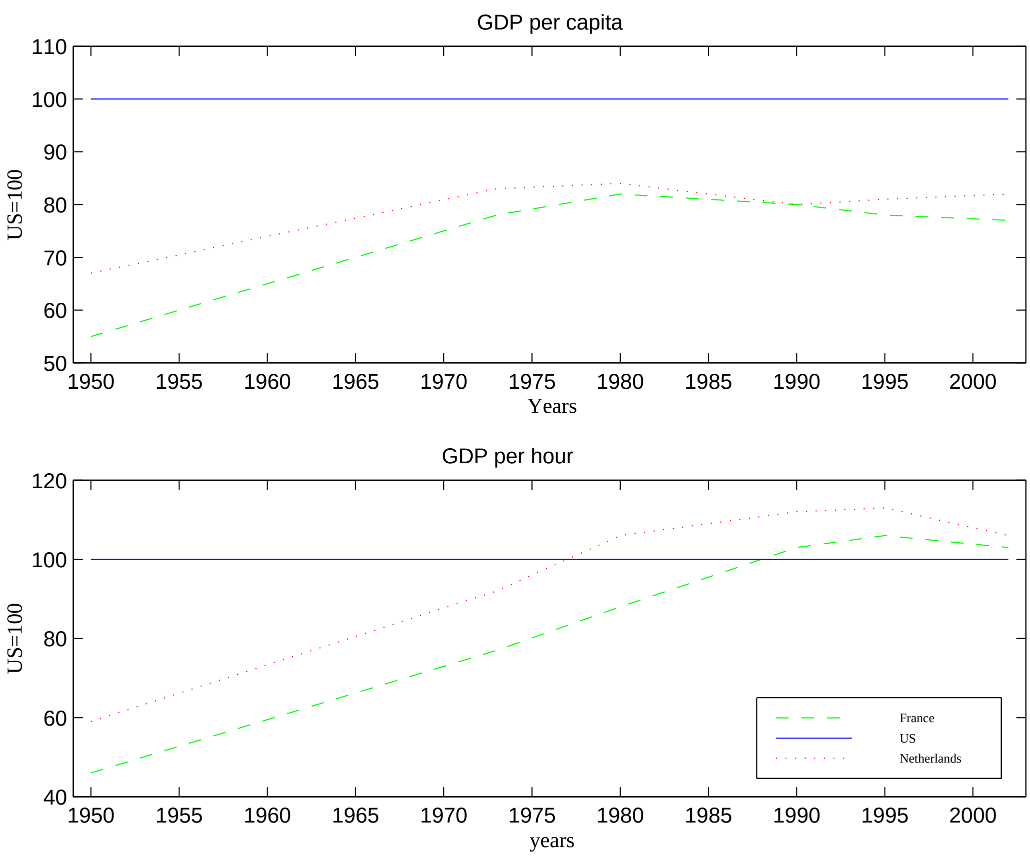
[4] Part-time jobs are defined by the OECD as jobs for which the laborer works less than 30 hours a week.

Figure 1: Economic Performance



Source: OECD Statistical database.

Figure 2: GDP per capita and per hour



Source: OECD Statistical database.

portion of part-time jobs in the whole population (33.9%).[5] In the other countries, part-time employment is much less prevalent. Another point can be made from looking across age categories. The use of part-time work is highest in the 15-24 age category. Partial work days are less present in the age category 25-54, and increases again in the population of age 55 and more.[6] Looking now across gender, one observes that the proportion of women employed part-time is higher than the proportion of men.

Table 2: Part-Time Jobs – Percent of all Jobs

	All			Age 15-24			Age 25-54			Age > 54		
Country	Men	Women	Share	Men	Women	Share	Men	Women	Share	Men	Women	Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(10)	(11)	(12)	(13)	(14)
France	5.2	24.1	79.5	9.7	26.6	67.6	4.1	23	82.7	10.6	32.3	71
Netherlands	14.7	58.8	75.4	49.3	60.7	53.5	5.9	57.1	88.3	26.8	71	56.8
USA	8.3	18.8	68.2	29.1	40	56.9	2.7	13.2	81.9	13	22.5	62.9

Notes: Data for 2002. Columns labelled 'Share' contain the share of women of total part-time work. Other columns contain the proportion of part-time work
Source: OECD Statistical database.

The evolution of the proportion of part-time jobs over time is also instructive. Over the last twenty years the Netherlands always have had the greatest proportion of part-time work in the whole population. This is mostly explained by the fact that part-time work is very prevalent for women in that country.[7] The importance of part-time work among women is true for other countries as well. Finally, except in the Netherlands, there is little change in part-time employment for the $25-54$ age group within the whole population. In the Netherlands, the proportion of part-time jobs has increased a lot for that category.[8]

2.2 Labor Market Institutions and Income Taxation

Countries differ greatly in terms of legislation on unions, wage setting, hours worked, and taxation. Some of these facts are reviewed. In particular, given that the model described below makes use of (*i*) varying average time between recontracting possibilities, (*ii*) choice of hours, (*iii*) taxation differences, and that, in addition, it is closely linked to other labor market institutions, the situation in the countries of interest is reviewed. It is argued that

[5] To a large extent, part-time work is chosen in accordance with the preferences of workers. For instance, 78 % of working part-time women in the Netherlands do not want to work full-time. See Nickell and van Ours (2000).

[6] In France and in the Netherlands, women of age 55 and more account for most of women part-time, with women in the 15-24 age category coming a close second.

[7] See Pissarides, Garibaldi, Olivetti, Petrongolo, and Wasmer (2005) for more on this topic.

[8] All this is documented in Figures 7, 8, and 9 in the Appendix.

Table 3: Labor market institutions

	France	Netherlands	United States
Union Density	9.7	23.2	12.8[a]
Wage Bargaining	**95**	80	**14**
Centralization	2	3	1
Coordination	2	**4**	1
Bargaining frequency (years)	1.5	**0.5**	No pattern
Weekly normal hour limits	**35–39**[b]	40[c]	40[d]
Maximum legal weekly hours[e]	41–47	41–47	**No limit**

Notes: Trade Union density: data from administrative sources except where stated. Data for 2000. Wage bargaining: percentage of employees covered by collective agreements as a percentage of the total number of employees. Data for 2000. Centralization and coordination: index from 1 (least centralization and coordination) to 5 (highest level of centralization and coordination). Data for 2000.
Source: OECD Statistical database (wage bargaining and union density), OECD Employment Outlook 2004. (centralization and coordination), Délégation du Sénat pour l'Union Européenne (1998) (bargaining frequencies), and McCann (2005) (restrictions on hours worked).
[a] Survey. [b] Labour Code, Decree No. 2002-1257, 2001. [c] Working Time Decree, 1995. [d] Fair Labor Standards Act, 1938. [e] Includes extra time.

the US is the country with the most flexible labor market characteristics and the lowest level of income taxation and that France is the opposite extreme. It is also shown that the Netherlands have level of income taxation similar to that in France but have implemented changes in the labor market legislation which have greatly increased the flexibility of the labor market.

2.2.1 Labor Market Settings

Table 3 displays data on union density, wage bargaining through collective agreements, indexes of centralization and coordination between unions, employers, and governments, frequencies of bargaining, and restrictions on hours worked for the same set of countries.

The US is a country characterized by the highest level of flexibility on the labor market. It has the lowest level of wage bargaining (collective bargaining coverage of 18%) among the three countries. Bargaining takes place exclusively at the firm level and with no particular pattern in terms of bargaining frequency. The situation is also very flexible regarding choices of hours worked. The normal work week in the US is similar to the one in the other two countries but there is no legal maximum number of weekly hours. In addition, evidence from weekly hour bands indicates that most people work full time in the US, and Americans tend to work long weeks (see Figure 10 in the Appendix).

France and the Netherlands have a collective bargaining coverage greater than 80%. This is true even though union density is relatively small (less than 20% in France and between 20% and 30% in the Netherlands). Wages are defined at the national level but within sectors in the Netherlands. Wage negotiation takes place within firms in France, but is sometimes framed by sectoral agreements. Negotiations take place every year and a half in France and twice a year in the Netherlands. The legal maximum number of weekly hours, which includes extra-time, is limited in both France and in the Netherlands. Data on weekly hour bands underline the fact that most people work full time in France, as is the case in the US, that Americans tend to work longer weeks, and that the population is spread out in most bands in the Netherlands (see Figure 10 in the Appendix).

A part from the frequency of negotiations, what distinguishes the Netherlands from France is the high level of centralization and of coordination between unions, employers and the government. This leads the Netherlands to have a higher degree of flexibility of the labor market. In that country, since the early 1980's, there have been important discussions between the government, the unions, and the employers which have led to a great level of coordination between all social partners. In 1982, the Wassenaar agreement marked a change in relations between Dutch unions, employers and the government. Unions agreed to more flexibility in wage setting and hours worked, and to give up resistance to part-time work. (See Nickell and van Ours (2000) for more details.) The Wassenaar agreement, as well as others that followed, have lead the unions to repeatedly accept greater flexibility in terms of choices of the working day, and to remove obstacles to part-time work. This process of improvement of flexibility is still taking place. For instance, the part-time Employment Act, passed by the lower house of the Dutch Parliament in February 2000, awards employees the right to increase or reduce their working hours.

To summarize, the US is characterized by a high level of decentralization, a low level of coordination between social partners and a relatively low level of coverage. Within Europe, one can distinguish France from the Netherlands. In France, negotiations are decentralized and not frequent, union density is small and coordination between social partners is small, but collective bargaining coverage is high. The Netherlands are characterized by a higher degree of centralization, more coordination and a high collective bargaining coverage with more frequent negotiations. The combination of these three elements greatly improves the flexibility of the Dutch labor market.

2.2.2 Labor Income Taxes

Labor income taxation is likely to influence labor/leisure decisions of households. Prescott (2003) discusses the effects of effective marginal tax rates on labor income in Germany, France, Italy, and the US. He shows that differences in tax rates account for most of the differences in labor supply in these countries (except Italy).

Effective income taxation levels are presented in Table 4. This table clearly shows that the labor income tax is much higher in France and in the Netherlands than in the US. Income taxes increase over time in all countries, and to a larger extent in the Netherlands and in France.

Table 4: Effective Tax Rates on Labor Income, 1965–1991

Countries	1965-1970	1971-1975	1976-1980	1981-1985	1986-1990	1991-1996
France	33.9	33.0	37.9	42.4	46.2	54.0
Netherlands	36.1	42.7	47.1	50.1	51.8	–
USA	20.1	23.0	26.1	28.3	28.6	27.7

Notes: Mendoza–Razin–Tesar effective tax rates updates through 1996 calculated using the method proposed in Mendoza, Razin, and Tesar (1994)

3 The Model

Ours is a quantitative two-sided search model with ex-ante heterogeneity in both worker and firm types and idiosyncratic shocks, as proposed in Danthine (2005), extended to include labor/leisure choices and bargaining frictions. Time is discrete. The economy is inhabited by heterogeneous and infinitely-lived workers and firms. A worker's productivity level is labelled by $z \in Z = \{z_1, ..., z_N\}$, while a firm's productivity is denoted by $x \in X = \{x_1, ..., x_M\}$. A worker of type z_k evolves to type z_l with transition probability $Z(l|k)$. Similarly, a firm's productivity evolves from x_i to x_j following the transition probability $X(j|i)$. When searching for a worker, a firm holding a vacancy meets a worker of type z_k with probability Ω_k. Similarly, an unemployed worker meets a firm of type x_i with probability Φ_i. A newly matched pair ik bargains over the hourly wage w_{ik} and the number of per period hours h_{ik}. If the two find a mutually agreeable arrangement, they produce using production function $F_{ik}(h_{ik})$. In that case, define the indicator function $I_{ik} = 1$. Otherwise, they lose a productive period, have to search once more next period and $I_{ik} = 0$. A previously matched pair composed of types ik, with previous contract (w, h), evolves to jl with probability $X(j|i)Z(l|k)$. With probability π, the pair can bargain over a new contract. If the two parties manage to agree on new terms, $I_{jl} = 1$ and the new contract is (w_{jl}, h_{jl}). Otherwise they lose a period, start searching again, and $I_{jl} = 0$. With probability $(1-\pi)$, they are not allowed to recontract. In that case, either they agree to remain together, allowing one to define an indicator function $J_{jl}(w, h) = 1$. If either member (or both) find that searching grants a higher value, they separate and $J_{jl}(w, h) = 0$.

3.1 Firms

A firm can be in any of three situations at the beginning of a period: matched with a worker and allowed to bargain again; matched with a worker and not allowed to bargain, in which case the worker and the firm must choose whether to remain matched at the previously set conditions or to split; vacant and in negotiation with a worker. Let V_i be the value for a firm of type i of remaining vacant and P_{ik} the value of a new contract for a firm of type i matched with a worker of type k. Finally, let $L_{ik}(w_{ik}, h_{ik})$ be the value for a firm of type i matched with a worker of type k of producing under a previous contract h_{ik}. Then,

$$\begin{aligned} P_{ik} = F_{ik}(h_{ik}) - w_{ik}h_{ik} + \beta \sum_j \sum_l X(j|i)Z(l|k)\Big[\pi\Big(I_{jl}P_{jl} + (1 - I_{jl})V_j\Big) \\ + (1-\pi)\Big(J_{jl}(w_{ik}, h_{ik})L_{jl}(w_{ik}, h_{ik}) + (1 - J_{jl}(w_{ik}, h_{ik}))V_j\Big)\Big]. \end{aligned} \tag{1}$$

Although complicated at first sight, this expression is straightforward. $F_{ik}(h_{ik}) - w_{ik}h_{ik}$ is just the net profit of the firm over the period. The pair ik then evolves to jl with probability $X(j|i)Z(l|k)$; with probability π, it can renegotiate and either decide to pursue their partnership $(I_{jl} = 1)$ or not. With probability $(1-\pi)$, the pair cannot renegotiate, and must decide whether to remain in partnership at the old contract $(J_{jl}(w_{ik}, h_{ik}) = 1)$ or not. The value of remaining vacant is simply given by

$$V_i = \beta \sum_j \sum_l X(j|i)\Omega_l\Big(I_{jl}P_{jl} + (1 - I_{jl})V_j\Big), \tag{2}$$

where $X(j|i)\Omega_l$ is the probability of evolving from type i to type j and to meet a worker of type l. Notice that a newly matched pair is always allowed to bargain. Finally,

$$L_{ik}(w,h) = F_{ik}(h) - wh + \beta \sum_j \sum_l X(j|i)Z(l|k)\Big[\pi\Big(I_{jl}P_{jl} + (1-I_{jl})V_j\Big) \\ + (1-\pi)\Big(J_{jl}(w,h)L_{jl}(w,h) + (1-J_{jl}(w,h))V_j\Big)\Big]. \quad (3)$$

The continuation part of this expression is identical to that in (1). The first part is just the net period profits given current types and past hours and wages.

3.2 Workers

A worker can be in the same three situations, and the expressions for workers' value functions are very similar to those of the firm. Denote the value of being employed at newly negotiated terms by E, the value of being employed at formerly negotiated terms by T, and the value of being unemployed by U. The value for a type k worker of being employed by a type i firm is given by

$$E_{ik} = u((1-\tau)w_{ik}h_{ik}, h_{ik}) + \beta \sum_j \sum_l X(j|i)Z(l|k)\Big[\pi\Big(I_{jl}E_{jl} + (1-I_{jl})U_l\Big) \\ + (1-\pi)\Big(J_{jl}(w_{ik},h_{ik})T_{jl}(w_{ik},h_{ik}) + (1-J_{jl}(w_{ik},h_{ik}))U_l\Big)\Big]. \quad (4)$$

It looks very much like equation (1), the difference being that workers have possibly non-linear utility $u(\cdot)$ and may be taxed at rate τ. The value of being unemployed is just

$$U_k = u((1-\tau)b + s, 0) + \beta \sum_j \sum_l Z(l|k)\Phi_j\Big(I_{jl}E_{jl} + (1-I_{jl})U_l\Big), \quad (5)$$

where b is unemployment benefits and s is home ("self") production. Finally, being employed by a type i firm but at past hours h and wage w yields

$$T_{ik}(w,h) = u((1-\tau)wh, h) + \beta \sum_j \sum_l X(j|i)Z(l|k)\Big[\pi\Big(I_{jl}E_{jl} + (1-I_{jl})U_l\Big) \\ + (1-\pi)\Big(J_{jl}(w,h)T_{jl}(w,h) + (1-J_{jl}(w,h))U_l\Big)\Big]. \quad (6)$$

3.3 Nash Bargaining

We now define two indicator functions, I and J. The first follows from the Nash Bargaining problem. A firm of type i and a worker of type k choose a wage w_{ik} and hours h_{ik} to maximize the product of their surpluses under the constraint that both surpluses must be non-negative:

$$\max_{h,w} \; [P_{ik}(w,h) - V_i] \quad \times \quad [E_{ik}(w,h) - U_k], \tag{7}$$

$$st.$$

$$P_{ik}(w,h) \geqslant V_i \quad \text{and} \quad E_{ik}(w,h) \geqslant U_k. \tag{8}$$

If a solution to this problem exists, then $I_{ik} = 1$, otherwise $I_{ik} = 0$. In similar fashion, $J_{ik}(w,h) = 1$ if, at the terms of the last negotiated contract (h,w), both the firm and the worker have a positive surplus, so that $L_{ik}(w,h) \geqslant V_i$ and $T_{ik}(w,h) \geqslant U_k$. Otherwise, if either or both prefer searching again, $J_{ik}(w,h) = 0$. With the existing distribution of workers and firms and with the newly defined indicator function, it is possible to update the distributions.

3.4 Updating the Distributions

Updating the probability of meeting a worker or a firm of a certain type involves counting. Let M^b_{ikop} be the measure of pairs of type ik who in the previous period were allowed to bargain and chose a contract (w_{op}, h_{op}).[9] Similarly, let M^n_{ikop} be the measure of pairs of type ik who did not bargain in the previous period, had a previously agreed upon contract (w_{op}, h_{op}), and remained together. Then $\sum_o \sum_p \left(M^b_{ikop} + M^n_{ikop}\right)$ is the measure of ik pairs who were matched in the previous period. Of these worker-firm pairs, a proportion π are allowed to renegotiate. In addition, there is a measure $\Phi_i \Omega_k N$ of ik pairs who meet in the market. If they can find a mutually agreeable contract (w_{ik}, h_{ik}), then they engage in production ($I_{ik} = 1$). Any pair consisting of types i and k evolves to types j and l with probability $X(j|i)Z(l|k)$. Hence, at the beginning of the next period, the measure of jl pairs who were matched with contract (w_{ik}, h_{ik}) is given by:

$$M^{b'}_{jlik} = \big[\big(\sum_o \sum_p M^b_{ikop} + M^n_{ikop}\big)\pi + \Phi_i \Omega_k N\big] I_{ik} X(j|i) Z(l|k). \tag{9}$$

In somewhat similar fashion, multiplying the measure of pairs of type ik who had contract (w_{op}, h_{op}) by $(1-\pi)$ yields the measure of ik firms who cannot renegotiate and have to decide whether or not to continue producing at the past contractual terms. If they decide it is worth to maintain their relationship, $J_{ikop} = 1$. The probability that they evolve to jl is given by $X(j|i)Z(l|k)$. Summing over all possible ik's leads to the measure of jl pairs who cannot

[9] In fact, this implies they were of type op in the previous period.

rebargain and carry over choice h from this period to the next:

$$M^{n\prime}_{jlop} = \sum_i \sum_k [M^b_{ikop} + M^n_{ikop}](1-\pi) J_{ikop} X(j|i) Z(l|k). \tag{10}$$

The probability of meeting a worker of type k is just the measure of unmatched workers of that type divided by the total number of unmatched workers. To obtain this, define A_{jl} as the measure of jl pairs who met in the previous period and did not find an agreeable contract, given that they were allowed to (re-)bargain. Similarly, define B_{jl} to be the measure of pairs jl who decided not to produce last period given that they could not renegotiate. These are given by

$$A_{jl} = \sum_i \sum_k \Big[\sum_o \sum_p \big(M^b_{ikop} + M^n_{ikop}\big)\pi + \Phi_i \Omega_k N \Big](1 - I_{ik}) X(j|i) Z(l|k), \tag{11}$$

and

$$B_{jl} = \sum_i \sum_k \Big[\sum_o \sum_p \big(M^b_{ikop} + M^n_{ikop}(1-\pi)(1 - J_{ikop})\big) \Big] X(j|i) Z(l|k). \tag{12}$$

It should be clear that the measure of unmatched workers or firms is given by the double sum

$$N' = \sum_l \sum_j \big(A_{jl} + B_{jl}\big). \tag{13}$$

Summing $A_{jl} + B_{jl}$, for each firm type, across worker types and dividing by N' yields the distribution of vacancy types. The distribution of unemployed is obtained in similar fashion. Formally,

$$\Phi'_j = \frac{\sum_l (A_{jl} + B_{jl})}{N'}, \tag{14}$$

and

$$\Omega'_l = \frac{\sum_j (A_{jl} + B_{jl})}{N'}. \tag{15}$$

3.5 Stationary Equilibrium

A stationary equilibrium is a set of value functions E, P, U, V, L, T, distributional functions $\Phi, \Omega, M^b, M^n, N$ and indicator functions I, J such that E, P, U, V, L, T satisfy equations (1)-(6), I, J are defined by (7), and the distributions are stationary.

4 Results

To evaluate the model, three steps are taken. First, functional forms are given and the parameters are chosen. Second, the properties of the numerical equilibrium and their sensitivity to parameter changes are discussed. Third, changes in income taxation coupled with changes in the probability of recontracting are introduced. This allows us to use the model to rationalize the differences in the economic performances of the United States, France, and the Netherlands documented above.

4.1 Parametrization

Functional forms for the production function, for individual preferences and for the idiosyncratic shocks must be specified. The production function is assumed to be a Cobb-Douglas with $\alpha = 0.4$ and $\mu = 0.8$, which implies diminishing returns to hours worked:

$$F_{ik}(h) = h^{\mu}(x_i^{\alpha} z_k^{1-\alpha}). \tag{16}$$

The utility function is assumed to be

$$u(c,h) = \frac{c^{(1-\sigma)}}{1-\sigma} - a\frac{h^{\nu}}{\nu}. \tag{17}$$

Preference parameters, like the technology parameters, are set following existing literature standards. The parameter of risk aversion, σ, is set to be 0.4. The parameter that fixes the level of consumption-leisure elasticity is set to $a = 2.5$ in the benchmark parametrization. Similarly, ν is set to 1 (hours enter linearly in the utility), as in Cooley and Hansen (1995).[10]

The rate of time preference is set to $\beta = 0.95$. This implies that the length of the period in the model is approximatively a year. The average time between contracts is not something readily available from the data. According to the French Senate, the average time between recontracting varies between two and eight quarters in European countries (see Délégation du Sénat pour l'Union Européenne (1998)). Counting one year for a period in the model implies that the average time is about two years and four months for $\pi = 0.3$, one year for $\pi = 0.5$, and it is about 4 months when $\pi = 0.7$. Home production is only introduced to prevent a log-utility specification from giving a highly negative utility, and it is set to a minimal level of 0.1. The preference and technology parameters used in the benchmark model

[10] The results are checked to be robust to variations in the parameters α, μ, σ, a, and ν.

are summarized in Table 5. Workers and firms have to choose one of four possible work days: $h \in \{0.25, 0.5, 0.75, 1\}$.

Table 5: Model Parameters

Parameter	Meaning	Value
β	discount factor	0.95
σ	risk aversion	0.4
ν	labor/leisure parameter	1
a	aversion to work	2.5
α	coefficient on firm type (production function)	0.4
μ	coefficient on hours (production function)	0.8
s	home production	0.1
π	recontracting probability	varies

The idiosyncratic shocks are set in the following way. Worker types are interpreted to correspond to education levels. It is assumed that there are ten different types of workers in the model. More precisely, the lower two types are labelled "less than high school", the next three are labelled "high school graduates" the next three are labelled "college grads" and finally the last two types represent postgraduates. The Markov matrix is then chosen so that (*i*) the distribution of diplomas in the model is roughly similar to that in the data, (*ii*) the movement between diplomas is not 'too' large. Ideally, a firm's type could be interpreted as its productivity. Unfortunately, data on firm productivity is hard to come by. There is some data available on the job skill requirement, in terms of education, of posted vacancies. Here, it is assumed that the type of a firm corresponds to its job skill requirement. The idiosyncratic shocks to the firms are then calibrated so that the distribution of job skill requirements among *all firms* in the model corresponds to the distribution, *among vacancies*, in the 1985 PSID (as reported by Handel (2000)).[11] The model distributions as well as those for the US (BLS 2001 (workers) and PSID 1985 (firms)) can be found in Table 6.

In the next section, the effects of changes in the probability of recontracting, as well as changes in labor taxation, are analyzed.

[11]Results are checked to be robust to changes in the parametrization of the idiosyncratic shocks.

Table 6: Distribution of Education Levels and Job Skill Requirements

	Worker		Firm	
Education	Model	USA	Model	USA
< HS	10%	10.1%	10%	13.2%
HS and some college	56%	59.6%	56%	56.3%
college	24%	20.0%	24%	23.4%
> college	10%	10.2%	10%	7.1%

Source: BLS 2001 (workers) and PSID 1985 (firms).

4.2 Contracts, Taxes and Labor Market Performance

The behavior of the model when the probability of recontracting changes, as well as when taxation varies, is examined. The results are then evaluated in light of the data discussed in Section 2.

4.2.1 Effects of Flexibility in Contracting

What happens, in this economy, when the probability of recontracting increases? Figure 3 plots GDP per capita, GDP per worker and GDP per hours worked in the benchmark economy.[12] When $\pi < 0.7$, GDP per capita increases with π. GDP per worker closely follows. Conversely, GDP per hours worked decreases. Once $\pi > 0.7$, the trend flattens out: GDP per capita and GDP per hours worked now remain relatively constant as a function of π.

Taking a look at Figure 4 helps understand what is happening. In that figure, employment and the share of part-time jobs are plotted against π.[13] As π increases, both the proportion of part-time jobs and employment increase at first but then flatten out. Until $\pi = 0.4$, however, the share of part-time jobs increases faster than employment. This implies that employment is increasing at a faster rate than the total number of hours worked. Hence, GDP per capita increases and GDP per hour decreases. This is the case until $\pi = 0.6$ at which point both employment and the share of part-time jobs become almost unaffected by further increase in flexibility. At that point ($\pi = 0.6$), GDP per capita and GDP per hour flatten out.

[12] GDP per capita and per worker are in right axis units in Figure 3, while GDP per hour is in terms of the left axis.

[13] The share of part-time jobs in all jobs is depicted against the left axis, employment against the right.

Figure 3: GDP's and recontracting probabilities

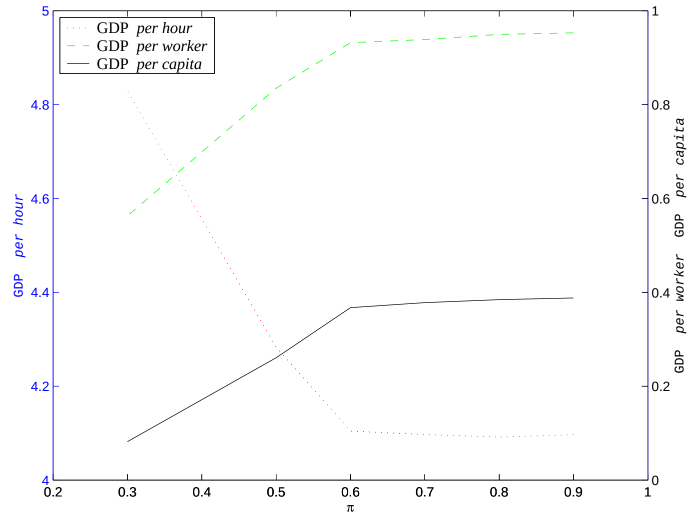
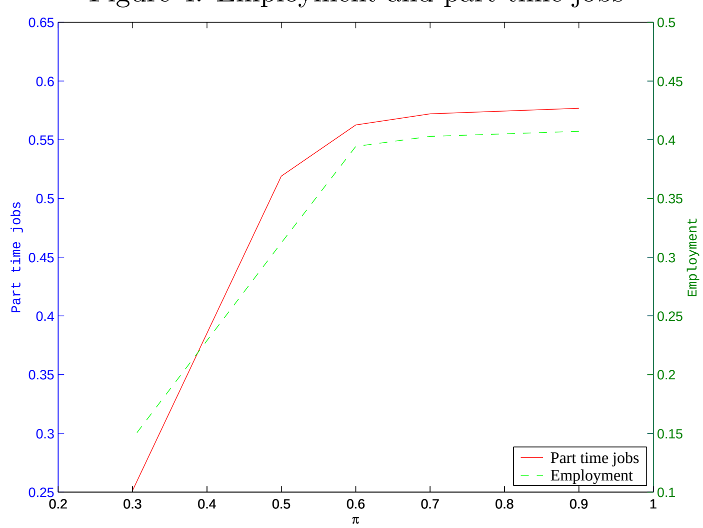


Figure 4: Employment and part-time jobs



Which pairs are affected by a change in the recontracting probability? The answer is obtained from Figure 5, where the hour choices made by worker and firm types for three different values of π – 0.3, 0.5 and 0.7– are depicted.[14] Firms are represented on the y-axis with firm type increasing from bottom to top. Workers are on the x-axis and increase in type from left to right. A black square represents a situation where the corresponding pair does not match. As the color lightens, the percentage of daily time devoted to work increases.[15] For instance, if $\pi = 0.3$, a worker of type 6 and a firm of type 7 decide to use 0.5 of a full day for production. Generally, as flexibility increases, both the white area, representing pairs engaged in full-time jobs, and the grey area, representing part-time relationships, increase, while the area in which pairs do not match (in black) decreases.

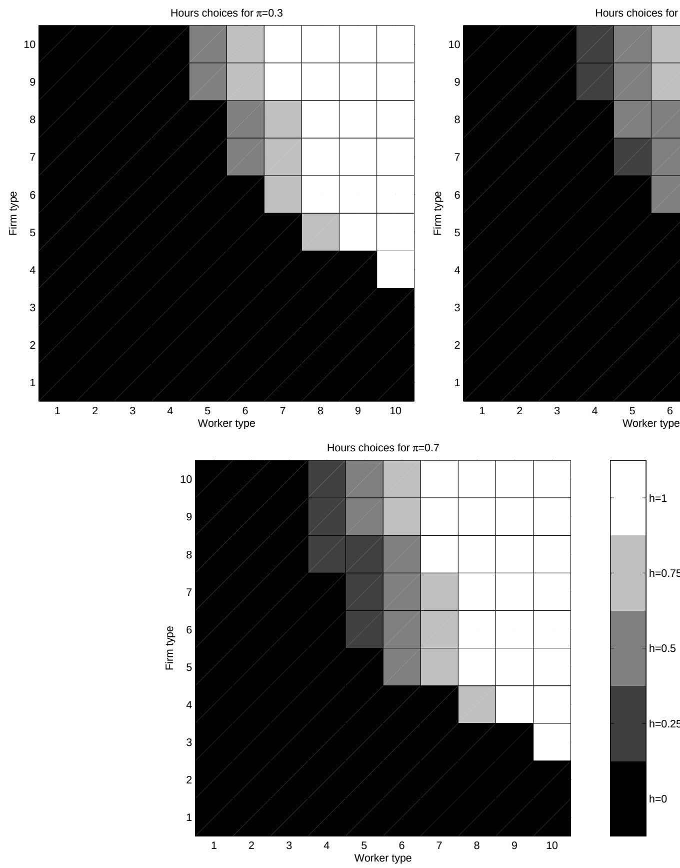
Going into more details, when π increases from 0.3 to 0.5, a number of pairs who do not match when $\pi = 0.3$ produce when $\pi = 0.5$. Workers of type 4 start working a quarter-time with firms of type 9 and 10. Workers of type 5 start working a quarter day with firms of type 7 and half-time with firms of type 8. Workers of type 6 now work half-time in firms of type 6, workers of type 7 work three-quarters of the day in firms of type 5, and workers of type 9 now work full time in firms of type 4. Moving from the intermediate case ($\pi = 0.5$) to the more flexible case ($\pi = 0.7$), workers of type 4 (respectively, 5, 6, and 8) start working with firms of type 8 (respectively, 6, 5, and 4) for a quarter of the day (respectively quarter, half, and three-quarter of the day). Workers of type 10 start working full-time with firms of type 3. In addition, workers of type 5 who would work half-time in the intermediate case now work quarter-time. Finally, workers of type 8 increase the length of their work-day in firms of type 5 from three-quarter to full time.

The intuition of what is going on is the following. In an economy with flexible wages and no leisure-labor decision, a number of pairs who cannot produce more together than individually decide to match anyway, while others do not. If the pair's joint evolution makes it likely enough to get better in the next period, and if this evolution is more likely than meeting a better partner in the future, the pair decides to lock up a match. The possibility of choosing part-time work makes it easier to lock-up a match. Once rigidities are introduced, and hours cannot be rebargained for sure in the future, locking a match with a small workday is not as profitable. This will deter some pairs from locking-up a partnership. As flexibility

[14]It is important to remember that the measure of each pair in these figures differ.

[15]h goes from 0.25 of the available time to 1.0 via 0.5 and 0.75.

Figure 5: Hour choices



increases, more and more pairs on the margin will go for part-time work. Hence employment increases and the proportion of part-time work increases, with the result that GDP per capita and the number of hours worked increase.

Are these results robust? Changes in the parameters of the production function, μ and α, do not qualitatively modify the results. As μ gets smaller, the differences between the economies with different degrees of rigidity become smaller. This is due to the diminishing marginal return of hours worked. Changes in α mostly modify the matching sets and therefore the unemployment rate. Variations in the idiosyncratic shocks have no qualitative effects on the results and small change in those shocks have no quantitatively significant effects either. The parameters of the utility function are closely linked. Increasing risk-aversion σ, for instance, while decreasing a and/or increasing ν, leaves the results qualitatively unchanged and yields very small quantitative variations. Changing one of these parameters while leaving the others constant mostly affect the level of flexibility, measured by π, at which an inflection point is observed in the GDP curves. For instance, as the parameter of risk aversion σ and the parameter a increase, the level of π at which GDP per capita starts to flatten decreases, and as ν increases, that level of π increases. If risk aversion increases a lot without changes in a or ν, it is even possible to obtain a situation in which a high level of rigidity leads to a high level of production. The reason is that working is very costly for workers, who thus wish to work only in the best firms. There is then a very low level of employment. Increasing flexibility slightly rises the employment level, and this has a positive effect on GDP per capita, but lowers the level of sorting, and this has a negative effect on GDP per capita. With a low level of employment, the second effect dominates.

4.2.2 Effect of Labor Income Taxation

As documented in Section 2, labor income taxation varies across countries. In general, the level of income taxation is much lower in the US than in Europe. It is possible to explain the different effects of increased taxation on economies with high or low rigidity. Income taxation distorts the marginal revenue of an extra hour of work. Hence, when tax increases, the workers wish to work less. Rigidity in both wage setting and hour choice decreases the long-term benefits of engaging in a part-time professional relationship. The result is that many potential part-time situations are converted in non-employment. As flexibility increases, the long-term cost associated with part-time work decreases, and the share of part-time work increases with taxation.

Table 7 displays GDP per capita, per worker and per hour, employment (E) and the proportion of part-time jobs (Part/Tot) for three levels of taxation and for three economies differing in their level of rigidity ($\pi = 0.3, 0.5, 0.7$). Call the first economy "rigid", the second "intermediate" and the third "flexible". All variables are normalized in terms of the situation in the flexible economy with no labor income taxation. This allows a comparison of the magnitude of the effects in the three economies. In all three economies, increasing taxation has the effect of decreasing GDP per capita, decreasing employment and increasing GDP per hour. The effects on the proportion of part-time jobs differ, however. Part-time jobs disappear in the economy with the highest rigidity when taxes are raised from 0 to 0.3, but the share of part-time jobs increases when taxes are raised from 0.3 to 0.5. This increase is largely due to the very low level of employment. It would be much smaller, and could even disappear, in a parametrization in which employment remains higher. The share of part-time jobs increases a lot in the economy with low rigidity. In the intermediate economy, the proportion of part-time jobs increases but less than in the most flexible case. As expected, employment decreases less with taxation when the economy is more flexible.

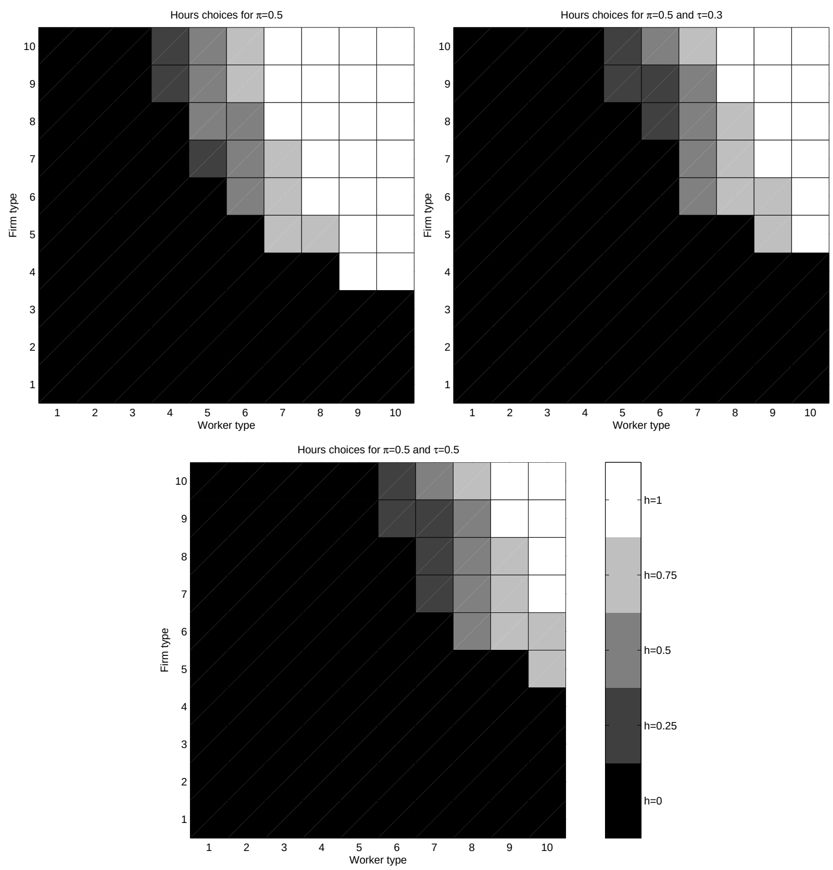
Table 7: Effect of variations in labor income tax

	$\pi = 0.3$			$\pi = 0.5$			$\pi = 0.7$		
τ	0.0	0.3	0.5	0.0	0.3	0.5	0.0	0.3	0.5
GDP pc	0.85	0.48	0.26	0.95	0.64	0.45	1.00	0.69	0.55
GDP pwk	0.58	0.16	0.04	0.83	0.36	0.19	1.00	0.47	0.33
GDP ph	1.07	1.31	1.53	1.02	1.21	1.37	1.00	1.18	1.28
E	0.68	0.32	0.17	0.87	0.56	0.42	1.00	0.67	0.60
Part/Tot	0.48	0.47	0.87	0.80	1.14	1.59	1.00	1.36	1.81

Notes: All variables are normalized by the values for parameters $\pi = 0.7$ and $\tau = 0.0$. E is employment, Part/Tot is part-time jobs divided by all jobs.

The effect of an increase in the labor income tax on pairwise hour choices in the economy with $\pi = 0.5$ is depicted in Figure 6. Clearly, many viable pairs in the economy with no taxation, the majority of them engaged in part-time contracts, are driven out by taxation. At the same time, a number of pairs who produce full-time when taxation is low reduce their work-days. It is then clear that there is a possibility that the proportion of part-time jobs increase or decrease as taxation increases, and that this depends on the total level of employment.

Figure 6: Hour choices



4.2.3 Can the Model Explain Cross-country Differences?

It is useful to summarize briefly the results described above. Starting from a situation where the probability of recontracting is low, an increase in that probability increases employment and GDP per capita, and decreases GDP per hour. When $\pi > 0.7$, these measures of GDP flatten out. In addition, as π increases, the proportion of part-time work in the economy keeps increasing. When looking at France, the Netherlands, and the US, it is striking that the Netherlands have a high proportion of part-time jobs while part-time jobs are less prevalent in France and in the US. Taking the model seriously, one expects employment and GDP per capita to be greater and GDP per hour to be lower in the former countries than in France, just as observed in the data. Without taxation, however, the model predicts that the most flexible country has the highest proportion of part-time jobs. Consequently, if one considers the US to be more flexible than France and the Netherlands, one should observe a greater proportion of part-time jobs in that country, which is clearly counterfactual. Adding taxes clears the picture. When taxation increases, the share of part-time work increases in the model in economies with high flexibility. The reverse is true in economies with low flexibility. The US has the lowest taxation rate. The Netherlands and France have high taxation rates. Hence, the model predicts that France has a lower fraction of part-time jobs, lower employment level and higher GDP per hour than the more flexible Netherlands. Finally, the model predicts that the Netherlands has a higher proportion of part-time jobs, similar employment levels and relatively high GDP per hour compared to flexible and low taxation countries like the US.

4.2.4 Further Validation

The model is not only consistent with static cross-country comparisons. The trend towards more employment and more production in the three countries under consideration can be explained by a change in the flexibility in contracting. For instance, the model can be used to understand the evolution of the measures of economic performance in the Netherlands in the last twenty years. In that country, high levels of cooperation between the government, the unions, and the employers have lead to repeated increases in labor market flexibility. Taxes, however, have remained high in that country. The model then predicts that both employment and the share of part-time jobs increase over the period. This implies that hours worked could either increase or decrease slightly, depending on whether the increase in employment dominates the increase in part-time or not. In terms of the measures of GDP, both GDP per capita and GDP per hour increase. GDP per hour should increase by a lot more, however. This is exactly what we observe in the Netherlands over the period.

Are both rigidities necessary for the results to arise in the model? To answer this question, the effects due to increased flexibility in wages from those due to increased flexibility in hours worked are separated in the next section.

4.3 Partial Rigidity in Contracts

In this section, situations in which the rigidity applies either only to wages or only to hours are discussed.

4.3.1 Fixed Hours – Flexible Wages

It is assumed first that there is a possibility of recontracting on the wage in every period, but that hours can be adjusted only with probability π. Once more, the effects of variations in the probability of recontracting and in the labor income tax are discussed. The results for $\pi = \{0.3, 0.5, 0.7\}$ can be seen in Table 8. In the economy with flexible wages but rigid hour choices, GDP per capita and per worker increase with the probability of recontracting. It is worth noting that GDP per hour also increases with π. These two features are explained by the behavior of employment and of the proportion of part-time jobs. Employment increases as the flexibility in adapting hours increases, but less and less. The proportion of part-time jobs increases when π increases from 0.3 to 0.5. It then decreases slightly. The result is that total hours worked is increasing in π, but by less than the increase in GDP per capita, which explains why GDP per hour is increasing in π.

Taxation has the same effect as in the economy of the previous section. GDP per capita is decreasing with τ while GDP per hour is increasing. Employment decreases when taxes are increased, and the proportion of part-time jobs increases. In terms of relative magnitude in the change in employment, taxes have the highest effect in the economy where $\pi = 0.3$. Part-time jobs increase relatively more in the economy with $\pi = 0.3$ than in the others when taxation moves from 0 to 0.5, but it increases more in the economy with $\pi = 0.7$ when τ increases from 0 to 0.3.

Table 8: Effects of variations in labor income tax - flexible wage

	$\pi = 0.3$			$\pi = 0.5$			$\pi = 0.7$		
τ	0.0	0.3	0.5	0.0	0.3	0.5	0.0	0.3	0.5
GDP pc	0.48	0.30	0.17	0.79	0.56	0.36	1.00	0.76	0.52
GDP pwk	0.37	0.09	0.03	0.93	0.32	0.14	1.00	0.52	0.25
GDP ph	0.61	0.98	1.27	0.71	1.03	1.23	1.00	1.25	1.45
E	0.76	0.29	0.16	0.99	0.57	0.39	1.00	0.68	0.49
Part/Tot	0.83	0.90	1.48	1.04	1.13	1.37	1.00	1.12	1.33

Notes: All variables are normalized by the values for parameters $\pi = 0.7$ and $\tau = 0.0$. E is employment, Part/Tot is part-time jobs divided by all jobs.

Notice that, as in the previous section, the model can account for differences across countries in terms of GDP per capita, employment, GDP per hour but the proportion of part-time jobs increases too much with a big increase in taxation in the most rigid economy.

4.3.2 Fixed Wages – Flexible Hours

When it is assumed that hours can be rebargained in every period, but that wage cannot necessarily be adjusted, the results are qualitatively similar to the situation in which both hours and wages are set in staggered fashion. Results pertaining to cases with $\pi = \{0.3, 0.5, 0.7\}$ can be found in Table 9.

Table 9: Effects of variations in labor income tax - flexible hours

	$\pi = 0.3$			$\pi = 0.5$			$\pi = 0.7$		
τ	0.0	0.3	0.5	0.0	0.3	0.5	0.0	0.3	0.5
GDP pc	0.94	0.54	0.32	0.99	0.71	0.54	1.00	0.83	0.65
GDP pwk	0.79	0.24	0.08	0.96	0.49	0.30	1.00	0.72	0.46
GDP ph	1.01	1.27	1.48	0.99	1.16	1.29	1.00	1.10	1.22
E	0.84	0.44	0.26	0.97	0.70	0.54	1.00	0.87	0.71
Part/Tot	0.80	1.08	1.26	0.96	1.26	1.36	1.00	1.17	1.32

Notes: All variables are normalized by the values for parameters $\pi = 0.7$ and $\tau = 0.0$. E is employment, Part/Tot is part-time jobs divided by all jobs.

As in the two previous cases discussed, GDP per capita and GDP per worker increase with flexibility, although the rate of increase is decreasing and almost nil when one moves from $\pi = 0.5$ to $\pi = 0.7$. GDP per hour is almost unchanged when π varies. Employment and the share of part-time jobs are increasing in π, but only slightly when one moves from $\pi = 0.5$ to $\pi = 0.7$. The effect of taxation is similar here than it is in the benchmark case, except for the behavior of the share of part-time jobs. Part-time jobs increase in the most rigid economy even with a small increase in taxation in this economy, while it varies very little and even decreases when τ increases from 0 to 0.3 in the benchmark case.

A special case in which there is rigidity in wages but in which the rigidity in hours does not matter is a situation in which part-time work is not an option. The results in that case are all similar to those in the benchmark case, with the exception of the effects on the proportion of part-time jobs in the economy.

Contrasting the results of the case in which only wages are rigid to those of the case where only hours are rigid is instructive. Note first that the rigidity in wages has very small effects on the various measures of GDP. The rigidity in hours has larger effects on these measures. The effects are similar in both partial-rigidity cases in terms of employment. Looking at the effects of partial rigidities on the share of part-time jobs, note that, when hours are flexible, the share increases at a decreasing rate as rigidities decrease. When wages are flexible, it increases at first and then starts decreasing. Finally, the effects of an increase in taxation are qualitatively similar in both cases, but the magnitude is smaller in the case where wages are flexible.

The interaction between the two types of rigidities is therefore necessary for the benchmark model to deliver the results described in Section 4.2.1.

5 Conclusion

Institutions explain performance. This paper shows that differences in labor market institutions and labor income taxation explain a constellation of measures of economic performance across countries. Our model economy is a two-sided matching model with ex-ante agent heterogeneity and idiosyncratic shocks in which labor/leisure choices and bargaining frictions are introduced. In such a model, a country with greater rigidity in wage setting and hour choices is stuck at a lower level of GDP per capita, lower level of employment and higher level of GDP per hour than a country with more flexibility. This arises because worker-firm pairs who would work part-time, were they given the possibility of changing the contract in the near future, are deterred from doing so by the rigidity. Hence, the proportion of part-time jobs is smaller in the economy with greater rigidity. On the other hand, the introduction of labor income taxes results in a smaller level of GDP per capita, a higher level of GDP per hour, a lower level of employment and a higher proportion of part-time jobs. The model, therefore, explains differences between the US, France, and the Netherlands.

More precisely, given that France is a country with higher wage and hour rigidities and high taxation, that the Netherlands is a country with less rigidity and high taxation, and the US has the lowest level of rigidity and the lowest level of income taxation, the model predicts that France has a low employment level, a low fraction of part-time jobs, a low GDP per capita and a high GDP per hour. It predicts that the Netherlands has a high employment level, an important fraction of part-time jobs, a low GDP per capita and a high GDP per hour, and it predicts that the US has a high employment level, a lower share of part-time jobs, a high level of GDP per capita and a low level of GDP per hour. All these features are clearly in the data.

References

Cooley, T. F., and G. D. Hansen (1995): "Money and the business cycle," in *Frontiers of business cycle research*, ed. by T. F. Cooley, chap. 7. Princeton University Press.

Danthine, S. (2005): "Two-Sided Search, Heterogeneity and Labor Market Performance," Discussion Paper 1572, IZA.

Délégation du Sénat pour l'Union Européenne (1998): "Quelles Politiques de l'Emploi pour la Zone Euro," Rapport d'Information 388.

Hall, R. E. (2003): "Modern Theory of Unemployment Fluctuations: Empirics and Policy Applications," *American Economic Review Papers and Proceedings*, 93(2), 145–150.

——— (2005): "Wage Determination and Employment Fluctuations," *American Economic Review*, 95(1), 50–65.

Handel, M. (2000): "Trends in Direct Measures of Job Skill Requirements," Working Paper 301, Jerome Levy Economics Institute.

McCann, D. (2005): *Working Time Laws: A global perspective. Findings from the ILO's Conditions of Work and Employment Database*. International Labour Organization.

Mendoza, E. G., A. Razin, and L. L. Tesar (1994): "Effective Tax Rates in Macro-economics: Cross-Country Estimates on Factor Income and Consumption," *Journal of Monetary Economics*, 34(3), 297–323.

Nickell, S., and J. van Ours (2000): "The Netherlands and the United Kingdom: a European Unemployment Miracle?," *Economic Policy*, 30, 137–180.

OECD (2004): *OECD Employment Outlook 2004*. OECD.

Pissarides, C., P. Garibaldi, C. Olivetti, B. Petrongolo, and E. Wasmer (2005): "Women in the labor force: How well is Europe doing?," in *European Women at Work*, ed. by T. Boeri, D. del Boca, and C. Pissarides, pp. 1–56. Oxford University Press.

Prescott, E. C. (2003): "Why Do Americans Work So Much More Than Europeans?," Staff Report 321, Federal Reserve Bank of Minneapolis.

Shimer, R. (2004): "The Consequences of Rigid Wages in Search Models," *Journal of the European Economic Association Papers and Proceedings*, 2, 469–479.

——— (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies: Evidence and Theory," *American Economic Review*, 95(1), 25–49.

Appendix

Figure 7: Part-time jobs – whole population

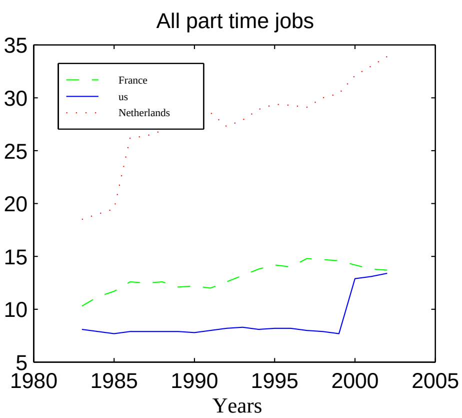
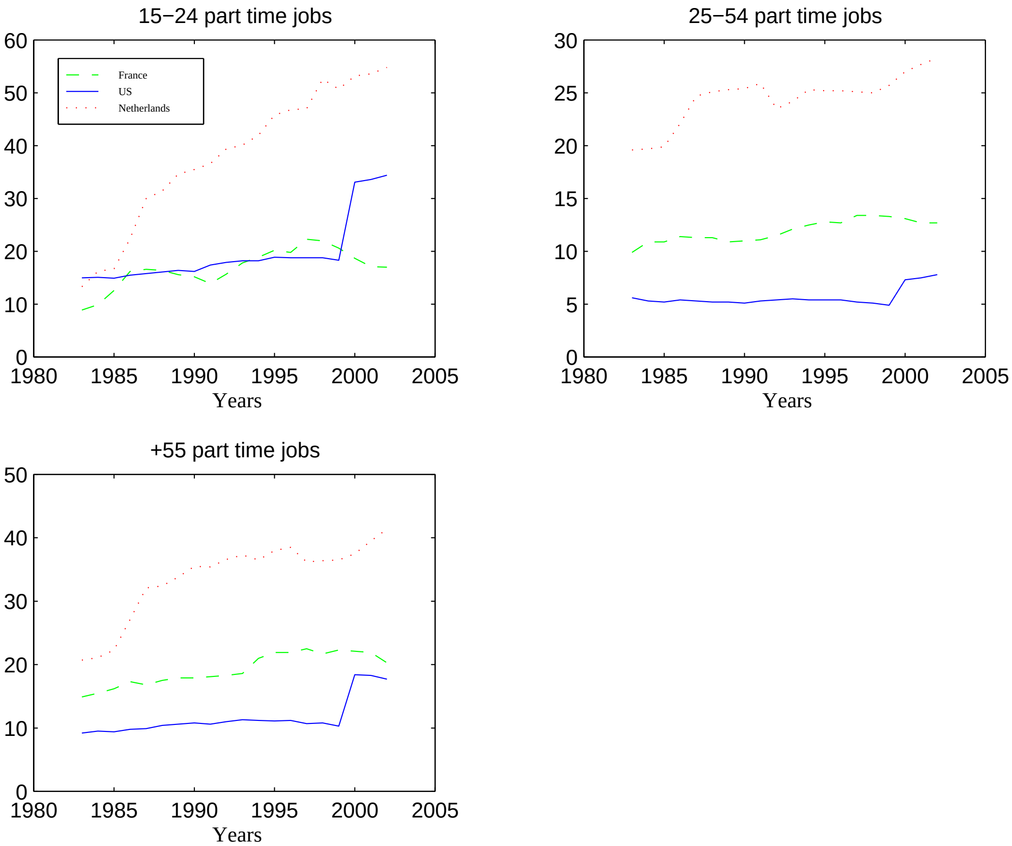


Figure 8: Part-time jobs – women

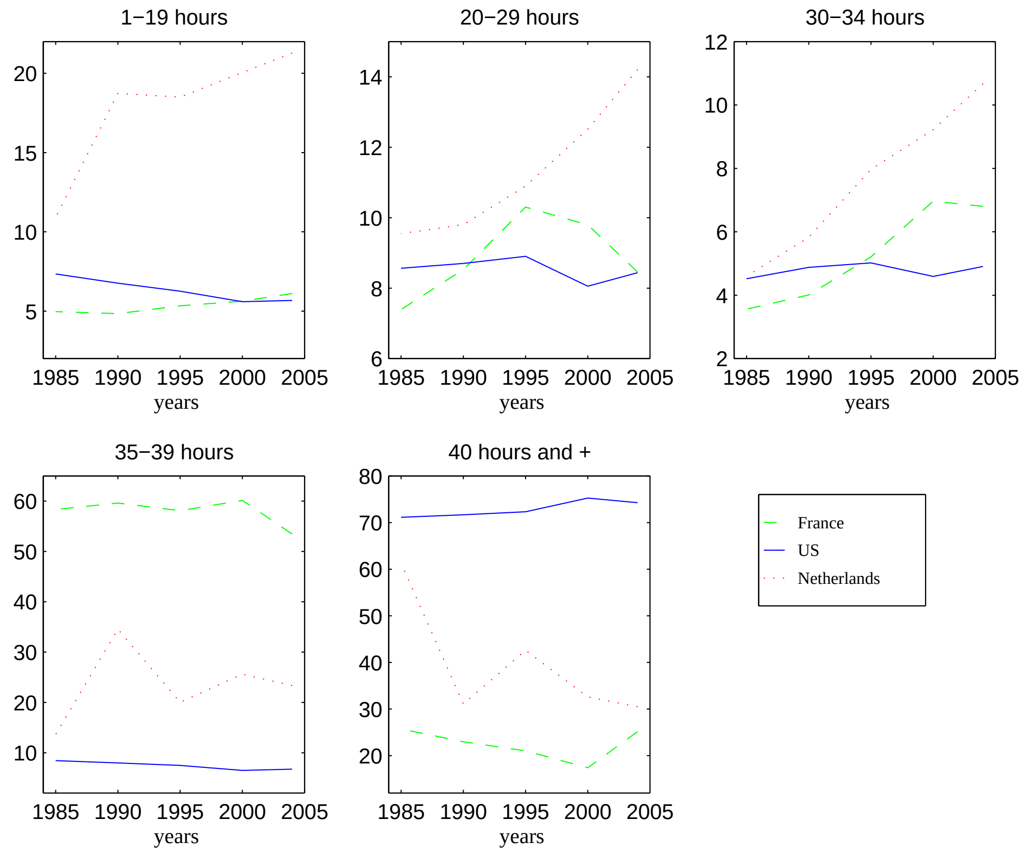
Source: OECD Statistical database.

Figure 9: Part-time jobs – differences across age groups



Source: OECD Statistical database.

Figure 10: Weekly hours band 1985–2004 (%)



Source: OECD Statistical database.

Women's College Choice: How Much Does Marriage Matter?

(Job Market Paper)

Suqin Ge*

University of Minnesota

and Federal Reserve Bank of Minneapolis

Revised January 2006

Abstract

This paper investigates the sequential college attendance decision of high school females and quantifies the impact of marriage on women's college choice. A dynamic choice model of school attendance, labor supply, and marriage is formulated and structurally estimated using panel data from the National Longitudinal Survey of Youth 1979 (NLSY79). Results show that when I zero out the benefits from marriage in the estimated model, the predicted college graduation rate drops by 6 percentage points, from 38% to 32%. Based on the estimated model, changes in family income, parental education, and cognitive ability predict an increase in enrollment of about 8 percentage points between the early 1980's to the early 2000's. Improvements in potential husbands' schooling predict an additional 4 percentage points increase. The dramatic increase in female's college premium accounts for only a 3 percentage points increase in enrollment, yet it accounts for a 10 percentage points increase in college graduation.

JEL classification: J12, J22, I21

Keywords: college choice, marriage, assortative mating, NLSY, women

*I am greatly indebted to Zvi Eckstein for his guidance and constant encouragement. I also thank Thomas Holmes, Sam Kortum, Larry Jones, Hanming Fang, participants in applied microeconomics workshop at University of Minnesota, and seminar participants at Federal Reserve Bank of Minneapolis for valuable comments and discussions. Financial support from Heller Dissertation Fellowship is gratefully acknowledged. All errors are mine.

1 Introduction

The primary motivation for going to college considered in the existing empirical literature is the increase in earnings power that college education provides (Willis and Rosen 1979, Keane and Wolpin 2001). The literature has ignored another potentially important benefit of college: college improves marriage opportunities by providing a social venue to meet potential spouses. Furthermore, a college-educated individual is substantially more likely to have a college-educated spouse. Thus, the individual enjoys educational balance in the household and benefits from the earnings power of the spouse. While this "marriage benefit" of college surely applies to both sexes, it is likely to be particularly important for women since married men on average have higher labor force participation rates and higher incomes than married women.[1]

This paper examines the choice of women on whether to attend and complete college, taking into account not only the labor market effects but also the marriage market effects. In addition, the effects of direct and opportunity costs of college as well as the effects of individual ability and family background are jointly considered.[2] To empirically disentangle all these effects, a dynamic choice model is formulated and estimated in which women decide whether to enroll and how long to stay in college, whether and when to work, when and to whom to marry. The novelty of this paper is in providing a quantitative assessment as to what extent women's college decision is determined by expectations of future marriage.

In the model, the college decision and marriage is linked in the following way. First, it is assumed that meeting technology is such that women who go to college get more marriage offers. For a college woman, any given offer is not necessarily more likely to be from a college man. But since she gets more offers, over any time interval she is more likely to have more offers of college educated men. Second, it is assumed that there exists a disutility from the educational imbalance in the household. When one spouse has a college degree and the other does not, it detracts from

[1]The "Marriage effect" on schooling choice, in general, seems to be larger for young women. When a sample of NLSY79 youths were asked the reason why they left school without a degree, approximately 25% of women chose the response categories "getting married, pregnancy, or home responsibilities" as the main reason as compared to only 5% of their male counterparts who chose the same categories.

[2]Cameron and Heckman (1998, 2001) emphasize effects of family background on college attendance; Cameron and Taber (2004) test for the importance of the credit market based on different impacts of direct schooling costs and opportunity costs if borrowing constraints were operative; Keane and Wolpin (2001) and Eckstein and Wolpin (1999) allow for joint decisions on schooling and labor supply.

marital bliss. Thus, highly educated women marry highly educated men. This phenomenon is known as educational assortative mating (Becker 1973) and is well documented.[3]

There are a number of difficulties with assessing the impact of marriage on college choice. The first difficulty is due to the dynamic simultaneity of college attendance, labor force participation, and marriage decisions. The dynamics of the decision process are due to the dependence of current choices on previous choices. For instance, whether or not one will complete the senior year of college depends crucially on if the individual finishes junior year. An example of simultaneous decisions is when a good job offer or marriage proposal comes along, it is likely to induce a woman to drop out of college. Without understanding the dynamic process by which individuals determine college enrollment and graduation, it is impossible to quantify the effects of alternative determinants, including expectations on future earnings and marriage.

The second difficulty is due to the endogenous self-selection of college, employment, and marriage decisions. The earnings gain from attending and completing college, known as the college premium, increases in individual skills or abilities, and those who have highest skills are the most likely to attend college. A statistical analysis could then attribute the effect of skills on college attendance to college premium. Individual skills are also likely to be correlated with preference for marriage. If it is the case that women who value marriage more have low skills systematically and are therefore less likely to attend college, then the estimated effect of marriage on college attendance would be downwardly biased. Self-selection is controlled by allowing for unobserved types in skills and in marriage,[4] and letting skills be correlated with background characteristics such as family income, parental education, and individual cognitive ability. Hence, the structural model implements a correction for selection biases which is based on an explicit dynamic decision process.

[3]For married women with college degrees, 60 percent of their husbands are also college graduates. On the other hand only 6 percent of married women with high school degrees marry to college men, according to a sample of high school young women from NLSY79. Similar pattern holds for different samples from census data. See Mare (1991) for trends in educational assortative mating from the 1930s to the 1980s. Pencavel (1998) studies the interaction between educational assortative mating and married couple's labor supply.

[4]Modelling skill as multidimensional is pioneered by Willis and Rosen (1979) and Heckman and Sedlacek (1985), formally incorporating Roy (1951)'s self-selection model. More recently, Keane and Wolpin (1997, 2001), Eckstein and Wolpin (1999) integrated ability selection in a dynamic setting of employment and schooling choices. Both unobserved skill and marriage types are used in broad sense in the paper. For example, skill types may differ in motivation, perseverance and tastes for school and marriage types may vary in attractiveness and preference for marriage.

The model is estimated by using a sample of high school white females from the National Longitudinal Survey of Youth 1979 (NLSY79). NLSY79 is a panel which provides dates for the beginning and ending of college, employment, and marriage. It also provides detailed information on an individual's background, wages if employed, and spouse's years of schooling and income if married. To empirically implement the model, it is first solved numerically. Then choice proportions, transitions, and wages over 10 years as well as the joint schooling distribution of married spouses are simulated and the method of moments is used to estimate the parameters. Empirical identification is secured from the conditional transitions from all of the chosen states to new states for 10 years.[5] For example, a college graduate woman's transition from employment to nonemployment following marriage would rationalize the marriage incentive for college attendance besides the earning incentive. Simulations using the estimated parameters show that the structural model fits the dynamics of college enrollment, dropout, and graduation, the transition from school to work, and the transition from school to marriage, etc.

To assess the importance of marriage on college attendance, a counterfactual economy is considered in which benefits from marriage are ruled out. Therefore, homogeneity in educational background between husband and wife does not increase the utility from marriage and college attendance does not increase the marriage offer rate. The equilibrium choices are numerically simulated in such a hypothetical world and a comparison is made of predicted college enrollment and graduation with the actual economy. In the real economy, the college enrollment rate is 61% for high school females and graduation rate is 38%. With no benefits from marriage, the college enrollment rate would drop to 58% and the graduation rate would drop to 32%.

The estimation of the model is based on a NLSY79 sample who were graduating from high school in the early 1980's. Between the early 1980's and the early 2000's, college enrollment rates increased from 61% to 80%.[6] I use the model which is estimated from the NLSY79 sample to account for this increase and have the following findings. Improved background on family income, parental education, and individual cognitive ability of a sample from the National Longitudinal

[5]Each year individuals choose from eight alternatives, so 64 transitions are observed from one year to another over a period of ten years.

[6]These are enrollment rates of white females with high school diploma based on the NLSY79 and the NLSY97 samples. The enrollment rates from the National Center for Education Statistics (NCES) and CPS are lower since their high school graduates include individuals who complete a GED (General Equivalency Diploma). It is well known that GED is not equivalent to high school diploma (Cameron and Heckman 1993).

Survey of Youth 1997 (NLSY97) would predict enrollment to increase by 8 percentage points. Changes in potential husbands' schooling imply an additional 4 percentage points increase in enrollment rate. The dramatic increase in female's college premium would increase college enrollment by 3 percentage points, yet it can account for a 10 percentage points increase in college graduation. On the other hand, the increase in college tuition would predict only 1 percentage point drop in enrollment. Overall the estimated model does well in predicting college enrollment behavior in the early 2000's, which is consistent with the stability of the structural model.[7]

This paper is organized as follows. In Section II, a dynamic discrete choice model is constructed which is designed to capture the interaction among women's schooling, labor supply, and marital choices. Section III describes the NLSY data from which the model is estimated and presents descriptive statistics. Section IV discusses the estimation method and identification issues. Estimation results are given in Section V. Section VI provides counterfactual simulations. Section VII presents the conclusion.

2 The Model

2.1 The Basic Structure

Choices about college attendance, employment, and marriage are made within social institutions. I first specify these social environment including the college, the labor market, and the marriage market, and describe how each choice is made.

The College and the Attendance Choice Consider a young woman who finishes high school and decides whether to enroll in college. She may attend college right after graduation or she may go later. If she enrolls, she must pay an annual cost of tuition and room and board $c_S = cs$ to accumulate one year of college. In the next year, she makes a decision on whether to continue studying. This decision is conditional on her previous schooling choices. When the woman is in college, she has the option to work, and/or get married at the same time. But employment and marriage in college may affect negatively the value of schooling due to time constraint. She may

[7]As discussed in Wolpin (1996), a major advantage of structural estimation is that it is capable of performing counterfactual policy experiments that entail extrapolations outside of the current policy regime. The out of sample prediction indicates that the structural model is likely stable across cohorts.

drop out of college altogether when she takes a job and/or marries a man. I assume that college degree is completed in four years from grade 13 to grade 16. Consider graduate school is going to grade 17, when the woman attends grade 17 or beyond, she pays an extra cost cg. Thus the cost of graduate school is $c_S = cs + cg$ for $S > 16$.

The Labor Market and the Employment Choice The woman can work independent of her attendance and marital status. She receives job offers at rate $p_{E_t}^{h_{t-1}}$, which depends on her schooling level $E_t \in \{hg, sc, cg\}$ and previous labor market attachment $h_{t-1} \in \{0, 1\}$. Wage offer varies if she works in college or she works during post-college time. Hourly wage offer w_t for employment in college is assumed to be log normal such that $\ln w_t = \beta_{0c} + \epsilon_{wct}$, where $\epsilon_{wct} \sim N(0, \sigma_{wc}^2)$ is an idiosyncratic shock. Hourly wage offer for post-college employment is assumed to depend on prior education and work experience as measured by cumulative years of schooling S_t, whether a college degree is received and cumulative years of experience H_t and on an idiosyncratic shock. The wage function thus follows Mincer's (1974) formulation:

$$\ln w_t = \beta_0 + \beta_1 S_t + \beta_2 H_t + \beta_3 H_t^2 + \beta_4 I(S_t \geq 16) + \epsilon_{wt},$$

where $I(\cdot)$ is an indicator function which equals one if the individual has a college degree and $\epsilon_{wt} \sim N(0, \sigma_w^2)$. The constant term β_0 can be interpreted as a composite of skill rental price and the level of individual (nonschool) premarket skill. β_1 and β_2 measure the effect of school attainment and work experience on the wage. β_4 is the wage premium due to college graduation. I allow for measurement error in observed wages, such that $\ln w^o = \ln w + u$, where w^o is the observed wage, w is the true wage and the error term is normally distributed: $u \sim N\left(0, \sigma_u^2\right)$.

The Marriage Market and the Marriage Choice Every year the woman receives a marriage proposal with some probability depending on her age, schooling level, and previous marital status. The marriage offer arrival rate when she is single has the following logistic form:

$$\Pr_t = \frac{\exp(b_0 + b_1 age_t + b_2 age_t^2 + b_3 I(S_t > 12))}{1 + \exp(b_0 + b_1 age_t + b_2 age_t^2 + b_3 I(S_t > 12))}.$$

A woman with college education should fare better on the marriage market. I model this in a simple way. A college educated woman gets a higher rate of offers so we expect b_3 to be positive. Marriage offers are not homogenous. They are made by men with different years of schooling S_t^H. The schooling distribution of potential husbands is assumed to be exogenous and discrete. This distribution remains the same independent of woman's characteristics. But my specification implies that over any given time interval, a college woman will get more offers of any kind, including offers from college men. Let the proportion of type g potential husband be μ_g, and the number of husband's types be G, so $\sum_{g=1}^{G} \mu_g = 1$. Then the probability of receiving a marriage proposal from a type g man is $\mu_g \Pr_t$. With probability $1 - \Pr_t$, no offer is received. If the woman is married, she always has the option to stay married. If she chooses to have a divorce, she will receive a random offer as a single woman next period.

Marriage decision is based on the woman's evaluation of marriage.[8] I use a function M to specify all emotional, biological and economic values for her related to marriage. This marriage value M at time t is assumed to depend on her own (S_t) and her husband's schooling $\left(S_t^H\right)$, on her age (age_t), on whether they have children (f_t), and on marriage duration $(mdur_t)$.

$$M_t = a_0 + a_1 \Delta S_t^2 + a_2 age_t + a_3 f_t + a_4 mdur_t,$$

where $\Delta S_t = S_t - S_t^H$ is the difference between spouses' years of schooling, f_t equals one if at least one child is in the household and zero otherwise. a_0 can be interpreted as permanent preference for marriage. A negative a_1 is consistent with positive assortative mating in education.[9] That is, educational imbalance in the household causes disutility, which could be due to disagreement on the consumption of public goods, etc. a_2 reflects the woman's varying preference for a stable relationship over time. a_3 and a_4 measure the impact of children and previous marriage choices. Children are likely to increase marriage utility. The dependence of marriage value on the duration of marriage reflects a possible increase in the bond between spouses. Value of marriage varies as

[8] For simplicity, I do not model marriage as a match outcome. Strategic behavior within the household is also not considered explicitly.

[9] I use a simple way to model educational assortative mating. In Becker (1973), mating is positive assortative if schooling levels are complements in production. Shimer and Smith (2000) derives more complex sufficient conditions for assortative mating under search costs. Wong (2003) specifies the production function as the product of the types (e.g. education) in her empirical study of marriage matching.

the marriage evolves. a new valuation of marriage could lead to a divorce.[10]

If the woman accepts a marriage offer from a man, part of the man's income is available for her consumption. The net transfer of income by the man to the woman depends on her work decision. We would expect the transfer to be smaller when the woman works. The model focuses primarily on female's decision process and assumes that married men always work full time in the labor market.[11] The earnings of (potential) husband is specified as

$$\ln y_t^H = \rho_0 + \rho_1 S_t^H + \rho_2 EX_t^H + \rho_3 EX_t^{H2} + \epsilon_{y^H t},$$

where S_t^H is his years of schooling, EX_t^H is potential experience,[12] and ϵ_{y^H} is the productivity shock. I also allow for measurement error in observed husband's income. When the woman is single and receives an offer from a man, she observes only his schooling thus mean incomes, and she knows the distribution of ϵ_{y^H} and uses it to predict his future income. While if she is married, she observes the husband's true income, that is, she knows both S_t^H and $\epsilon_{y^H t}$.

Choice Set At the beginning of each school year if the woman has a job offer and a marriage proposal, she chooses whether to attend school, whether to work in the labor market and whether to get or stay married. The choice set for her thus consists of eight mutually exclusive and exhaustive alternatives. Let s_t, h_t, m_t be indicators for school attendance, employment, and marital status respectively, each alternative will be a triple $(s_t, h_t, m_t) \in J = \{(s_t, h_t, m_t) : s_t \in \{0,1\}, h_t \in \{0,1\}, m_t \in \{0,1\}\}$, i.e., not attend school, not work, and single $(0,0,0)$, or attend school, not work, and single $(1,0,0)$, or not attend school, work, and single $(0,1,0)$, or attend school, work, and single $(1,1,0)$, not attend school, not work, and married $(0,0,1)$, or attend school, not work, and married $(1,0,1)$, or not attend school, work, and married $(0,1,1)$, or attend school, work, and married $(1,1,1)$. If she receives no job offer then she chooses among only four alternatives: $\{(s_t, 0, m_t) : s_t \in \{0,1\}, m_t \in \{0,1\}\}$ and if she has no marriage offer, she also

[10]There are two general causes for divorce. First, search is costly and meeting occurs randomly. Second, match quality is uncertain. Dissolving a marriage may be costly. (Weiss 1997, Weiss and Willis 1997). This issue deserves separate study and I leave it for future research. Here I assume the direct divorce cost is zero, what a woman gives up is the value of marriage when she has a divorce.

[11]As argued by Van Der Klaauw (1996), given that 95% of male population works in a representative sample, this is not a very restrictive assumption.

[12]In the computation, I use married women's age as a proxy for husbands' age and potential experience. This reduces one dimension in the state space and it does not change the main results.

chooses only among four options: $\{(s_t, h_t, 0) : s_t \in \{0,1\}, h_t \in \{0,1\}\}$. If neither job or marriage offers are received, her choice set is reduced to two alternatives: $\{(s_t, 0, 0) : s_t \in \{0,1\}\}$.

The Arrival of Children In general, both the number and ages of children may be important in determining female's choices. However, I assume that the fertility effect can be adequately captured by a single indicator of the presence of any children f_t. The stochastic process that governs f_t over time is characterized by the specification of the exogenous probability of a first birth at t.[13] I specify this process as following logit form:

$$\begin{aligned}
\Pr(f_t &= 1|f_{t-1} = 0) = \frac{\exp\{c_0 + c_1 S_t + c_2 m_{t-1} + c_3 age_t + c_4 age_t^2 + c_5 mdur_t\}}{1 + \exp\{c_0 + c_1 S_t + c_2 m_{t-1} + c_3 age_t + c_4 age_t^2 + c_5 mdur_t\}}, \\
\Pr(f_t &= 1|f_{t-1} = 1) = 1.
\end{aligned}$$

The annual rate for the first birth depends on the female's education, her marital status in the previous period, her age and the marriage duration. Note that the fertility rate is not necessarily zero for single women. A single mother is observed if this woman gives birth to a child before marriage or she is the custody parent after a divorce.

Preferences and Constraints The woman has preferences over choice variables, i.e. consumption c_t, school attendance s_t, labor force participation h_t, and marital status m_t, conditional on the state space Ω_t, which is specified later. The utility per period at time t is given by $U_t(c_t, s_t, h_t, m_t|\Omega_t)$.

Let U_t^{shm} be the utility associated with choice (s, h, m) at period t, which is known to the individual at time t but is random from the perspective of periods prior to t. U_t^{shm} is given by:

$$\begin{aligned}
U_t^{shm} &= (\alpha_1 + \alpha_2 s_t + \alpha_3 h_t + \alpha_4 m_t)c_t \\
&\quad + v_1 s_t(1-h_t)(1-m_t) + v_2 s_t h_t (1-m_t) + v_3 s_t (1-h_t) m_t + v_4 s_t h_t m_t \\
&\quad + v_5(1-h_t) f_t + v_6 (1-h_t)(1-f_t) + M_t m_t + \epsilon_t^{shm}.
\end{aligned}$$

[13]In this model fertility is exogenous. It is clear that a more complete model should explicitly incorporate fertility decision as choice variable. However, to avoid the modeling and estimation complications resulting from an increase in the choice set and the dimension of the state space, the focus here will be on the interaction of schooling, employment and marriage decisions conditional on fertility in each period.

The utility function is assumed to be linear in consumption. The marginal utility of consumption depends on current school, work and marital status of the individual. v_1 to v_4 evaluates the net utility of attending school given employment status h_t and marital status m_t. The utility of school interacts with labor supply since more involvement in the market work may prevent individuals from engaging in school activities, representing the time constraint. It also depends on marital status if marriage requires leaving school or simply school utility is lower if married. The value of nonemployment is assumed to depend on children as represented by v_5 and v_6. M_t is the utility value of marriage as previously specified. Finally, ϵ_t^{shm}'s are alternative-specific random components representing random variations in the individual's preference for school and work, as well as changes in the utility derived from getting married or being married.

The choice decision is subject to the female's budget constraint given by:

$$c_t + c_S \cdot s_t + cc \cdot f_t = y_t h_t + \psi(h_t) y_t^H m_t.$$

c_S is the direct cost of schooling, $c_S = cs$ for $12 < S \leq 16$ and $c_S = cs + cg$ for $S > 16$. cc is the total cost related to having children in the household. y_t denotes the annual earnings of the female. y_t^H is the husband's income and $\psi(h_t)$ represents the fraction of his income that is available for the woman's consumption, which depends on her employment status. This transfer may be interpreted as the woman's share of the accumulated common property. In this specification, there is no borrowing and saving decisions. The budget constraint is assumed to be satisfied period by period.[14]

Optimization Problem The objective of the female is to maximize the expected present discounted value of utility over a finite horizon from the first year after high school graduation to a known terminal time T, i.e.,

$$\max_{\{c_t, s_t, h_t, m_t\}} E\left[\sum_{t=1}^{T} \beta^{t-1} U_t(c_t, s_t, h_t, m_t | \Omega_t)\right],$$

[14]The introduction of savings and borrowing decisions in a model like this is not straightforward, and will generally lead to a considerable expansion of the choice set and the state space. Keane and Wolpin (2001) used a model with borrowing and lending to study the effect of parental transfers on educational attainment.

where $\beta > 0$ is the woman's subjective discount factor and Ω_t is the state space at time t. The state space consists of all factors, known to the female, that affect current utilities or the probability distribution of any of the future utilities. As the model is specified, the state variables include years of schooling, years of working experience, marriage duration, all previous choices, age, fertility, and the contemporaneous shocks, the ϵ_t's. The random shocks $\epsilon_t = \{\epsilon_t^1, \cdots, \epsilon_t^8, \epsilon_{wt}, \epsilon_{wct}, \epsilon_{y^H t}\}$ are jointly serially independent, noncorrelated and have a joint normal distribution $F(\epsilon_t)$. They are known to the female in period t, but unknown before t. Choice of the optimal sequence of control variables $\{c_t, s_t, h_t, m_t\}$ for $t = 1, \cdots, T$ maximizes the expected present value given current realization of the state space.

2.2 Heterogeneity

Initial Conditions and Heterogeneity The basic model I consider above corresponds to the decision problem of a representative female. However young women differ in many aspects at high school graduation. They may differ in family background as measured by parental schooling, number of siblings, family income etc. They may differ in cognitive background as measured by *AFQT* test scores. They may also have different high school grades and *SAT* scores. The abilities and preferences of individuals are likely to vary, too, in unobserved ways (like motivation, perseverance or ambition) that are both persistent and correlated with observed traits (like test scores). All these characteristics, both observed and unobserved, may affect youth's college decisions. For example, those with greater endowments of unobserved skills may be more likely to attend college and postpone marriage and workforce entry. They may also have better family background and higher test scores. To consistently estimate the parameters, the model takes into account both the unobserved heterogeneity and its correlation with observed background.[15]

Assume that there exist $k = 1, 2, \cdots, K$ different skill types (Heckman and Singer 1984).[16] The *ex ante* probability that a female i is of type k is denoted by P_i^k. To capture the correlation between a woman's unobservable type and her background, I allow P_i^k to depend on her observed

[15] There are different ways to introduce observed and unobserved heterogeneity into a schooling model. Cameron and Heckman (2001) uses background variables as explanatory variables in their econometric model. Eckstein and Wolpin (1999) treats heterogeneity as unobservable and finds that unobserved types are correlated with observed background variables. Keane and Wolpin (2001) includes the joint distribution of unobserved type and some observables in the likelihood function. Cameron and Taber (2004) adopts all of them in different model specifications.

[16] I choose $K = 3$ after sensitivity analysis.

initial traits at high school graduation, namely, mother's schooling S_i^m, father's schooling S_i^f, number of siblings N_i^{sib}, household structure at 14 HH_i, net family income Y_i^0, AFQT score $AFQT_i$ and age at high school graduation AGE_i^0, in the form of a multinomial logit. For $k = 2, \cdots, K$,

$$P_i^k = \frac{\exp\begin{bmatrix} \lambda_0^k + \lambda_1^k S_i^m + \lambda_2^k S_i^f + \lambda_3^k N_i^{sib} + \lambda_4^k HH_i \\ +\lambda_5^k Y_i^0 + \lambda_6^k AFQT_i + \lambda_7^k AGE_i^0 \end{bmatrix}}{1 + \sum_{l=2}^K \exp\begin{bmatrix} \lambda_0^l + \lambda_1^l S_i^m + \lambda_2^l S_i^f + \lambda_3^l N_i^{sib} + \lambda_4^l HH_i \\ +\lambda_5^l Y_i^0 + \lambda_6^l AFQT_i + \lambda_7^l AGE_i^0 \end{bmatrix}},$$

and normalize P_i^1 as

$$P_i^1 = \frac{1}{1 + \sum_{l=2}^K \exp\begin{bmatrix} \lambda_0^l + \lambda_1^l S_i^m + \lambda_2^l S_i^f + \lambda_3^l N_i^{sib} + \lambda_4^l HH_i \\ +\lambda_5^l Y_i^0 + \lambda_6^l AFQT_i + \lambda_7^l AGE_i^0 \end{bmatrix}}.$$

Achievement scores like high school grades and SAT scores may affect college entrance indirectly by the correlation with ability types like other background variables. They may also affect college choice directly if college acceptance depends on the grades or SAT scores. Due to data limitation as explained in the next section, I leave the introduction of grades to a schooling model like this to future research.

Further Parameterization with Heterogeneity I allow women of different skill types to have distinct taste for school and for nonemployment, different skill rental price and returns to schooling. In my estimation, these parameters will be type specific. The type specific utility function for individual i of type k when choosing alternative (s, h, m) at time t becomes

$$\begin{aligned} U_{it}^{shm} &= (\alpha_1 + \alpha_2 s_{it} + \alpha_3 h_{it} + \alpha_4 m_{it}) c_{it} \\ &\quad + v_1^k s_{it}(1 - h_{it})(1 - m_{it}) + v_2^k s_{it} h_{it}(1 - m_{it}) + v_3^k s_{it}(1 - h_{it}) m_{it} + v_4^k s_{it} h_{it} m_{it} \\ &\quad + v_5^k (1 - h_{it}) f_{it} + v_6^k (1 - h_{it})(1 - f_{it}) + M_{it} m_{it} + \epsilon_{it}^{shm}. \end{aligned}$$

Women with different skills also have different wage offer distributions. For skill type k, the wage offer when working in college is given by $\ln w_{it} = \beta_{0c}^k + \epsilon_{iwct}$, and the wage offer when working after college is determined by

$$\ln w_{it} = \beta_0^k + \beta_1^k S_{it} + \beta_2 H_{it} + \beta_3 H_{it}^2 + \beta_4 I\left(S_{it} \geq 16\right) + \epsilon_{iwt}.$$

We expect high skill women have both higher skill rental price and higher returns to schooling.

Furthermore women may also differ in taste for marriage and marriageability in the marriage market. I assume that there exist $m = 1, 2, \cdots, M$ different marriage types. A woman of skill type k has probability π_k^m of being marriage type m, so $\sum_{m=1}^M \pi_k^m = 1$ for all k. The value of marriage is then type specific:

$$M_{it} = a_0^m + a_1^m \Delta S_{it}^2 + a_2 age_{it} + a_3 f_{it} + a_4 mdur_{it}.$$

Women with high a_0 are more family oriented. If a_1 is negative and large in absolute value, it indicates that this type of women care much about schooling balance with their husbands. Moreover the marriage offer probability is also type specific such that:

$$\Pr_{it} = \frac{\exp(b_0^m + b_1 age_{it} + b_2 age_{it}^2 + b_3 I(S_{it} > 12))}{1 + \exp(b_0^m + b_1 age_{it} + b_2 age_{it}^2 + b_3 I(S_{it} > 12))}.$$

2.3 Solution to the Decision Problem

To solve the optimization problem, I define the value function $V_{it}(\Omega_{it})$ as the maximal value of the individual i's optimization problem at t:

$$V_{it}(\Omega_{it}) = \max_{\{c_{it}, s_{it}, h_{it}, m_{it}\}} E\left[\sum_{\tau=t}^{T_i} \beta^{\tau-t} U(c_{i\tau}, s_{i\tau}, h_{i\tau}, m_{i\tau} | \Omega_{it})\right].$$

The current utility at time t is defined as before, and the female maximizes the expected present value of her life time utility by the choice of $\{c_{it}, s_{it}, h_{it}, m_{it}\}$ for all $t = 1, 2, \cdots T_i$. Ω_{it} is the state space at t.

The value function can be written as the maximum over alternative-specific value functions

$V_{it}(\Omega_{it}) = \max_{(s_t,h_t,m_t)\in J}\{V_{it}^{shm}(\Omega_{it})\}$, which obeys the Bellman equation:

$$V_{it}^{shm}(\Omega_{it}) = U_{it}^{shm} + \beta E[V_{it+1}(\Omega_{it+1})|\Omega_{it},\ (s_t, h_t, m_t) \text{ is chosen at } t].$$

The alternative-specific value function assumes that future choices are optimally made for any given current decision. The randomness in utility arises from the fact that Ω_{it+1} is observable to the individual at time $t+1$ but unobservable at time t or before. We can separate the state space into a nonstochastic part and a stochastic part. Let $\overline{\Omega_{it}}$ be the nonstochastic part of the state space, which includes types, years of schooling, years of experience, marriage duration, age, choices, fertility and husband's schooling in the previous period, that is, $\overline{\Omega_{it}} = [type_i, S_{it}, H_{it}, mdur_{it}, age_{it}, s_{it-1}, h_{it-1}, m_{it-1}, f_{it-1}, S_{it-1}^H]$. Some of these state variables evolve endogenously: $S_{it} = S_{it-1} + s_{it}$, $H_{it} = H_{it-1} + h_{it}$, $mdur_{it} = m_{it}[mdur_{it-1} + m_{it}]$. Some of them are assumed to evolve exogenously: individual type, age, fertility status. The stochastic part of the state space includes the vector of the random shocks $\epsilon_{it} = [\epsilon_{i1t}, \cdots, \epsilon_{i8t}, \epsilon_{iwct}, \epsilon_{iwt}, \epsilon_{iy_Ht}]$, as well as job offer, marriage offer, and fertility realizations.

The model does not have an analytical solution, but it can be solved backwards numerically. To simplify the model, I assume that the optimization problem is divided into two sub-periods, as in Eckstein and Wolpin (1999). During the first $T_i - 1$, for each individual i, the model is solved explicitly. At the terminal period T_i, the current utility is given by $U_{iT_i}^{shm}$, and the expected future utility is assumed to be a given linear function of Ω_{iT_i}.

$$V_{iT_i}^{shm}(\Omega_{iT_i}) = U_{iT_i}^{shm} + \beta\{V_{iT_i+1}(\Omega_{iT_i+1})|\Omega_{iT_i},\ (s_{T_i}, h_{T_i}, m_{T_i}) \text{ is chosen at } T_i\}.$$

The present value of individual i's utility at $T_i + 1$ has following linear functional form in the state variables, $V_{iT_i+1}(\Omega_{iT_i+1}) = \delta_1 S_{iT_i+1} + \delta_2 H_{iT_i+1} + \delta_3 H_{iT_i+1}^2 + \delta_4 I(m_{iT_i} = 1)$.

Using the end condition, and assuming a known distribution of ϵ_{it}, the individual's optimization problem is solved recursively from the final period T_i. The numerical complexity arises because the value function requires high dimensional integrations for the computation of the "$E\max$ function" at each point of the state space. Following the procedure proposed in Keane and Wolpin (1994), I use Monte Carlo integrations to evaluate the integrals.

2.4 A Simple Example

I use a simple two-period model to illustrate how the marriage market, through the marriage offer rate, assortative mating and husband's income, affects women's college choice. In the model, high school women decide whether to attend college in the first period and whether to marry in the second period. The model does not distinguish between college attendance and graduation. It sheds some light on the importance of unobserved heterogeneity and the interplay between the marriage market and the college choice.

There is a continuum of individuals in the model. I normalize the size to one. Each individual i is a woman with a high school degree. Her ability is δ_i which is randomly drawn from some known, fixed distribution $\Phi(\cdot)$. Individuals in the model live for two periods. In the first period, all of them are single and they simply decide whether to attend and graduate from college. To acquire college education, they need to pay a fixed cost cs. In the second period, everyone has finished formal schooling and the only remaining choice is marital status. Schooling and labor supply are mutually exclusive. Individuals either attend school or work in each period. Labor earnings depend on both individual schooling S_i and ability δ_i and take the form of $\ln y(S_i, \delta_i) = \beta_0(\delta_i) + \beta_1 S_i$. Skill rental price β_0 depends on ability, $\beta_0'(\delta) > 0$, and β_1 measures the effect of education on earnings.

Let p be the marriage offer arrival rate in the second period. There exist measure one potential husbands in the economy with proportion μ being college graduates and proportion $1-\mu$ being high school graduates. Each woman then has probability μp to receive an offer from a college man, $(1-\mu)p$ to receive an offer from a high school man and $1-p$ probability with no offer. College is assumed to be an active matching place such that marriage offer rate of college women p^1 is greater than that of high school women p^0. A marriage is formulated only if a woman receives and accepts an offer. A married woman benefits from her marriage in two ways: first, marriage provides utility denoted by M; second, a fraction ψ of her husband's income y^H is available for her consumption. Husband's income increases in his schooling S^H.

The preference of individual i is linear and of the form $U_{it} = c_{it} + M_{it}m_{it}$, where c_{it} is the consumption at t, m_{it} denotes marital status ($m_{i1} = 0$) and M_{it} is the net utility value of marriage. Let s_{it} denote the schooling choice for individual i. s_{it} equals one if she attends school and zero otherwise. Each individual i solves the following problem:

$$Max_{\{s_{i1},m_{i2}\}} E\{c_{i1} + \beta[c_{i2} + M_{i2}m_{i2}]\}$$
$$\begin{aligned} s.t.\;\; c_{i1} + cs \cdot s_{i1} &\leq (1 - s_{i1})y(1, \delta_i) \\ c_{i2} &\leq y(1 + s_{i1}, \delta_i) + \psi y^H \left(S_{i2}^H\right) m_{i2}, \end{aligned}$$

where the expectation is taken over the marriage offer probabilities.

The value of marriage is specified as $M_{i2} = a_0 - \gamma(S_{i2} - S_{i2}^H)^2$, where S_{i2} and S_{i2}^H are individual i and her husband's schooling. There are only two schooling levels in the model and I use $S = 1$ to denote high school and $S = 2$ to denote college. The value of marriage first depends on some permanent utility value a_0 through affection, children etc. The value of marriage also depends on the couple's homogeneity in education background. If they are different in schooling level, disutility γ occurs.

The model can be solved analytically. In the second period, the marital choice depends on total marriage value $\psi y^H \left(S^H\right) + a_0 - \gamma(S - S^H)^2$. A marriage is formulated if this value is positive. College choice at $t = 1$ depends on individual ability. Since earnings increase with ability, all individuals with ability above some threshold attend college.

The benefits of marrying a high school man and a college man are $\psi y^H(1) + a_0$ and $\psi y^H(2) + a_0$ respectively. If a woman marries a man with different schooling level, then cost γ occurs. When $\gamma < \psi y^H(1) + a_0$, no matter which type the man is, the net value of marriage is always positive. Therefore all women accept offers from both types of men and match is random.[17] When $\psi y^H(1) + a_0 < \gamma < \psi y^H(2) + a_0$, a high school woman accepts all offers since even if she marries a college man and γ occurs, college husband's higher income compensates for the cost. However college women are more selective in this case, they choose college men only. Positive assortative matching thus appears. When $\gamma > \psi y^H(2) + a_0$, the cost of marrying someone with different schooling level is so high that a woman will only choose a man with the same educational attainment. Then the assortative mating is perfect. Therefore we have the next proposition providing sufficient

[17] This result holds for the two-period model or a multiple-period model with on the marriage search. For a finite horizon model with more than two periods and married women receive no outside offers, it may be optimal for women to wait until a good match comes to marry.

conditions under which positive assortative mating appears.

Proposition 1 *(Assortative mating) Degree of assortative mating in education depends positively on* γ*. When* $\gamma < \psi y^H(1) + a_0$*, the match is random; when* $\psi y^H(1) + a_0 < \gamma < \psi y^H(2) + a_0$*, positive assortative mating exists; when* $\gamma > \psi y^H(2) + a_0$*, the match is perfectly assortative.*

The next proposition shows how future marriage affects college choice.

Proposition 2 *(Marriage ⇒ College) Given* $\alpha_1 > \ln(\frac{1+\beta}{\beta})$*, the more college men available, the more women attend college; the more college increases marriage offer rate, the more women attend college. College enrollment also increases in* γ *when* $\psi y^H(1) + a_0 < \gamma < \psi y^H(2) + a_0$*.*

Proof. *See Appendix A.* ■

3 Data

3.1 The Sample

The micro data are taken from the 1979-98 waves of the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14-22 years old when they were first surveyed in 1979. These individuals were interviewed annually through 1994 and are now interviewed on a biennial basis. The sample contains a core random sample and oversamples blacks, Hispanics, "disadvantaged" whites, and members of the military. A key feature of this survey is that it gathers information in an event history format, in which dates are collected for the beginning and ending of important life events such as employment, education and marriage. I follow each individual in my sample for up to ten years since they received their high school diploma.

The sample used in the present analysis consists of white females from the core random sample of 2,279 individuals. I keep those who have received a high school diploma and reported graduation date. All women in the sample graduated from high school during May to August between 1980-1983. I further restrict my sample such that every woman graduated from high school between age 17 and 19, single and with no children.[18] I dropped 89 individuals from the sample because of

[18] Complete schooling history is not available before 1980, therefore the sample is restricted to high school grad-

inconsistent or incomplete observations on schooling, employment or marital choices. This leaves me with a sample of 582 women born between 1961-1964. Another 95 women are excluded from this study since their family background information is not complete. Selected individuals stay in the sample up to ten years as long as consecutive annual schooling, employment and marriage profiles are observed. The empirical analysis is based on this sample of 487 females with a total of 4,770 person-year observations. Data based on answers to retrospective questions are aggregated as described in Appendix B. Some descriptive statistics are shown in Table 1. On average, this sample of high school graduate women completed more than 14 years of schooling.

3.2 Women's Choices and Transitions

Figure 1 presents the proportions of women who are in college, who are employed, who are married, and who have children for the first ten years since graduating from high school. Conditional on high school graduation, 61 percent of my sample acquired at least some post secondary education and 38 percent had at least four years of post secondary schooling. 49 percent of the sample attend college in the first year after high school. Attendance falls by 4 to 5 percent annually throughout the next three years. Then we observe a more than 15 percent discrete drop in attendance after the fourth year, corresponding to typical college graduation. The attendance rate continues to fall but stays around 9 percent after seven years. This reflects the fact that some women return to school. About one third of women in the sample have the experience of leaving and subsequently returning to school.[19] This may due to female's low returns to experience. The labor force participation rate increases from 43 percent to about 80 percent in the first six years after high school. Then it becomes flat and declines slightly, reflecting the well-known hump-shaped female labor supply profile. This is consistent with women getting married and withdrawing from the labor market. By the tenth year the percentage of women who are married has increased to about 66 percent and the percentage of women who have children has increased to 45 percent.

Table 2 shows the disaggregate choice proportions. Each entry of the table is the proportion

uates after 1980. 7 individuals graduated after 1983. 9 individuals graduated before 17 or after 19. More than 96% of the sample receive high school diploma during May to August. 24 women were married or had children at graduation.

[19]This is very different from men. In Cameron and Heckman (2001), it is documented that only 2-6 percent of high school graduates and 6-12 percent of dropouts report at least one episode of leaving and then returning to school.

of women who chose one of the eight alternatives in each year after high school. Conditional proportions can be easily calculated from this table. The participation rate of married women is significantly lower than that of single women except for the first few years when few women are married. Another interesting observation is that very few married women stay in school. It indicates low complementarity between marriage and school. The attendance rate for married women is always below 10 percent as compared to single women whose attendance rate is as high as 51 percent in the first year after high school.

Even though the sample women are in their twenties, many of them have already undergone one or more changes in marital status. 142 women (29%) remained single throughout the sample period, 25 (5%) married twice, 54 (11%) experience at least one divorce. Most (about 60 percent) of divorced women never went to college. Marriage seems to be more stable for well educated women.[20] Table 3 shows mean transitions between attendance and non-attendance; between employment and non-employment; and between being single and being married. Each row presents the probability of moving to one state at $t+1$ given the choice at t. Persistence in choices is indicated by the high probability on the diagonal.

As a parsimonious way of describing the joint patterns of school attendance, marriage, and employment, Table 4 and 5 present probit results. Women, who are younger, who are single, who have no children, and who do not work, are more likely to attend college. Women, who have children, who are older, who do not work, and who are not at school, are more likely to be married. Women who have more experience, who are younger and not in school, who are single with no children, and more schooling are more likely to work. Furthermore column (2) and column (3) of Table 5 display employment probits for married women with two different specifications. Comparing these results, we observe the following: first, women with low income husbands are more likely to work;[21] second, without controlling for husband's income, schooling has less effect on employment. This is probably due to the fact that a highly educated woman marries more

[20]From life cycle perspective, this number is probably biased since college graduates get married much later. Therefore it is less likely for us to observe their divorce over the same period of time. However some aggregate data show the same pattern. Based on data from the National Survey of Family Growth (NSFG), "Cohabitation, Marriage, Divorce, and Remarriage in the United States" Table 25 reports that among non-Hispanic 20 to 44 years old white women in 1995, the probability of first marriage disruption after 15 years is 55% for high school dropouts, 45% for high school graduates, and 36% for women with more than high school education.

[21]For example, in the 10th year of the sample, 86 percent of women with husband's earnings less than 20 thousand worked and only 55 percent of those with husband's earnings over 80 thousand worked.

often to a man with higher education and income, which induces her to work less. Therefore without controlling for husband's income, the effect of schooling on employment probability is under estimated.

3.3 Women's Wages

Real hourly wages are obtained as explained in Appendix B. Table 6 reports both the mean and deciles of hourly wages. The mean wage more than doubles over ten years. Except for the first year mean wage is always greater than the median wage showing the wage distribution is skewed. The observed wage distribution shifts to the right and becomes more dispersed as women accumulate more schooling and experience.

In solving the dynamic programme, actual hours worked are ignored. Potential annual earnings, obtained by multiplying hourly wage by 2000 hours, is used. Each woman is essentially assumed to be deciding about full-time work and the wage rate is assumed to be independent of hours worked. Among all the wage observations, wages of women who work when at school are much lower and less dispersed. Following the convention, I use wage observations after formal schooling to run an OLS log wage regression on years of schooling and experience. The regression yields the following coefficients with standard errors in parentheses: $\beta_0(\text{constant}) = 0.712$ (0.051), $\beta_1(\text{schooling}) = 0.081\ (0.004)$, $\beta_2(\text{experience}) = 0.122\ (0.009)$, $\beta_3(\text{experience}^2) = -0.005$ (0.001). The concavity of the experience profile and the positive schooling effect are consistent with many other studies.

3.4 Marriages

Mean age at the first marriage is approximately 22, 23, and 25 for high school women, some college, and college women respectively. At the time of first marriage, men are on average three years older than women, nevertheless women have slightly higher schooling. Mean annual income of husbands is around 35 thousand. It increases from 21 thousand to 38 thousand during the sample period.

Married couples tend to share a common schooling background. The correlation between women's highest grades completed (HGC) and their husbands' HGC at the first marriage is as much as 0.55. At time of the first marriage, 42 percent of the couples have the same educational attainment. About 80 percent of the couples have at most two years schooling difference. Table 7 shows schooling homogamy by husband's schooling distribution conditional on married women's education. High school women are very likely to marry high school men. As a matter of fact, 78 percent of them marry men with a high school education or less. As women accumulate more schooling, they tend to marry men with more education. For college women, 60 percent of their husbands are college graduates as compared to for high school women, less than 7 percent of their spouses are college graduates. Women with some college education, but who never finish 4 years of college, seem to be more similar to high school graduates. If schooling homogamy provides positive value to marriage, we expect marriages in which partners share similar education background be more stable. Due to lack of observations, however, distributions of schooling difference are not statistically different for marriages survived and divorced during the sample periods.

3.5 Background at High School Graduation

I use both family and cognitive background variables as initial conditions in the schooling model. Highest grade completed of a woman's mother and father, number of siblings, and whether the woman came from a broken family (i.e. one or both biological parents were absent) are measured at age 14. Family income measures parental income for dependent respondents. A dependent is defined by NLSY as a person living at home or not at home but living in a dorm or military barrack. Thus family income is generally not known for older NLSY respondents. A two year average was constructed for family income at ages 15 and 16 if available. Family income at age 14 and age 17 is used if the data are missing at age 15 or 16. Family income is measured in 2000 dollars.

Three surveys, conducted independently of the regular NLSY79 interviews, collected aptitude and intelligence score information: (1) The Armed Services Vocational Aptitude Battery ($ASVAB$), a special survey administered in 1980 to NLSY79 respondents (94% of the 1979 sample participated); (2) the 1980 survey of high schools, which collected scores from various aptitude/intelligence tests and a variety of college entrance exams such as the Preliminary Scholastic Aptitude Test ($PSAT$), the Scholastic Aptitude Test (SAT), and the American College Test

(ACT); and (3) the 1980-83 collection of high school transcript information. The type of information gathered for each of up to 64 courses included grade level at which the course was taken, a code for the high school course, the final or computed grade for that course, the source of the final grade, and the credits received.[22] The $ASVAB$ consists of a battery of 10 tests that measure knowledge and skill in 10 different areas. Armed Forces Qualifications Test score ($AFQT$) is a composite score derived from 4 sections of the battery (namely arithmetic reasoning, word knowledge, paragraph comprehension and math knowledge) and widely used as cognitive ability indicator. $AFQT89$ percentile scores are used in this study.

College entrance examination scores may be important for college application. They are not included in the analysis since the number of respondents for whom these scores are available is low. Consider three major college entrance exams, namely $PSAT$, SAT and ACT, within my sample, 93 individuals report SAT scores, 109 report $PSAT$ scores and 102 report ACT scores, overall only 40% of the sample has at least one usable test score. When evaluating applications, schools use an SAT type of achievement score as a signal for individual ability. This study assumes that an SAT score is of second order importance conditional on ability.

Table 8 illustrates the potential importance of family and cognitive background in determining school, employment and marriage outcomes. As the first panel of the table shows, the difference in completed schooling between high school women whose mother did not complete high school and women whose mother completed college is over 3 years. Of the former group, 64 percent of them never attend college while about 82 percent of the latter completed college. Similar patterns hold for women's schooling conditional on father's education. Given women's schooling differences, labor market and marriage outcomes are also significantly related to parents' schooling. The real hourly wage rate over the ages of 25 and 28 for those who are employed increases more than half over the range of parents' schooling. Much fewer women whose mother or father completed college marry at the age of 25. The third and fourth panels of the table show outcomes in school, employment and marriage conditional on number of siblings and household structure. Number of sibling has small effect on schooling outcome if it is less than four. Having more than four siblings reduces years of schooling and real hourly wage. 24 percent of women with no sibling are

[22]High school grades may be important for college decision, which I leave for future research due to the complications in collecting data (see Eckstein and Wolpin 1999).

married at age 25 while 59 percent of women with more than four sibling are married at the same age. Women who live with both parents obtain half year more schooling, marry slightly less at age 25 and their hourly wage is lower. The fifth and sixth panels show the well known correlation between family income, cognitive ability and youths' outcomes. Women whose family income is greater than twice the median obtain almost two years more schooling than women whose family income is less than half of the median. Women from rich family also perform significantly better in the labor market and fewer of them are married at age 25. $AFQT$ scores are strongly correlated with schooling outcome. 79 percent of women with top 20 percentile of $AFQT$ scores complete college while 77 percent of women with bottom 20 percentile $AFQT$ scores never attend college. The former group's hourly wage almost doubles as compared to the latter. The last panel of the table shows that on average women finish high school at 18. Those who graduate at age 17 obtain 1.8 more years of schooling and do significantly better in the labor market than those who graduate at age 19.

4 Estimation

After solving the optimization problem, I generate data from the behavior model and use simulated method of moments (McFadden 1989, Pakes and Pollard 1989) to estimate parameters in the economy.[23]

4.1 Simulated Method of Moments Estimator

Specification I restrict the model to have exogenous processes on fertility and exogenous schooling distribution of potential husbands. The discount factor β is set to be 0.96, i.e. an annual rate of time preference of 4 percent. Parameters to be estimated consist of parameters that correlate observed background with unobserved types, utility parameters, parameters in the budget constraint, parameters in the marriage offer function and the value of marriage, women's wage equations, men's earning equation, end condition and the variance covariance matrix of idiosyncratic shocks.

[23] Simulated maximum likelihood estimator is efficient but it requires the number of simulation be large. SMM estimator, however, is asymptotically normal as long as the number of observations is large.

Data I have a sample of white female high school graduates indexed by $i = 1, \cdots, 487$. I observe their family and cognitive background (mother's schooling, father's schooling, number of siblings, household structure at 14, net family income, $AFQT$ score and age at high school graduation), their schooling, employment and marital status every year $\left(s_{it}^D, h_{it}^D, m_{it}^D\right)$, their observed wages if employed (w_{it}^{oD}) and characteristics of the first marriage (woman i's own schooling S_i^D, her husband's years of schooling S_i^{HD}, and annual income y_i^{HD}) if married, for $i = 1, \cdots, 487$ and $t = 1, \cdots, T_i$, where the superscript D denotes the data.

Simulations I simulate individual choices, wages, husbands' schooling and income from the model in a consistent way as in the data. All women start with 12 years of schooling, no labor market experience, and having never married or had children. For each individual i, I first simulate her type conditional on her background. At the beginning of the first year, $t_i = 1$, all other uncertainty in the economy is also realized. Preference and productivity shocks are known to woman i, as well as whether job offer, marriage offer or child arrives and the type of offers received. Using the distribution of the shocks, she also forms expectations on future utility and earnings given her current decision. She makes joint decision on schooling, employment and marriage $\left(s_{i1}^S, h_{i1}^S, m_{i1}^S\right)$. Her wage w_{i1}^S is recorded if employed and her husband's schooling S_i^{HS} and income y_i^{HS} are recorded if married. The states are then updated. Now at $t_i = 2$, conditional on the current states and all the idiosyncratic shocks, $\left(s_{i2}^S, h_{i2}^S, m_{i2}^S\right)$, w_{i2}^S, S_i^{HS}, y_i^{HS} are simulated. If a woman is working and her wage is observed, I simulate the measurement error to obtain the "observed" wage according to $w_{it_i}^{oS} = w_{it_i}^S \exp(u)$. Observed husband's income is simulated in a similar way. Given the value of parameters, I simulate data from the model for $N^S = 25$ times for each individual.

Moments The moments used include the proportions of women who choose each of the eight alternatives in each year as in Table 2 and the aggregated proportion attending college, working and married as shown in Figure 1; the proportions of high school graduate, some college and college graduate women; transitions moments as in Table 3; husband's schooling distribution conditional on married women's education as in Table 7; mean and standard deviation of husband's annual income; as well as observed mean wage and wage decile moments as in Table 6. In total, there are 254 data moments. For each simulation, the same moments are computed and the simulated moments are averaged over all simulations.

Implementation Simulated method of moments is implemented by using these moments. Let m_j^D be moment j in the data and $m_j^S(\theta)$ be moment j from the model simulation given the parameter vector θ. The moment vector is

$$g'(\theta) = [m_1^D - m_1^S(\theta), \cdots, m_j^D - m_j^S(\theta), \cdots, m_J^D - m_J^S(\theta)],$$

where J is the total number of moments and $J = 254$. I minimize the objective function $J(\theta) = g(\theta)' W g(\theta)$ with respect to θ, where the weighting matrix W is set to be the identity matrix. I bootstrap the standard errors.

4.2 Identification Issues

In general the non-linearity makes it difficult to establish theoretical and practical identification. Although the model is complex it is nevertheless possible to provide intuition concerning the identification of the importance of the three sources, namely, the background, the earnings, and the marriage.

Background variables enter the model as covariates of discrete types. First, given the assumption that types are discrete and uniformly distributed, we can identify the joint distribution of types, i.e. the proportions, from a cross section as in Roy model (Heckman and Honore, 1990). Second, type specific parameters are like persistent individual effect, which are identified by repeated observations on individuals. Third, in the data, girls whose mother went to college go to college and girls with high test score go to college. Mother's schooling, test scores, etc. are exogenous, so it is easy to identify the correlation between these background variables and unobserved types, given the type distribution.

In order to identify the earnings effect on college decision, we need to estimate a nonlinear simultaneous equation system. We need to first identify the causal effect of college on earnings, then we need to identify the causal effect of expected earnings on college decision. The college wage premium is identified from the wage data on those who enrolled college, those who graduated from college, and those who didn't attend college, like in a Heckman selection model. Given that the wage is log normal, the treatment effect of college is identified. The dynamic programming model provides the decision rule for college attendance and graduation conditional on wage premium. The decision rule is essentially a structural probit equation. *i.i.d.* wage shocks provide exogenous variation in wage premium.

Similarly we need to estimate another nonlinear simultaneous equation system to identify the marriage effect on college decision. We do not observe marriage utility but we observe marriage outcome. Like in McFadden's random utility model, parameters in the marriage utility function can be identified. The observations that high educated women marry highly educated men provide the variation to identify the disutility of school difference. The transition from singlehood to marriage conditional on schooling identifies effect of college on marriage offer rate. Women's own wage can be used as an exclusion restriction as it affects schooling but not the marriage utility. Given the dynamic programming model, the structural probit equation of college decision thus also includes the expected value of marriage. Then the idiosyncratic shocks on the value of marriage play exactly the same role as the wage residuals.

We can write down the wage equation, the marriage utility function, and structural probit equations for each of the joint choices (8 choices) for every year (10 years). These are approximations of the dynamic joint decision process without coherent cross equation restrictions. With the exclusion restrictions from the model, coefficients from reduced form estimates are functions of the structural parameters.

As another way to think about identification, a necessary condition is that each parameter should affect some moments in the distribution. Consider first the parameters in the labor earning processes. The identification apparently rests on the wage data. We observe only accepted wages and high skill women obtain more schooling than low skill women. Conventional OLS regression thus suffers sample selection and endogeneity biases. The solution to the optimization problem provides the sample selection rules, which serves the same purpose as would a sample selection correction in a two-step procedure. Unobserved heterogeneity is explicitly specified so that the endogeneity bias is also corrected. Similarly, men's earning process is identified by observed husband's earning data.

Consider next the identification of the utility function parameters. If the model were static, the proportion of the choices identify the constants in the utility function. The value of schooling v_1 to v_4 are identified by the attendance conditional on labor supply and marital status. Similarly the value of nonemployment v_5 to v_6 are identified by the participation rate conditional on fertility. As an example, write alternative specific utility function incorporating the budget constraint as following (given $f_t = 0$ and $S_t < 16$):

$$
\begin{aligned}
U_t^1 &= v_{6t} + \epsilon_t^1, \\
U_t^2 &= -(\alpha_1 + \alpha_2)cs + v_{1t} + v_{6t} + \epsilon_t^2, \\
U_t^3 &= (\alpha_1 + \alpha_3)y_t + \epsilon_t^3, \\
U_t^4 &= (\alpha_1 + \alpha_2 + \alpha_3)(y_t - cs) + v_{2t} + \epsilon_t^4, \\
U_t^5 &= (\alpha_1 + \alpha_4)\psi(0)y_t^H + v_{6t} + M_t + \epsilon_t^5, \\
U_t^6 &= (\alpha_1 + \alpha_2 + \alpha_4)(\psi(0)\, y_t^H - cs) + v_{3t} + v_{6t} + M_t + \epsilon_t^6, \\
U_t^7 &= (\alpha_1 + \alpha_3 + \alpha_4)(y_t + \psi(1)\, y_t^H) + M_t + \epsilon_t^7, \\
U_t^8 &= (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)(y_t + \psi(1)\, y_t^H - cs) + v_{4t} + M_t + \epsilon_t^8.
\end{aligned}
$$

The maximization problem written in this "reduced form" representation is similar to a static multinomial choice model. The coefficients are functions of the utility and budget constraint parameters. All parameters in these equation are identified by the proportions of choices and the variation in women's earnings and husbands' earnings except for the cost of college cs. cs enters the model linearly with the value of schooling v_1 to v_4. In the estimation, I set $cs = 7,515$ in 2000 dollars[24]. The cost of graduate school, however, can be identified by the discrete drop in attendance after four years. With regard to parameters in the value of marriage, the effects of age, fertility, and marriage duration on the value of marriage can be identified by conditional marriage proportions and marriage transitions. The disutility from the difference in schooling is identified by husband's schooling distribution conditional on married women's schooling.

The panel data provides conditional transitions from eight states to eight new states from

[24]According to National Center for Education Statistics (NCES, Digest of Education Statistics, 1990, pp285, Table 281), the average total tuition room and board cost was $7,515 (dollars of 2000) during 1980-1988.

one year to another for 10 years. These transitions would allow us to identify parameters in our structural model which characterize the dynamics: the job offer probabilities, the marriage offer function, etc. The terminal value parameters are identified by their joint restrictions on the transitions between states over time and the cross section choice.

5 Estimation Results

The model is estimated by minimizing the squared difference between data and simulated moments as previously defined. In this section I discuss the estimation results and their economic interpretation, as well as the fit of the model to observed moments, followed by an out of sample prediction.

The probability of the first birth is estimated separately and used as inputs to the estimation algorithm.[25] I estimate a logit using individual's characteristics to determine the probability of having first child for each period. The results are presented in Appendix C. Schooling has a negative effect on the probability of having children, which is consistent with the observation that highly educated women tend to have fewer children and have the first birth at older age. The estimates also show that married women are more likely than single women to have children and as they become older, their probability of having at least one child increases but at a diminishing rate. Potential husbands' schooling distribution is treated as exogenous in the model. I calculate the schooling distribution of 22 to 35 years old white males between 1980 to 1983 from CPS and use it as non-parametric estimates of potential husbands' schooling distribution. These estimates are also presented in Appendix C.

5.1 Parameter Estimates

The model's estimated parameters are reported in Appendix D with standard errors in parentheses. In total, there are 102 parameters. The first panel reports the estimates of the parameters in the utility function. Marginal utility from consumption is estimated to be 1.158 for single,

[25]The probability of the first birth depends on schooling and marital status, which are correlated with unobservables (ability, taste for marriage, etc.). Therefore, the logit estimates may be biased and inconsistent. With unobserved heterogeneity, the two step procedure is in general not consistent. I assume the potential bias is small and adopt two step procedure as in Van Der Klaauw (1996).

non-employed women, who are not in college. The negativity of α_2, α_3 and α_4 implies that the utility gains from consumption decrease when women attend school, work or stay married. Estimated utility values of school and nonemployment indicate significant heterogeneity between women's skill types. With respect to tastes for school, type one likes school the least, type two the next and type three is the type who likes school the most independent of working and marital status, as can be observed from the rank order of values of v_1, v_2, v_3 and v_4's. Attending school when married brings disutility for all types, which is consistent with women leaving school after marriage. The value of nonemployment is higher when children are not present. Type two's have the highest value for nonemployment.

Average annual expenses for graduate school are around \$35,552. The estimated cost of children is \$42,370. Although this amount may seem large, it is important to remember that in the model, child care does not take time but it should be included in the estimated cost. This estimate also sums over the total number of children women may have which is around three. Without specifying the strategic behavior within the household, the model predicts that nonemployed married women receive 48.4 percent of their husbands' income, while employed married women get 32.4 percent. This is consistent with the case when married couples share their income. If the man specializes in the marketplace and the woman at home, the stay home married woman is compensated by the husband. Higher income the husband earns, *ceteris paribus*, less likely the married woman works at the marketplace.

According to the estimated correlation between background and type, λ's, higher parental education, fewer siblings, living with both parents at 14, higher family income, good $AFQT$ score and graduate high school at an early age increase the probability of being skill type two. Similarly parental schooling, family income and $AFQT$ score also have positive (but less) impact on the probability of being skill type three. We expect these two types have higher skills relative to the first type. As seen from the estimates of conditional marriage type proportions, π's, each skill type has different distribution of marriage types. For example, all skill type 2's are marriage type 1 but only 66% of skill type 3's belong to marriage type 1.

Job offer rates depend positively on labor market attachment in the previous year. Offer rate for employed women is higher than for non-employed women independent of schooling. Women with some college or college degree always receive more job offers than high school women. According to the estimates of the wage equation parameters, both skill rental price and return to schooling are the lowest for the first type. Type two have the highest skill rental price while type three have the highest return to schooling. Each additional year of schooling increases wages by 4.2%, 5.5%, 6.2% respectively for each type. Note that the estimated return to schooling are much lower than the OLS estimates, providing evidence that without controlling for self selection, the returns to schooling is upward biased.[26] Wages increase by 10.1% with each additional year of after school experience and the depreciation is 0.08%. The return to experience seems high since my sample is at the beginning of their labor market experience, when they accumulate skills fast and shop for jobs frequently. College graduation increases wage by 29.6% conditional on years of schooling and experience. Even though skill type 1's have much lower skill rental price and returns to schooling for the formal labor market, they seem to have comparative advantage for jobs available at school as indicated by the highest β_{0c}. Wages offers received in school are much less dispersed than those received out of school.

Based on the estimates from the marriage evaluation rule, the negative a_1 shows that education attainment of both spouses are complements within the family. Women value marriage more when they grow older, which is captured by positive a_2. The value of marriage also depends positively on the presence of child and marriage duration. Children provide large utility for women, which is consistent with the argument that the main reason why people marry is to have their own children. Positive dependence on marriage duration reflects the likely accumulation of physical and emotional bond between the spouses associated with marriage. The estimates of the marriage offer probability function show that age and college attendance have significant effect on the arrival probability of marriage offers. We also observe considerable heterogeneity between marriage types. Women from various marriage types differ in their permanent value for marriage a_0, marriage offer arrival rate b_0, and their preference for husbands' schooling a_1. Marriage type 1's fixed value for marriage is the lowest and the difference in schooling with the husbands gives them the largest disutility. Interestingly, they receive marriage offers most often among all types. Potential husband's earning function is increasing in schooling and concave in experience.

[26] See Card (2001) for a recent survey on the complexity in estimating the return to schooling.

5.2 Within-Sample Fit

Choice Proportions and Transitions Given the estimated parameters of the model, I calculate the predicted proportions of women who choose each alternative in each year after high school. Figure 2 (a) and (b) depict the fit of the model to the choice proportions. Each of the profiles implied by the estimated model has approximately the right shape and matches the levels of the data quite closely. More formally, Table 9 presents the within-sample χ^2 goodness of fit statistics for the model with respect to choice proportions, by years after high school graduation. The model prediction is statistically the same as the data moments at the five percent level. As for the overall schooling distribution, the model predicts 61.0% of the sample attend college and 38.0% finish four-year college as compared in the data 61.4% attend at least one year college and 37.8% complete four years.

Table 10 presents the predicted mean transitions based on the same simulations that generated the choice distributions in Figure 2. The model can match transitions reasonably well. The data demonstrates much persistence in each state, the model recovers persistence in attendance status and marital status but somewhat underpredicts the persistence in nonemployment. Individual heterogeneity and state dependency generate persistence in the model.

Wage Moments The estimated model fits well the trend and the level of the mean and deciles of accepted wages. Figure 3 (a) compares the model's mean wage and median wage profiles, with the parallel moments in the data. With the accumulation of both schooling and experience, mean wage doubles from \$6.2 to \$12.7. Mean hourly wage jumps by \$2 in the 5th year, reflecting college women joining the labor force with one year lag. The model does not capture this feature but it well captures the trend in both mean wage and median wage. The model predicts median wage lower than mean wage, indicating a skewed wage distribution as we observed in the data. Moreover, the growing distance between median and mean reveals wages to be more dispersed. Figure 3 (b) compares the mean and standard deviation for husband's annual income. Again the model is able to fit the level and the trend of both.

Assortative Matching in Education As Table 11 presents, the predicted husband's schooling distribution conditional on married women's schooling level matches data closely. In the model, high school graduate women like to marry college men for their high income, but they suffer from the difference in education background and they receive fewer marriage proposals. The model underpredicts their probability of marrying college men. Women who attend college but never finish four years behave more like high school graduates. Overall the model can fit the conditional schooling distribution of husbands.

Comparison with Reduced Form Model A comparison of within sample prediction from the structural model and a multinomial probit model is another method of assessing the fit of the model. The reduced form parameters of the probit are unspecified functions of the structural parameters of the optimization model. Likelihood ratio chi-square tests yield values of 1485, 1684 and 1771 for attendance, marriage and employment probit models in Table 4 and 5. If we use reduced form model to fit the observed distribution of data with all the rich transitions, we potentially need to specify eight nested multinomial simultaneous equations for every year, and all of them should include individual fix effect. This system of equations would probably have more parameters than the structural model I consider here.

5.3 Observed and Unobserved Heterogeneity

As can be seen in Appendix D, there is considerable variation in type-specific skill endowments and preferences. Table 12 presents selected characteristics at the end of the sample period on the basis of simulations using the estimated model. I first consider unobserved types in skills. Skill types differ substantially in their highest grades completed, work experience, marriage duration and choices in the 10th year after high school graduation. Among type 1's, 74% never attend college and none of them finish four years college for those 26% who attend. Only 5% of type 2's are high school graduates only and 72% graduate from college. An overwhelmingly 98% of type 3's graduate from college. Basically Type 1 is the high school type, type 2 is the college type and type 3 is the graduate school type. High skill type also tend to work more after leaving school. The hourly wage for the second type is 48% higher than the hourly wage for the first type. The hourly wage for the third type is low because many of them are working when in graduate school.

Not only are these unobserved skill traits related to school performance and work experience, they are also related to marriage experiences. Each skill type consists different composition of marriage types. Skill type 1's consist of mostly marriage type 1 and 2. In particular, 73.2% of skill type 1's belong to marriage type 1 and 26.0% of them are marriage type 2. Almost all skill type 2's belong to marriage type 1. For skill type 3's, 64.0% are marriage type 1, 14.9% are marriage type 2 and the rest 21.1% are marriage type 3. The estimates of type specific b_0's show that marriage type 1 receive much more offers than type 3, even more than type 2. Figure 4 shows how the value of marriage depends on school differences of the couple for each type and for different age groups. The top panel plots the value of marriage M by the difference in years of schooling with potential husband for single 25-year old women with no children. Type 1's are the most choosy type. When husband's income is not taken into consideration, type 1's accept marriage proposals only from men with the same schooling whereas type 2's also accept offers from men with one year schooling difference and type 3's marry men with up to 4 years difference in schooling. The dotted line illustrates potential husband's incomes, which are increasing in their schooling. Under the assumption that married women can consume part of their husbands income, marriage opportunity would be more attractive when husband's income is taken into account. The bottom panel in figure 4 shows the age effect on marital choice. As women grow older, they become less selective but the effect is relatively small.

Given the importance of unobserved types in determining women's schooling, employment and marriage choices, and observed correlation between family and cognitive background with outcomes, the model predicts strong correlation between observed background variables and unobserved types. Although I cannot determine each individual's actual type, I can assign a set of type probabilities conditional on her family and cognitive background. Table 13 shows the correlation between observed background with unobserved skill types. Since background variables are likely correlated with each other, for example, if both parents are college graduates, family income is probably in top quantile, table 8 cannot separate the influence from each background variable. I consider the marginal contribution of each variable on the skill type distribution. For example, to study the correlation between mother's schooling and young women's skill type, I fix other background variables at the sample means and then compute the type probabilities conditional on mothers being high school incomplete, high school graduates, some college or college graduates. As Table 13 shows, family and cognitive background variables have strong predictive power on the probabilities of being skill type 1 and type 2. Higher completed schooling of parents, family income and $AFQT$ score, and lower number of siblings imply higher proportion of skill type 2 and lower proportion of skill type 1. Mother's schooling has stronger correlation with the skill types than father's schooling. Conditional on everything else, family income is not a strong predictor for skill types at least if the income is below twice of the median income.[27] Furthermore living with both parents at age 14 and graduating high school at younger age increase the probability of being type 2. The probability of being type 3, however, is not strongly correlated with background variables.

5.4 Out of Sample Predictions for NLSY97 Sample

The National Longitudinal Survey of Youth 1997 (NLSY97) is designed to be representative of the U.S. population in 1997 born during the period 1980 to 1984. NLSY97 sample consists of 8,984 youths age 12-16 as of December 31, 1996. Two subsamples comprise the NLSY97 sample: a cross section random sample and supplemental oversamples of Hispanics and blacks. NLSY97 gathers information in an event history format like NLSY79. Since these two surveys ask the same questions to respondents I can use them to compare college enrollment behavior in the early 1980s and in the early 2000s.

NLSY97 rounds 1-6 with event history released in October 2004 is used in this study. I construct comparable data from the NLSY97 using the same restrictions as for the NLSY79 sample. First I restrict my sample to white females in the cross section sample with 2,317 individuals. I keep those who have received a high school diploma between 1997-2000. All women in the sample graduated from high school between 17 to 19, were never in the military, and were single with no children at graduation. Individuals with incomplete observations on schooling, employment, or marital choices are dropped from the sample. Finally I keep women with complete background information. Selected 537 individuals born between 1980-1983 are observed up to five years.

Figure 5 compares college attendance profiles between NLSY97 and NLSY79. Note that only four-year data are available for NLSY97 sample conditional on having enough number of observations. College enrollment increased by 19 percentage points, from 61% to 80% between

[27] This is consistent with Cameron and Taber (2004)'s finding that liquidity constraints have little impact on schooling attainment.

these two cohorts.[28] Is the estimated schooling model able to predict this change in college enrollment? To answer this question, I simulate the estimated model for the NLSY97 sample.

The observed changes over this period are (1) changes in background such as family income, parental education and cognitive ability; (2) changes in the schooling distribution of potential husbands; (3) changes in college premium for both males and females; (4) changes in direct cost of college. Table 14 compares the background of the NLSY79 sample and the NLSY97 sample[29]. Even though the NLSY97 sample have more siblings and are more likely from broken families, they have better background in terms of family income, parental education and cognitive ability. On average their parents obtain more than one year of schooling, family income is 13 thousand dollars higher, percentile cognitive score is 10 percentage points higher and they also graduate high school 0.1 years younger. I assume that the NLSY97 sample use the schooling distribution of 22-to-35-year old white males between 1997 and 2000 to predict their future husbands' schooling distribution. As shown in Table 15, college enrollment of white males increased by 5 percentage points.[30] Between 1980s to 2000s, skill premium for both males and females increased dramatically.[31] To estimate the changes in the college premium is complex. The returns to schooling for both males and females are estimated in the structural model to control for selection. Without a similar structural model estimated for the new cohort, we can not obtain a consistent estimate of the returns to schooling. I adopt a much more parsimonious method. As Figure 6 shows, the relative wage between some college and high school graduate females increased by 50%, while the relative wage between college graduate and high school graduate females doubled between the early 1980s and the early 2000s.[32] These premiums can be attributed to the returns to ability

[28]At the same time, labor force participation pattern stays the same. The young cohort tends to marry less or later. But if we take the cohabitation into account, the proportion of having a partner/spouse converge to the marriage profile of the old cohort. To consider cohabitation as a separate choice variable is left for future research.

[29]I use the same variable definitions except for cognitive ability percentile scores. For NLSY79, AFQT percentile score generated by the department of defense is presented. For NLSY97, however, ASVAB math and verbal percentile score generated by NLS is used. It is an age-adjusted, weighted average percentile score of four batteries from ASVAB: Mathematical Knowledge (MK), Arithmetic Reasoning (AR), Word Knowledge (WK), and Paragraph Comprehension (PC). The formula is similar to AFQT score and is the most comparable variable.

[30]CPS changed schooling classification in 1992. Prior to 1991, we have information on the number of grades attended and completed upto 18 years. After 1992, however, we only have information on an individual's highest degree received. I classify those who have some college but no degree as completed 13 years, those who have bachelors degree as completed 16 years, those who have masters degree as completed 17 years and those who have professional or doctorate degrees as completed 18 years.

[31]The increase in skill premium is well documented in the literature, see Katz and Murphy (1992), Card and DiNardo (2002) and Eckstein and Nagypál (2004).

[32]Similar pattern holds for men's college premium. Here, I only consider the effects of changes in female's college

or the returns to college (Taber 2001). I simply treat them as college premium to have an upper bound for the changes in college premium for the new cohort. I assume in the model, for the new cohort, the returns to each additional year of schooling (β_1's) increases by 50% and the returns to college graduation (β_4) doubles for all types. Figure 7 shows the time trend of changes in direct cost of college inclusive of tuition, room and board according to National Center for Education Statistics. The average total cost between 1980-1988 is used as the cost for NLSY79 sample and I use the average total cost between 1997-2003 to approximate the cost for NLSY97 sample.

Given these changes, I predict enrollment behavior of the NLSY97 sample using the estimated model based on the NLSY79 sample. In the first simulation, potential husbands' schooling distribution, the earning processes, and the cost of college are fixed at the levels for NLSY79 sample, and the NLSY97 sample's background variables are used. College enrollment would increase by 8 percentage points, from 61% to 69%. This increase is simply because women who have better background have a higher probability of being high skilled. In the second simulation, the females face potential husbands with better education, the model predicts female's college enrollment would increase by an additional 4 percentage points due to educational assortative mating. In the third simulation, females expect the dramatic increase in their own college premium. The college enrollment increases by 3 percentage points more. In the last simulation, the NLSY97 sample have to pay the new average cost of college (around \$11,030 in 2000 dollars). College enrollment would drop by 1 percentage point. The model, which is estimated based on a sample attending college in the early 1980s, can predict the enrollment behavior in the early 2000s. Since most school premium is conditional on college graduation, the effect of increasing college premium on college enrollment is relatively small but it has a large effect on college graduation. In fact, it can account for a 10 percentage points increase in the college graduation rate. Figure 8 depicts college attendance profiles conditional on each exogenous change. It will be very interesting to see if the model can predict the college graduation behavior well for the new cohort when data is available.

premium. Since available earnings data for husbands is not as good as the wage data for females, the specification of husband's earnings does not take graduation premium into account. When I use OLS estimates of men's earnings equation for both samples using CPS data, changes in men's earning process have negligible effects on women's college enrollment rate.

6 Simulations

6.1 How Much Does Marriage Matter to College Decision?

I run counterfactual simulations to study the effects of marriage on women's college decision. I compare women's schooling distribution from each simulation with the baseline given estimated parameters. Table 16 presents the simulation results.

The first simulation analyzes the case when women do not care about the relative schooling background of husbands. Setting $a_1 = a_2 = 0$, the model predicts no correlation between couples' education because matching is random. The only gain through the marriage market in this case is that college attendance increases marriage offer rate. Therefore we observe that women cluster at the level of some college. College enrollment would increase by 14.7 percentage points and graduation would drop by 6.9 percentage points. Type 1's have more incentive to attend college and type 2's have less tendency to graduate.

In the second simulation, I assume that college does not increase the marriage offer rate, i.e. $b_3 = 0$. College graduation rate increases slightly by 1.4 percentage points but college enrollment drops by 6.6 percentage points. Based on the type specific simulation, type 1's are the type who attend college for more marriage opportunities. If college has no effect on the marriage offer rate, their enrollment rate drops by half. The marriage offer rate has almost no effect on type 2's enrollment. In fact, setting b_3 to zero increases their college graduation rate simply because they are less likely to get married and drop out of college when the marriage offer rate is lower.

Women benefit from expected marriage from educational assortative mating and the marriage offer rate. When I zero out both benefits in the third simulation, college enrollment drops by around 3 percentage points, from 61% to 58% and the college graduation rate drops by 6 percentage points, from 38% to 32%. The drop in enrollment is mainly due to type 1's stopping going to college to meet more potential spouses. On the other hand, the fact that type 2's have less incentive to graduate to match their schooling with college graduate men attributes to the drop in graduation.

If the marriage option is not available altogether, the only incentive to attend college is to increase future earnings. Then the benefits of higher wages become more important in the college decision. Furthermore the estimated utility of schooling is negative and large when women are married. Even if women do not attend college to gain from a higher marriage offer rate and assortative matching, they tend to drop out college following marriage. Therefore when women are always single, they will invest more and stay longer in college. In the simulation where marriage offers are never received, enrollment would increase by 1.7 percentage points and graduation would increase by 11 percentage points. The effects are the strongest for type 2's, 99.8% of whom would graduate from college.

6.2 The Impact of the Return to Schooling

Table 17 shows the impact of the return to schooling.[33] With 10% increase in the return of each additional year of schooling (β_1's), the enrollment rate would increase by 0.2 percentage points and the graduation rate would increase by 0.3 percentage points. If the return of each additional year of schooling increases by 50%, college enrollment and graduation rates would increase by 1.5 percentage points and 1.8 percentage points, respectively. Enrollment increases are mainly from type 1's and graduation increases are mainly from type 2's. On the other hand, a 10% increase in returns to college graduation (β_4) would have almost no effect on enrollment and increase graduation by 0.4 percentage points. Even with 50% increase β_4, college enrollment would increase only by 0.2 percentage points and graduation would increase by 2 percentage points. These effects are due to the response of type 2's.

6.3 Education Policy Experiments

In Table 18, I present evidence on the impact of two policy interventions to increase educational attainment: college tuition subsidies and college graduation bonus. These education policy experiments assume the impact of policy-induced skill supply responses on equilibrium skill rental prices are negligible.[34]

[33] This exercise considers the wage elasticity of college enrollment. The wage elasticity of labor supply has been a topic of considerable interest in both labor and macro economics and it correlates with both marriage and schooling choices. In Van Der Klaauw (1996), marital status is a choice variable. Eckstein and Wolpin (1989) and Imai and Keane (2004) include post school human capital accumulation in a life cycle labor supply model.

[34] Two recent papers, Donghoon Lee (2005) and Heckman et al (1998), have made a start at developing solution and estimation methods that can account for the general equilibrium feedbacks. However, their results are very divergent.

College Tuition Subsidies I first simulate the effect of an experiment that provides a 50% tuition subsidy (a reduction in cs by 50%) for each year of college attendance. Average completed schooling level increases by 0.1 years, from 14.3 to 14.4 years. College attendance rate increases from 61% to 63% and graduation rate increases from 38% to 39%. Because college graduation is so prevalent among type 3's regardless of the subsidy, the increases in college graduation rates are mostly from type 2's. At the same time more type 1's attend college with the tuition subsidy.

Graduation Bonuses In contrast to tuition subsidies, which are based only on attendance, graduation bonuses rewards individuals for years of schooling that are completed. Graduation bonus schemes provide monetary payment for college graduation. In the second policy experiment, reported in panel (2) of Table 18, the effect of $5000 graduation bonus is presented. College attendance rate increases slightly by 0.2 percentage points and graduation rate increases from 38% to 40.3%. The low skill type 1's are not affected by the policy variation.

7 Concluding Remarks

In this paper, I have formulated and empirically implemented a structural dynamic model of high school graduate women's sequential decisions on college attendance, work, and marriage. The model is estimated on longitudinal data that includes information about school attendance, labor force participation, marital status, wages, and spousal characteristics. The estimates of the model are used to quantify the importance of alternative reasons for college attendance and graduation, and in particular, the estimates of the model are used to assess the effect of the expectations of marriage on college choice due to educational assortative mating and potential husband's income and due to the marriage offer rate.

The main results can be summarized as follows: First, marriage plays a significant role in a female's college choice. When the benefits from marriage are ruled out in the estimated model and everything else is kept the same, the predicted college enrollment drops by 3 percentage points, from 61 percent to 58 percent, and college graduation drops by 6 percentage points, from 38 percent to 32 percent. This prediction is for women graduating from high school in the early 1980s, as is the sample used to estimate the model. Second, the estimated model from the early

1980s does well in predicting college enrollment behavior in the early 2000s. College enrollment rates increased from 61 to 80 percent over this period. The model predicts the following: given (1) changes in family income, parental education and individual cognitive ability, (2) changes in potential husbands' schooling, (3) changes in female's college premium, (4) changes in the direct cost of college, college enrollment would increase by 8 percentage points, 4 percentage points, and 3 percentage points and decrease by 1 percentage point, respectively. The prediction for the new cohort is not only a validation of the model, it also provides evidence of the stability of the structural model for policy analysis.

The U.S. labor market has experienced some dramatic changes over the past few decades. First of all, female's college enrollment and graduation rates have been expanding constantly. At the same time, labor force participation rate of married females increased from 40 percent to 71 percent between 1964 and 2003. These two trends are consistent with each other because as women become more educated, the returns from working are higher. However, for cohorts born since the mid 1950s and the early 1960s, the women's college enrollment rate and graduation rate exceed those of men but their labor force participation is much lower than men's labor force participation, especially for married women. If the increase in earnings power were the only gain from investing in education and there were no discrimination towards females, we would not expect female's labor force participation rate to be much lower than male's. This paper provides a mechanism which is consistent with this puzzling fact. Suppose some women attend to college only to improve their future marriage, they would withdraw from the workplace following marriage. Therefore married women's labor force participation is low. An open question would be what is the socially optimal level of schooling when some people invest in education, but do not work.

References

[1] Becker (1973), "A Theory of Marriage: Part I" *J.P.E.* Vol. 81, Iss. 4: 813-846.

[2] Behrman, Jere R., Pollak, Robert A. and Taubman, Paul (1989), "Family Resources, Family Size, and Access to Financing for College Education", *J.P.E.* Vol. 97: 398-419.

[3] Brien, Michael J., Lillard, Lee A. and Stern, Steven (1999), "Cohabitation, Marriage, and Divorce in a Model of Match Quality", *International Economic Review*, forthcoming.

[4] Cameron, Stephen V., and Heckman, James J. (1993), "The Nonequivalence of High School Equivalents", *Journal of Labor Economics*, Vol. 11, No.1, Part 1: 1-47.

[5] —— (1998), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *J.P.E.* Vol. 106: 262-333.

[6] —— (2001), "The Dynamics of Educational Attainment for Black, Hispanic, and White Males" *J.P.E.* Vol. 109, no. 3: 455-499.

[7] Cameron, Stephen V., and Taber, Christopher (2004), "Estimation of Educational Borrowing Constraints Using Returns to Schooling", *J.P.E.* Vol. 112, no. 1: 132-182.

[8] Card, David (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, Vol. 69, No. 5: 1127-1160.

[9] Card, David, and DiNardo, John E. (2002), "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles", *NBER working paper 8769*.

[10] Carrasco, M. and Florens J.P. (2002), "Simulation-Based Method of Moments and Efficiency", *Journal of Business and Economic Statistics* 20(4): 482-92.

[11] Eckstein, Zvi, and Nagypál, Éva (2004), "The Evolution of U.S. Earnings Inequality: 1961-2002", *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 28, No. 2, December 2004: 10-29.

[12] Eckstein, Zvi, and Wolpin, Kenneth I. (1989), "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", *The Review of Economic Studies*, Vol. 56, No. 3: 375-390.

[13] —— (1999), "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities", *Econometrica* 67: 1295-1339.

[14] Gould, Eric D. (2003), "Marriage and Career: The Dynamic Decisions of Young Men", Mimeo, Hebrew University.

[15] Heckman, James J., Lochner, Lance, and Taber, Christopher (1998), "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents", *Review of Economic Dynamics*, Vol. 1, No. 1: 1-58.

[16] Heckman, James J. and Sedlacek, Guilherme (1985), "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *J.P.E.*, Vol. 93, No. 6: 1077-1125.

[17] Heckman, J. and Singer B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52(2), 271-320.

[18] Imai, Susumu, and Keane, Michael P. (2004), "Intertemporal Labor Supply and Human Capital Accumulation", *International Economic Review*, 45:2: 601-641.

[19] Katz, Lawrence, and Murphy, Kevin M. (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, 107: 35-78.

[20] Keane, Michael P., and Wolpin, Kenneth I. (1994), "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence", *Review of Economics and Statistics*, 76: 648-672.

[21] —— (1997), "The Career Decisions of Young Men", *J.P.E.*, Vol. 105, No. 3: 473-522.

[22] —— (2001), "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", *International Economic Review*, Vol. 42, No. 4: 1051-1103.

[23] Lee, Donghoon (2005), "An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice", *International Economic Review*, Vol. 46, No. 1: 1-34.

[24] MaFadden, Daniel (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", *Econometrica* 57: 995-1026.

[25] Mare, Robert D. (1991), "Five Decades of Educational Assortative Mating", *American Sociological Review*, Vol. 56, No. 1: 15-32.

[26] Mincer, Jacob (1974), *Schooling, Experience, and Earnings*. New York: NBER.

[27] Pakes, Ariel, and Pollard, David (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica* 57: 1027-1057.

[28] Pencavel, John (1998), "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands", *The American Economic Review*, Vol. 88, No. 2: 326-329.

[29] Roy, Andrew D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers* Vol. 3, No. 2: 135-146.

[30] Shimer, Robert and Smith, Lones (2000), "Assortative Matching and Search", *Econometrica*, Vol. 68, No. 2: 343-369.

[31] Taber, Christopher (2001), "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *The Review of Economic Studies*, Vol. 68, No. 3: 665-691.

[32] Van Der Klaauw, Wilbert (1996), "Female Labour Supply and Marital Status Decisions: A Life-Cycle Model", *The Review of Economic Studies*, 63, 199-235.

[33] Weiss, Yoram (1997),"The Formation and Dissolution of Families: Why Marry? Who Marries Whom? and What Happens upon Marriage and Divorce?", In *Handbook of Population Economics*, edited by Robert Rosenzweig and Oded Stark. Amsterdam Elsevier Science.

[34] Weiss, Yoram and Willis, Robert J. (1997), "Match Quality, New Information, and Marital Dissolution", *Journal of Labor Economics*, Vol. 15, No. 1: S293-S329.

[35] Willis, Robert J. and Rosen, Sherwin (1979), "Education and Self-Selection", *J.P.E.* Vol. 87, No.5: S7-S36.

[36] Wolpin, Kenneth I. (1996), "Public-Policy Uses of Discrete-Choice Dynamic Programming Models", *The American Economic Review*, Vol.86, No. 2: 427-432.

[37] —— (2003), "Wage Equations and Education Policy", in *Advances in Economics and Econometrics*, (ed. by M. Dewatripont, L.P. Hansen and S.J. Turnovsky), Cambridge, UK: Cambridge University Press.

[38] Wong, Linda (2003), "Structural Estimation of Marriage Models", *Journal of Labor Economics*, Vol. 21, No. 3: 699-727.

AppendixA: Proofs

Proof of Proposition 2:

The model can be solved backwards. At $t = 2$, alternative specific value functions conditional on female's schooling $S_i = k$ and male's schooling $S_i^H = j$ can be written as following:

$$\begin{aligned} V_{i2}(1; k, j) &= y(k, \delta_i) + \psi y^H(j) + a_0 - \gamma(k - j)^2, \\ V_{i2}(0; k) &= y(k, \delta_i), \quad k = 1, 2, j = 1, 2. \end{aligned}$$

$V_2(1)$ and $V_2(0)$ are values of being married and being single respectively. The problem is solved separately in three cases.

Case 1: If $\gamma < \psi y^H(1) + a_0$, every woman marries when an offer arrives so match is random.

At $t = 1$, value of not attending college is

$$\begin{aligned} V_{i1}(0) &= y(1, \delta_i) + \beta E \max[V_{i2}(0), V_{i2}(1)] \\ &= y(1, \delta_i) + \beta \left[\mu p^0 V_{i2}(1; 1, 2) + (1 - \mu) p^0 V_{i2}(1; 1, 1) + \left(1 - p^0\right) V_{i2}(0; 1)\right]. \end{aligned}$$

Similarly value of attending college is

$$\begin{aligned} V_{i1}(1) &= -cs + \beta E \max[V_{i2}(0), V_{i2}(1)] \\ &= -cs + \beta \left[\mu p^1 V_{i2}(1; 2, 2) + (1 - \mu) p^1 V_{i2}(1; 2, 1) + \left(1 - p^1\right) V_{i2}(0; 2)\right]. \end{aligned}$$

Individual i attends college if and only if $V_{i1}(1) \geq V_{i1}(0)$. Assume $\beta_1 > \ln(\frac{1+\beta}{\beta})$, then attending college is the dominant strategy if and only if $\delta_i \geq \delta_1^* = \beta_0^{-1} \left(\ln \frac{A_1}{e^{\beta_1}[\beta e^{\beta_1} - (1+\beta)]}\right)$, where

$$A_1 = cs + \beta \left(p^0 - p^1\right) \left[\mu \psi y^H(2) + (1 - \mu) \psi y^H(1) + a_0\right] + \beta[(1 - \mu) p^1 - \mu p^0]\gamma.$$

All individuals with ability above threshold δ_1^* choose to attend college. Therefore a fraction $\Phi_1 = \Phi(\delta_1^*)$ of women are high school graduates and the rest are college graduates.

It is straightforward to show that $\partial A_1/\partial(cs) > 0$, $\partial A_1/\partial\mu < 0$, that is, college enrollment decreases in the cost of college and increases in the number of college men available. Furthermore

$$\frac{\partial A_1}{\partial p^1} = -\beta\{\mu\psi y^H(2) + \mu a_0 + (1-\mu)[\psi y^H(1) + a_0 - \gamma]\} < 0,$$

so college enrollment also increases in p^1. But the sign of $\partial A_1/\partial\gamma = \beta[(1-\mu)p^1 - \mu p^0]$ is ambiguous.

Case 2: If $\psi y^H(1) + a_0 < \gamma < \psi y^H(2) + a_0$, high school graduate women accept all marriage offers but college graduate women marry only if offers are from college men. We can derive similar threshold condition in ability δ_2^*, such that individuals attend college if $\delta_i \geq \delta_2^*$. $\delta_2^* = \beta_0^{-1}\left(\ln \frac{A_2}{e^{\beta_1}[\beta e^{\beta_1} - (1+\beta)]}\right)$, where

$$A_2 = cs + \beta\left(p^0 - p^1\right)\mu[\psi y^H(2) + a_0] + \beta(1-\mu)p^0[\psi y^H(1) + a_0] - \beta\mu p^0\gamma.$$

Therefore $\partial A_2/\partial\mu < 0$, $\partial A_2/\partial p^1 < 0$ and $\partial A_2/\partial\gamma < 0$. That is, college enrollment increases if there are more college men available, if college enhances access to marriage offers more, and if women care more about the homogeneity in schooling.

Case 3: If $\gamma > \psi y^H(2) + a_0$, whenever there is a miss match in education, the disutility is overwhelming. High school women only marry high school men and college women only marry college men so the sorting is perfect. The threshold ability for college is $\delta_3^* = \beta_0^{-1}\left(\ln \frac{A_3}{e^{\beta_1}[\beta e^{\beta_1} - (1+\beta)]}\right)$, where

$$A_3 = cs + \beta\{(1-\mu)p^0[\psi y^H(1) + a_0] - \mu p^1[\psi y^H(2) + a_0]\}.$$

And again $\partial A_3/\partial\mu < 0$, $\partial A_3/\partial p^1 < 0$.

Appendix B: Data Construction

Recall that in the model each period is a year. This characterization of the decision process implies that some of the data must be aggregated to match the model. The details of the data construction follow.

Timing: I follow each woman in the model after she graduates from high school. A year in the model is defined as a school year from September to August. Suppose a woman received her high school diploma in June 1985, the first year corresponds to calendar month September 1985 to August 1986.

Schooling: In order to construct the annual school attendance, I first derive monthly attendance on the basis of a question concerning whether the youth enrolled in regular school in each month of the previous year. This question started in 1981. I thus have individual's monthly schooling status from January 1980. I treat a woman as an attendee if she reported having attended school for at least 6 months in the school year[35]. Questions on month and year respondents receive high school diploma are used to determine the graduation date. Combined this date with respondent's date of birth, her age at graduation is computed.

Employment and wage: NLSY79 workhistory records weekly hours worked for each week since the beginning of 1978. Annual hours worked is based on accumulating weekly hours worked over the school year. A woman in the model is defined as employed if her working hours are reported at least in 26 weeks of the year, and annual hours worked at least 1000 hours.

The employment history information is employer-based. All references to a "job" should be understood as references to an employer. The variable "hourly rate of pay job #1-5" in the work history file provides the hourly wage rate for each job. The associated wage on multiple jobs held is the average and data are constructed that maximum number of jobs held in a year is five. I use coded real hourly wage in 2000 dollars. Nominal wage data are deflated by CPI from BLS CPI-U. The hourly wages are top coded at $300 and bottom coded at $1.

Marital status and fertility: Month/year in which the first, the second and the third marriage began and month/year in which the first and the second marriage ended are recorded in NLSY79. I aggregate monthly marital status into annual status according to the following: an individual is

[35]For simplicity, I do not consider measurement error on choices.

defined as married in a year if she is married for at least 6 months in the year. This definition of marriage does not include those who cohabit. Detailed cohabitation information is not available in NLSY79 until the 1990 survey and the decision to cohabit is quite different from the decision to marry (see Brien et al 1999). Cohabitation is not treated as a separate choice to limit the state space. Based on a question about the birth date of the first child born to NLSY79 respondents, the fertility history on the first child can be constructed. For simplicity I follow the birth of the first child only and ignore child mortality.

Spouses' characteristics: NLSY ask every year how much respondent's spouse received from wages, salary, commissions or tips from all jobs before deductions for taxes or anything else. I use this question to construct husbands' annual earnings and they are converted to real income in 2000 dollars. NLSY household roster provides each family member's highest grade completed and their relationship to the youth respondent. I first obtain the spouse's household number, then link it to corresponding family member's characteristics such as age and highest grades completed.

Appendix C: Inputs of the Model

Tabel C1: Logit Estimates of the Arrival Probability of theFirst Child

Coefficient	Estimates (Std. Err.)
c_0 : Constant	-12.385 (3.460)
c_1 : Education	-0.343 (0.026)
c_2 : Last period's marital status	1.461 (0.126)
c_3 : Age	1.041 (0.291)
c_4 : Age^2	-0.018 (0.006)
c_5 : Marriage duration	0.310 (0.031)

Table C2: Potential Husbands' Schooling Distribution

Years of schooling	11 or less	12	13	14	15	16	17	18 or more
Percentage	6.88	41.69	8.65	11.03	5.30	16.16	3.53	6.77

Appendix D: Parameter Estimates

Parameters		Parameters		Parameters	
Utility function		Type proportions		Marriage value	
α_1	1.158 (1.96e-3)	λ_0^2	-6.729 (1.00e-2)	a_0^1	-7.878e+3 (1.32e+1)
α_2	-0.025 (4.14e-5)	λ_1^2	0.410 (1.90e-3)	a_0^2	-1.111e+3 (1.90)
α_3	-0.016 (2.38e-5)	λ_2^2	0.120 (1.99e-4)	a_0^3	5.953e+4 (2.35e+2)
α_4	-0.004 (5.94e-6)	λ_3^2	-0.565 (8.20e-5)	a_1^1	-6.060e+3 (2.94e+1)
v_1^1	-6.154e+4 (2.16e+2)	λ_4^2	1.077 (2.17e-3)	a_1^2	-4.908e+3 (1.43e+1)
v_1^2	3.952e+4 (6.11e+1)	λ_5^2	0.063 (7.70e-5)	a_1^3	-3.305e+3 (5.89)
v_1^3	6.255e+4 (8.84e+1)	λ_6^2	0.066 (2.14e-4)	a_2	0.841e+3 (1.96)
v_2^1	-1.125e+5 (2.72e+2)	λ_7^2	-0.188 (5.09e-4)	a_3	1.396e+4 (4.58e+1)
v_2^2	2.537e+4 (4.05e+1)	λ_0^3	-7.150 (9.47e-3)	a_4	5.154e+3 (7.27)
v_2^3	1.095e+5 (2.34e+2)	λ_1^3	0.240 (6.20e-4)		
v_3^1	-5.154e+5 (1.02e+3)	λ_2^3	0.108 (1.94e-4)	Marriage offer	
v_3^2	-2.778e+5 (3.63e+2)	λ_3^3	-0.115 (1.76e-4)	b_0^1	-6.185 (1.16e-2)
v_3^3	-2.060e+4 (5.99e+1)	λ_4^3	-0.318 (6.05e-4)	b_0^2	-12.53 (1.76e-2)
v_4^1	-6.041e+5 (1.83e+3)	λ_5^3	0.026 (3.60e-5)	b_0^3	-7.870 (2.48e-2)
v_4^2	-6.268e+5 (1.90e+3)	λ_6^3	0.033 (4.55e-5)	b_1	0.219 (3.31e-4)
v_4^3	-2.290e+4 (3.58e+1)	λ_7^3	0.017 (2.86e-5)	b_2	-0.368e-3 (1.20e-6)
v_5^1	3.490e+3 (5.12)	π_1^1	0.732 (2.43e-3)	b_3	0.640 (1.02e-3)
v_5^2	9.956e+3 (1.99e+1)	π_1^2	0.260 (2.41e-3)	Husband's earning	
v_5^3	8.116e+3 (1.35e+1)	π_2^1	1.000 (2.97e-6)	ρ_0	9.379 (1.23e-2)
v_6^1	1.432e+4 (4.72e+1)	π_2^2	0.000 (2.11e-6)	ρ_1	0.043 (1.10e-4)
v_6^2	2.481e+4 (3.59e+1)	π_3^1	0.659 (5.72e-4)	ρ_2	0.058 (8.50e-5)
v_6^3	9.622e+3 (2.91e+1)	π_3^2	0.137 (2.82e-4)	ρ_3	-0.144e-2 (2.70e-6)
		Earnings		σ_{y^H}	0.550 (9.32e-4)
		β_0^1	1.132 (1.80e-3)	σ_{μ_y}	0.158 (3.24e-4)
Budget constraint		β_0^2	1.217 (2.42e-3)	Shocks	
cg	3.555e+4 (9.71e+1)	β_0^3	1.193 (2.23e-3)	σ_1	2.879e+4 (6.10e+1)
cc	4.237e+4 (1.31e+2)	β_1^1	0.042 (5.97e-5)	σ_2	1.302e+4 (3.80e+1)
$\psi(0)$	0.324 (1.04e-3)	β_1^2	0.055 (6.90e-5)	σ_3	1.122e+4 (1.86e+1)
$\psi(1)$	0.484 (1.51e-3)	β_1^3	0.062 (8.60e-5)	σ_4	1.298e+4 (2.07e+1)
		β_2	0.101 (2.05e-4)	σ_5	4.129e+4 (1.21e+2)
		β_3	-0.751e-3 (1.17e-6)	σ_6	1.684e+5 (3.22e+2)
Job offers		β_4	0.296 (8.63e-4)	σ_7	7.642e+4 (1.17e+2)
p_{hg}^0	0.772 (2.38e-3)	β_{0c}^1	2.362 (6.67e-3)	σ_8	2.159e+5 (7.73e+2)
p_{sc}^0	0.760 (1.34e-3)	β_{0c}^2	1.929 (2.27e-3)	End condition	
p_{cg}^0	0.707 (2.20e-3)	β_{0c}^3	2.214 (6.85e-3)	δ_1	2.484e+4 (9.33e+1)
p_{hg}^1	0.957 (6.56e-6)	σ_w	0.366 (7.05e-4)	δ_2	2.715e+4 (6.55e+1)
p_{sc}^1	0.999 (1.88e-6)	σ_{wc}	0.113 (2.09e-4)	δ_3	-0.265e+2 (3.99e-2)
p_{cg}^1	1.000 (2.63e-7)	σ_u	0.165 (4.53e-4)	δ_4	2.662e+4 (9.20e+1)

Table 1: Descriptive Statistics

Variable	Mean	Standard deviation	Number of observations
	Sample of 487 individuals		
Years in sample	9.79	1.09	487
Age at high school graduation	17.89	0.44	487
Highest grade completed (HGC)	14.31	2.39	487
Years of total experience	6.73	2.66	487
Marriage duration in years	3.84	3.25	487
	Sample of 345 at the first marriages		
Women's age	23.06	2.64	345
Men's age	26.08	4.23	345
Women's HGC	13.56	1.87	345
Men's HGC	13.28	2.27	345
	Sample of 4,770 person-year observations		
Age	23.34	2.90	4770
Attend school (percent)	24	42	4770
Work (percent)	69	46	4770
Married (percent)	38	49	4770
Child (percent)	22	41	4770
Hourly wage*	9.88	8.27	3126
Husband's annual earnings*	34,896	41,124	1858

* In 2000 dollars.

Table 2: Choice Proportions by Years After High School

Year	No. Obs	NNS	ANS	NWS	AWS	NNM	ANM	NWM	AWM
1	(487)	15.2	37.6	30.6	10.5	3.5	0.2	2.3	0.2
2	(486)	9.5	31.7	32.7	11.5	4.1	0.4	9.5	0.6
3	(485)	8.0	26.8	33.0	10.5	6.6	1.9	13.0	0.2
4	(481)	6.2	21.0	33.9	10.8	7.5	1.5	18.5	0.6
5	(478)	4.2	6.7	44.6	7.9	8.4	1.5	25.1	1.7
6	(475)	4.0	4.4	42.1	5.7	10.7	1.1	30.9	1.1
7	(472)	4.0	3.0	38.6	3.8	13.6	0.8	33.9	2.3
8	(470)	3.2	1.9	32.8	4.0	14.0	1.3	40.4	2.3
9	(469)	3.2	1.9	30.1	3.4	16.2	1.5	41.2	2.6
10	(467)	3.0	0.9	27.0	3.4	18.6	1.9	42.6	2.6

Note:
NNS denotes not-attend, not-work, single; ANS denotes attend, not-work, single;
NWS denotes not-attend, work, single; AWS denotes attend, work, single;
NNM denotes not-attend, not-work, married; ANM denotes attend, not-work, married;
NWM denotes not-attend, work, married; AWM denotes attend, work, married.

Table 3: Mean Transitions

From\To	Attend	Not-Attend
Attend	66.45	33.55
Not-Attend	5.31	94.69
From\To	Work	Non-employed
Work	88.76	11.24
Non-employed	34.77	65.23
From\To	Single	Married
Single	87.71	12.29
Married	3.63	96.37

Table 4: Attendance and Marriage Probits

	Attendance probit	Marriage probit
Constant	1.949 (0.198)	-3.480 (0.191)
Age	-0.069 (0.009)	0.132 (0.008)
Participation	-1.242 (0.052)	-0.090 (0.055)
Presence of children	-1.044 (0.089)	1.236 (0.056)
Marital status	-0.629 (0.062)	
Attendance		-0.749 (0.064)
Log likelihood (No. of obs.)	-1863.0 (4770)	-2334.4 (4770)
Likelihood Ratio χ^2	1485.2	1684.1

Table 5: Employment Probits

	All women	Married women	Married women
	(1)	(2)	(3)
Constant	1.396 (0.265)	2.440 (0.439)	2.235 (0.464)
Experience	0.616 (0.034)	0.499 (0.052)	0.457 (0.056)
Experience squared	-0.041 (0.004)	-0.025 (0.006)	-0.021 (0.006)
Age	-0.125 (0.016)	-0.170 (0.022)	-0.161 (0.024)
Schooling	0.128 (0.017)	0.114 (0.023)	0.128 (0.025)
Attendance	-1.275 (0.057)	-0.834 (0.136)	-0.820 (0.144)
Presence of children	-0.967 (0.065)	-0.961 (0.075)	-0.980 (0.078)
Marital status	-0.322 (0.059)		
Husband's earnings (thousands)			-0.003 (0.0008)
Log likelihood (No. of obs.)	-2079.8 (4,770)	-870.9 (1,831)	-805.1 (1,711)
Likelihood Ratio χ^2	1771.0	489.6	433.7

Table 6: Mean and Deciles of Women's Hourly Wage

Year	Mean (no. of obs.)	Wage Deciles 10%	20%	30%	40%	50%	60%	70%	80%	90%
1	6.17 (175)	4.64	5.33	5.87	5.98	6.24	6.39	6.51	6.90	7.54
2	6.81 (248)	5.02	5.76	5.96	6.15	6.38	6.76	7.22	7.73	8.83
3	6.93(262)	5.12	5.79	5.98	6.36	6.69	7.14	7.50	8.27	9.53
4	7.59(295)	4.97	5.61	6.07	6.62	7.13	7.60	8.44	9.47	10.72
5	8.01(359)	5.26	5.73	6.38	6.96	7.46	8.16	9.05	10.26	11.89
6	10.01(367)	5.84	6.56	7.31	8.03	8.80	9.47	10.52	12.30	14.62
7	11.38(358)	6.28	7.46	8.39	9.17	10.19	11.20	12.57	14.40	17.28
8	12.66(367)	6.06	7.38	8.33	9.30	10.61	11.83	13.68	15.59	18.90
9	12.65(350)	6.23	7.66	8.47	9.46	11.04	12.37	14.30	15.96	19.36
10	12.70(345)	6.25	7.75	8.85	9.88	11.49	12.91	14.80	17.11	19.58

Table 7: Assortative Mating in Education at the First Marriage

Married Women's Schooling	Husband's Schooling: HS or less	Some College	College Graduates
HS Graduates	77.7	15.7	6.6
Some College	42.9	38.5	18.7
College Graduates	19.5	20.7	59.8
Correlation in Years of Schooling: 0.55			

Table 8: Background and Outcomes

	No. of Obs.	HGC	% HS graduate	% some college	% college graduate	Mean hourly wage (std. err.) at 25-28	% married at 25
All	487	14.3	38.6	23.6	37.8	12.7 (0.7)	52.0
Mother's Schooling:							
Non-high school graduate	100	12.9	64.0	24.0	12.0	11.3 (1.0)	58.0
High school graduate	267	14.2	39.7	25.5	34.8	12.1 (0.7)	53.9
Some college	60	15.2	21.7	28.3	50.0	12.4 (0.7)	53.3
Collge graduate	60	16.3	8.3	10.0	81.7	17.9 (3.7)	31.7
Father's Schooling:							
Non-high school graduate	114	13.2	59.7	21.0	19.3	10.3 (0.8)	64.9
High school graduate	205	13.9	44.4	26.8	28.8	12.0 (0.7)	53.2
Some college	64	14.8	23.4	31.3	45.3	11.3 (0.6)	56.3
College graduate	104	16.1	13.5	15.4	71.1	17.4 (2.5)	32.7
Number of Siblings:							
0	17	14.7	35.3	17.7	47.0	17.3 (4.9)	23.5
1	94	14.6	33.0	25.5	41.5	12.4 (0.8)	48.9
2	144	14.5	36.1	24.3	39.6	12.7 (0.8)	49.3
3	104	14.5	35.6	21.1	43.3	13.7 (2.2)	54.8
4+	128	13.7	48.4	24.2	27.4	11.5 (1.3)	58.6
Household Structure at 14:							
Live with both parents	68	13.9	41.2	33.8	25.0	15.8 (3.3)	48.5
Not live with both parents	419	14.4	38.2	22.0	39.8	12.2 (0.5)	52.5
Net Family Income:							
$Y <= 1/2$median	40	13.8	47.5	25.0	27.5	12.0 (2.4)	60.0
$1/2$median$< Y <=$median	204	13.9	44.1	29.4	26.5	12.8 (1.3)	56.4
median$< Y <= 2$median	210	14.6	33.8	19.5	46.7	12.6 (0.8)	46.7
$Y > 2$median	33	15.6	24.2	12.1	63.6	13.6 (1.3)	48.5
AFQT Percentile Score							
AFQT<=20	48	12.5	77.1	18.7	4.2	9.8 (1.1)	58.3
20<AFQT<=50	173	13.3	57.2	26.0	16.8	11.5 (1.1)	57.8
50<AFQT<=80	191	14.9	25.1	25.7	49.2	12.2 (0.5)	52.9
AFQT>80	75	16.3	5.3	16.0	78.7	18.5 (3.1)	32.0
Age at High School Graduation							
17	76	15.0	28.9	21.1	50.0	14.3 (1.5)	57.9
18	389	14.2	39.6	23.6	36.8	12.6 (0.8)	51.2
19	22	13.2	54.6	31.8	13.6	9.1 (0.8)	45.5

Table 9: Chi-Square Goodness-of-Fit Tests of the Within-Sample Choice Distribution

	Choices								
Year	NNS	ANS	NWS	AWS	NNM	ANM	NWM	AWM	χ^2 Row
1	0.71	0.73	0.13	0.00	1.09	0.01	0.10	0.01	2.77
2	0.01	0.06	0.21	0.02	0.00	0.04	0.18	0.01	0.52
3	0.02	0.04	0.15	0.01	0.03	1.27	0.01	0.46	1.99
4	0.00	0.10	0.04	0.04	0.19	0.25	0.02	0.24	0.87
5	0.69	0.86	0.55	0.01	0.60	0.12	0.12	0.18	3.13
6	0.24	0.01	0.07	0.02	0.01	0.07	0.02	0.01	0.44
7	0.01	0.06	0.01	0.45	0.05	0.26	0.02	0.92	1.78
8	0.17	0.19	0.11	0.22	0.04	0.04	0.01	0.44	1.21
9	0.27	0.79	0.00	0.41	0.06	0.02	0.09	0.78	2.42
10	0.01	0.39	0.07	0.61	0.06	0.01	0.02	0.97	2.13

Note: $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed frequency for bin i and E_i is the expected frequency for bin i, $\chi^2_7(0.05) = 14.07$.

Table 10: Fit of the Mean Transitions

From\To	Attend	Not-Attend
Attend	63.23 (66.45)	36.77 (33.55)
Not-Attend	5.97 (5.31)	94.03 (94.69)

From\To	Work	Not Work
Work	83.47 (88.76)	16.53 (11.24)
Not Work	46.60 (34.77)	53.40 (65.23)

From\To	Single	Married
Single	87.31 (87.71)	12.69 (12.29)
Married	4.75 (3.63)	95.25 (96.37)

Note: Data moments are in parentheses.

Table 11: Predicted Matching in Education at The First Marriage

Married Women's	Husbands' Schooling		
Schooling	HS Graduates	Some College	College Graduates
HS Graduates	69.6 (77.7)	27.2 (15.7)	3.2 (6.6)
Some College	44.7 (42.9)	36.4 (38.5)	18.9 (18.7)
College Graduates	11.9 (19.5)	32.5 (20.7)	55.6 (59.8)

Note: Data moments are in parentheses.

Table 12: Selected Characteristics in the 10th Year After Graduation by Skill Type

	Skill Type One	Skill Type Two	Skill Type Three
Sample proportions	49.6	40.6	9.8
Proportions of			
High school graduates	73.8	4.7	0.0
Some college	26.2	23.5	2.2
College graduates	0.0	71.8	97.8
Years of			
Schooling	12.3	15.6	19.9
Experience (after school)	7.8	5.1	1.7
Marriage duration	3.7	4.0	3.3
Proportion who			
Attend school	0.3	6.8	78.8
Work	75.6	74.4	86.0
Marry	62.6	70.8	43.0
Mean hourly wage ($2000)	11.7	17.3	12.0

Note: Sample proportions are based on 5,000 simulations and other characteristics are based on simulations for 5,000 individuals of each type.

Table 13: Relationship of Selected Family Background Characteristics to Skill Types

	% Skill Type 1	% Skill Type 2	% Skill Type 3
All	49.6	40.6	9.8
Mother's Schooling:			
Non-high school graduate	69.4	20.3	10.3
High school graduate	51.9	35.1	13.0
Some college	37.1	49.2	13.7
Collge graduate	19.3	68.2	12.5
Father's Schooling:			
Non-high school graduate	59.7	29.5	10.8
High school graduate	51.0	36.1	12.9
Some college	45.8	40.2	14.0
College graduate	37.1	47.1	15.8
Number of Siblings:			
0	20.1	72.4	7.5
1	29.3	61.1	9.6
2	40.2	47.9	11.9
3	51.6	35.0	13.4
4+	61.6	24.0	14.4
Household Structure at 14:			
Live with both parents	48.6	38.5	12.9
Not live with both parents	61.2	16.6	22.2
Net Family Income:			
$Y <= 1/2$median	55.0	31.9	13.1
1/2median$< Y <=$median	52.0	34.8	13.2
median$< Y <= 2$median	47.5	39.3	13.2
$Y > 2$median	38.0	49.3	12.7
AFQT Percentile Score			
AFQT<=20	88.6	5.1	6.3
20<AFQT<=50	70.9	18.5	10.6
50<AFQT<=80	33.5	53.9	12.6
AFQT>80	10.8	80.2	9.0
Age at High School Graduation			
17	46.3	41.5	12.2
18	49.6	37.0	13.3
19	53.0	32.6	14.4

Note: Results are based on 5,000 simulations.

Table 14: Background Differencs: NLSY79 v.s. NLSY97

Variable Name	NLSY79	NLSY97
Highest grade completed of mother at 14	12.3 (0.09)	13.6 (0.10)
Highest grade completed of father at 14	12.6 (0.13)	13.8 (0.12)
Number of siblings at 14	2.8 (0.08)	3.4 (0.10)
Broken home at 14	0.14 (0.01)	0.16 (0.02)
Family income (in thousands 2000 dollars)	65.3 (1.50)	78.5 (2.68)
AFQT score	53.9 (1.08)	63.5 (1.00)
Age at high school graduation	17.9 (0.02)	17.8 (0.02)

Note: Standard errors of the means are in parentheses

Table 15: Schooling Distribution of NLSY79 and NLSY97 Sample's Potential Husbands

Cohort\Yrs of school	11 or less	12	13	14	15	16	17	18 or more
NLSY79	6.88	41.69	8.65	11.03	5.30	16.16	3.53	6.77
NLSY97	6.40	37.34	21.75	4.79	3.89	20.48	3.46	1.89

Note: statistics are based on 22 to 35 years old white males whose years of schooling are at least 10 years from CPS 1980-1983 and 1997-2000.

Table 16: The Impact of Marriage Expectations on Education Outcome by Skill Types

	All	Type 1	Type 2	Type 3
Baseline Model				
Mean HGC	14.3	12.3	15.6	19.9
% HS Graduate	39.0	73.8	4.7	0.0
% Some College	23.0	26.2	23.5	2.2
% College Graduate	38.0	0	71.8	97.8
(1) No Educational Assortative Mating ($a_1 = a_2 = 0$)				
Mean HGC	14.3	12.6	15.2	19.4
% HS Graduate	24.3	45.4	3.7	0
% Some College	44.6	54.6	43.5	1.8
% College Graduate	31.1	0	52.8	98.2
(2) College Does Not Increase Marriage Offers ($b_3 = 0$)				
Mean HGC	14.3	12.1	15.7	20.3
% HS Graduate	45.6	86.9	5.3	0.0
% Some College	15.0	13.1	19.2	2.4
% College Graduate	39.4	0	75.5	97.6
(3) Both (2) and (3) Hold ($a_1 = a_2 = 0,\ b_3 = 0$)				
Mean HGC	14.2	12.2	15.3	19.7
% HS Graduate	42.5	80.8	5.6	0
% Some College	25.3	19.2	38.1	1.6
% College Graduate	32.2	0	56.3	98.4
(4) No Marriage Offers ($\Pr_t = 0$)				
Mean HGC	15.0	12.3	16.9	21.8
% HS Graduate	37.3	74.5	0	0
% Some College	13.7	25.5	0.2	0
% College Graduate	49.0	0	99.8	100

Table 17: The Impact of Returns to Schooling on Education Outcome by Skill Types

	All	Type 1	Type 2	Type 3
Baseline Model				
Mean HGC	14.3	12.3	15.6	19.9
% HS Graduate	39.0	73.8	4.7	0.0
% Some College	23.0	26.2	23.5	2.2
% College Graduate	38.0	0	71.8	97.8
(1) 10% Increase in Return to One Year of Schooling (β_1's)				
Mean HGC	14.3	12.3	15.8	19.8
% HS Graduate	38.8	73.5	4.6	0.0
% Some College	22.9	26.5	22.8	2.3
% College Graduate	38.3	0	72.6	97.7
(2) 50% Increase in Return to One Year of Schooling (β_1's)				
Mean HGC	14.3	12.3	15.6	18.9
% HS Graduate	37.5	71.4	4.2	0.0
% Some College	22.7	28.6	19.0	2.7
% College Graduate	39.8	0	76.8	97.3
(3) 10% Increase in Return to College Graduation (β_4)				
Mean HGC	14.3	12.3	15.6	19.9
% HS Graduate	38.9	73.8	4.7	0.0
% Some College	22.7	26.2	22.4	2.2
% College Graduate	38.4	0	72.9	97.8
(4) 50% Increase in Return to College Graduation (β_4)				
Mean HGC	14.3	12.3	15.7	19.8
% HS Graduate	38.8	73.8	4.4	0.0
% Some College	21.2	26.2	18.5	2.1
% College Graduate	40.0	0	77.1	97.9

Table 18: Education Policy Experiments

	All	Type 1	Type 2	Type 3
Baseline Model				
Mean HGC	14.3	12.3	15.6	19.9
% HS Graduate	39.0	73.8	4.7	0.0
% Some College	23.0	26.2	23.5	2.2
% College Graduate	38.0	0	71.8	97.8
(1) 50% College Tuition Subsidy				
Mean HGC	14.4	12.3	15.7	20.1
% HS Graduate	37.1	70.3	4.3	0
% Some College	23.9	29.7	21.3	1.9
% College Graduate	39.0	0	74.4	98.1
(2) $5000 Graduation Bonus				
Mean HGC	14.3	12.3	15.6	19.9
% HS Graduate	38.8	73.8	4.3	0
% Some College	20.9	26.2	18.0	2.1
% College Graduate	40.3	0	77.7	97.9

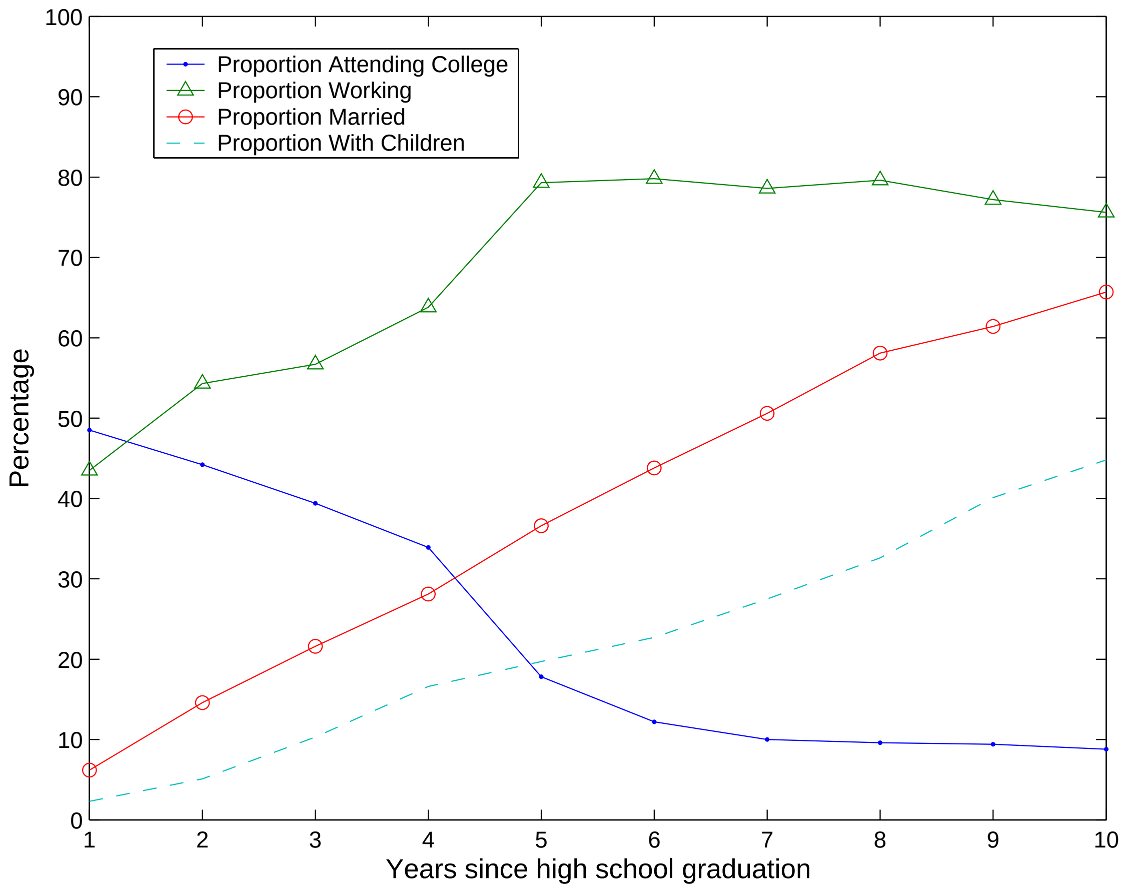


Figure 1: Proportions Attending College, Working, Married and Having Children

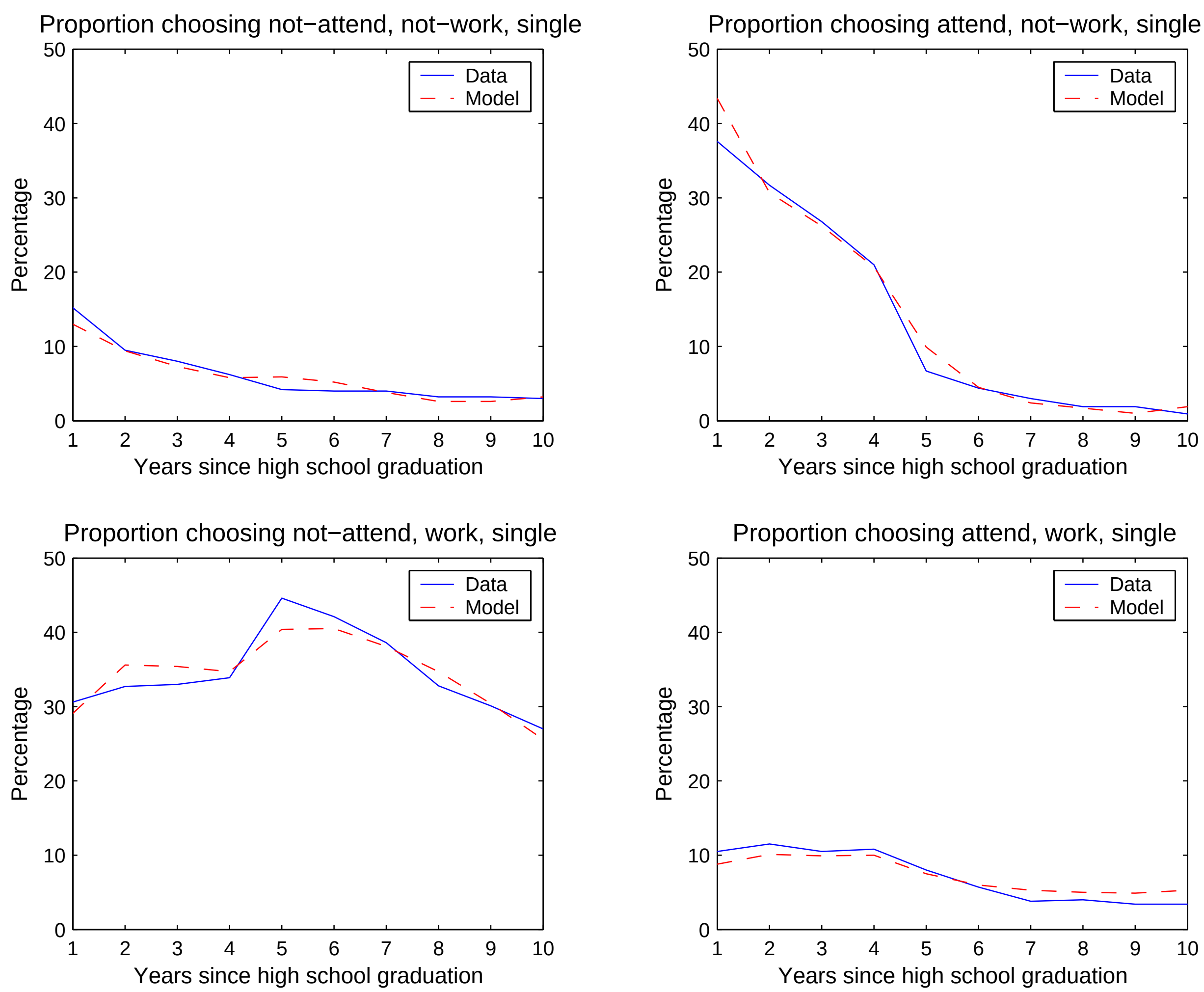


Figure 2 (a): Fit of Choice Proportions When Single

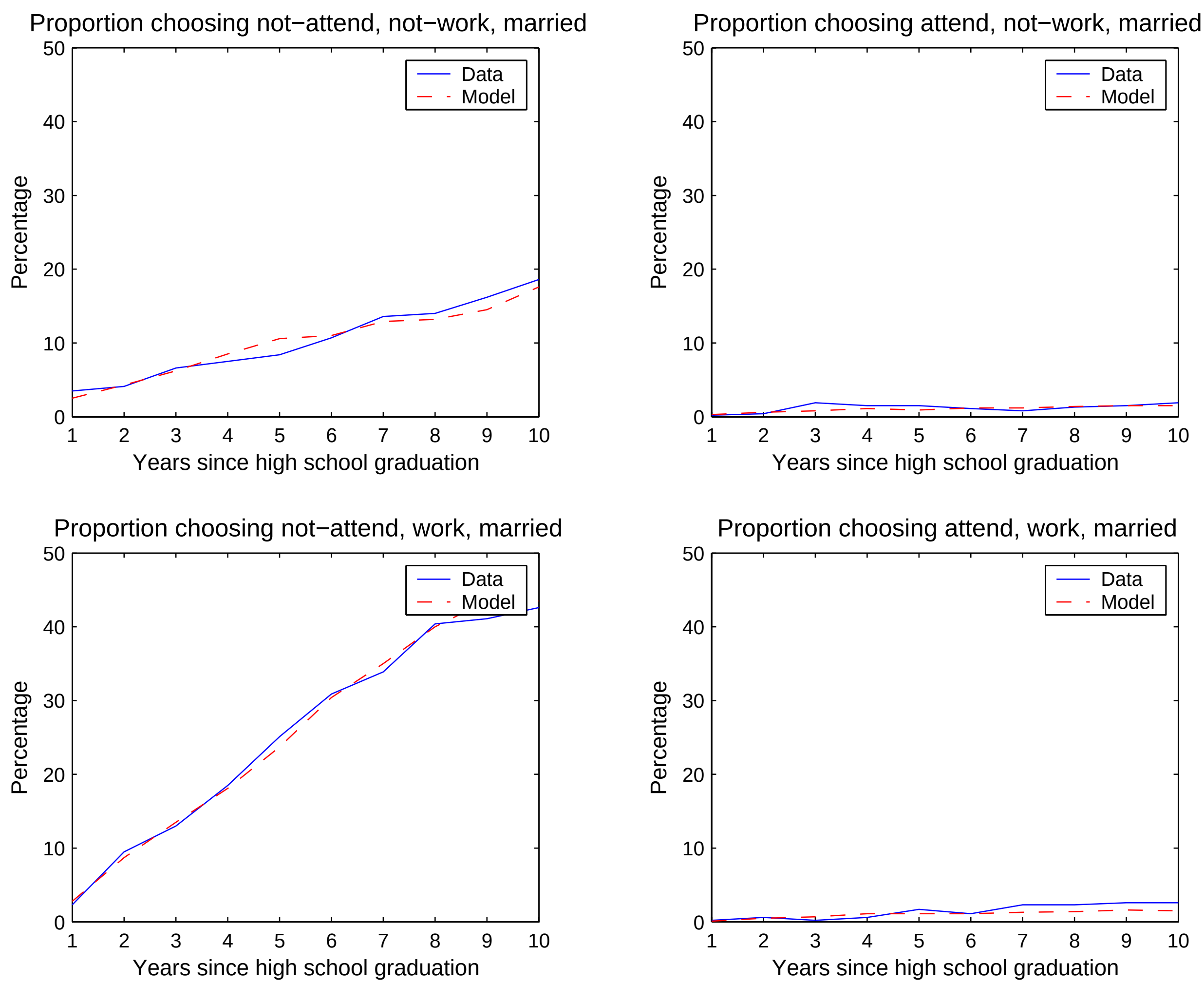


Figure 2 (b): Fit of Choice Proportions When Married

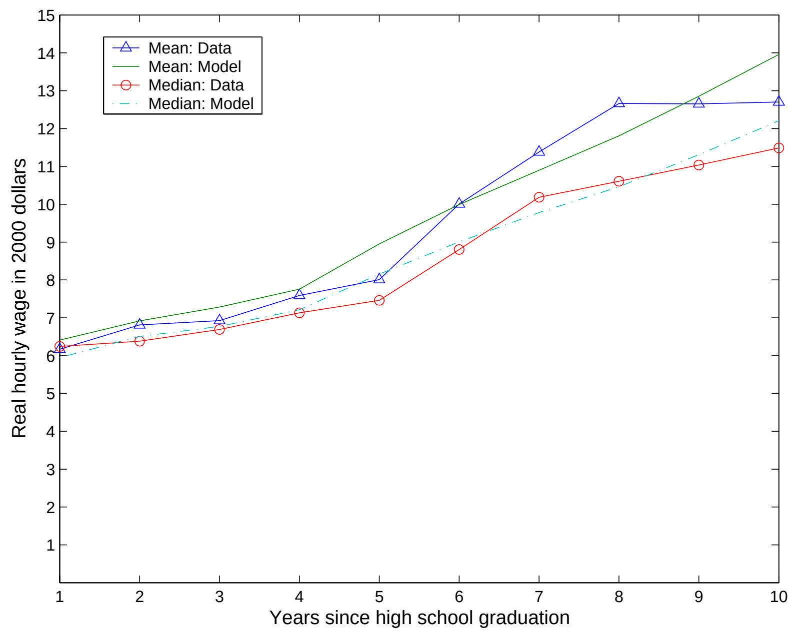


Figure 3 (a): Fit of Women's Hourly Wage: Mean and Median

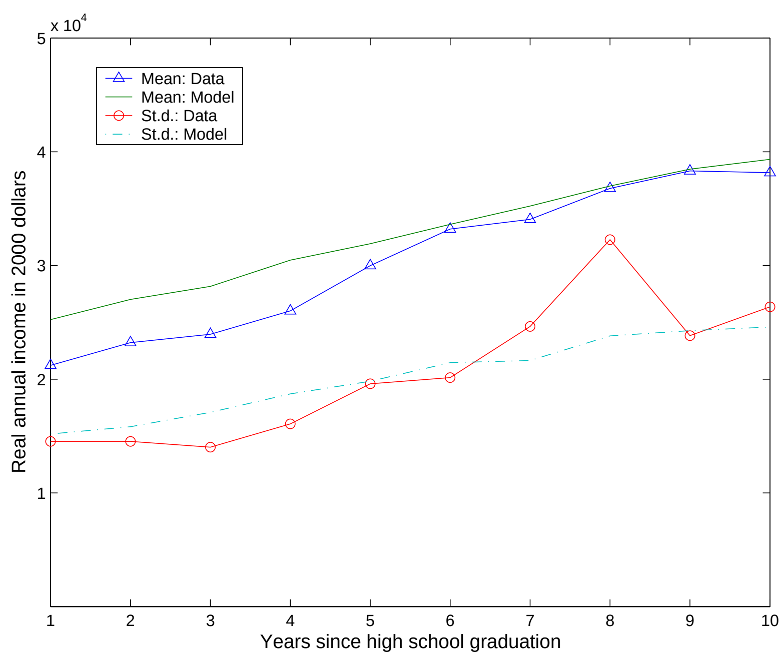


Figure 3 (b): Fit of Husband's Annual Income: Mean and Standard Deviation

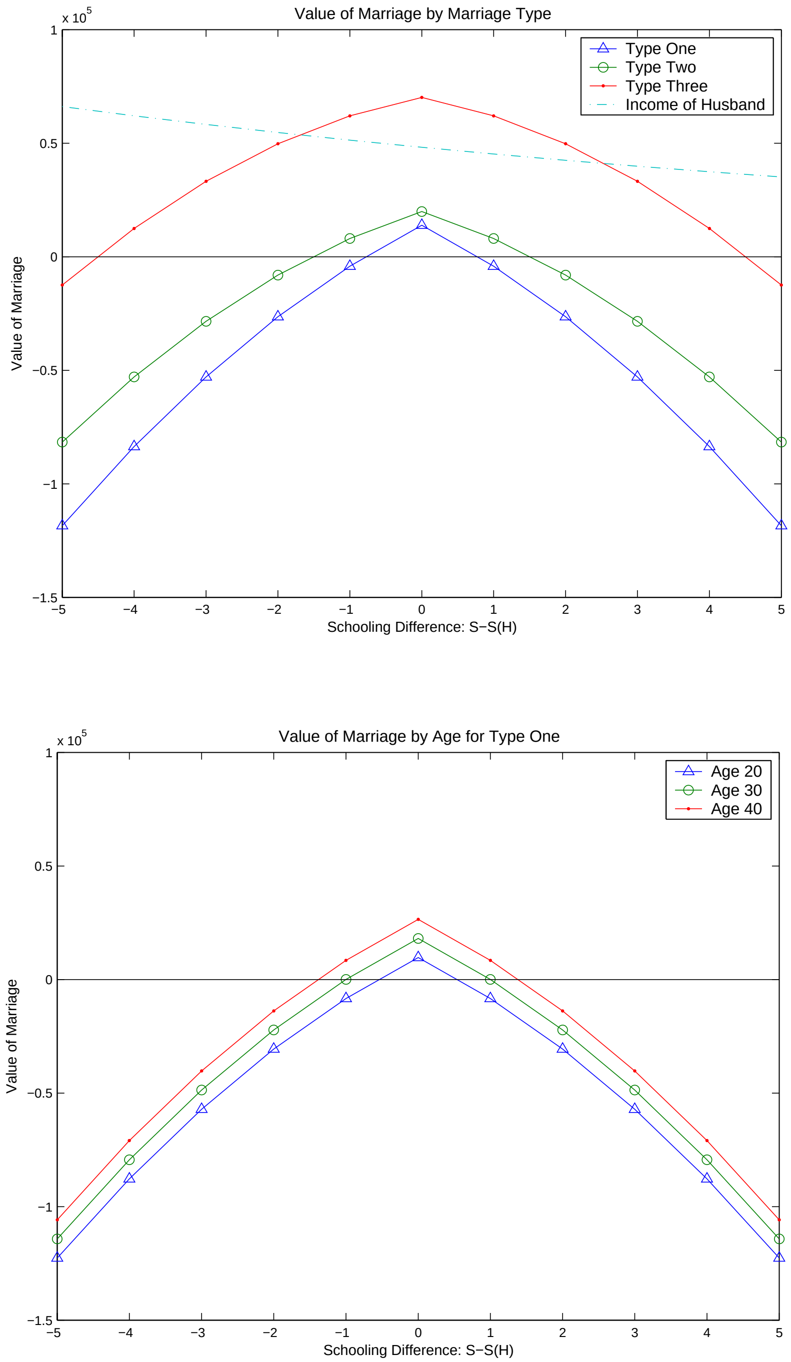


Figure 4: Value of Marriage Conditional on Schooling Differences by Marriage Type and by Age

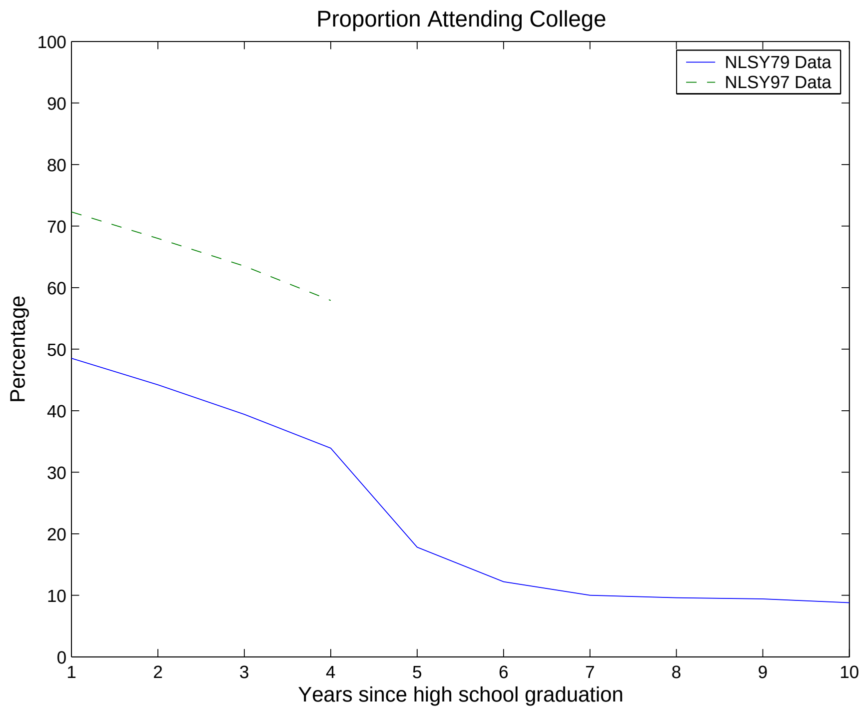


Figure 5: College Enrollment: NLSY79 and NLSY97 Samples

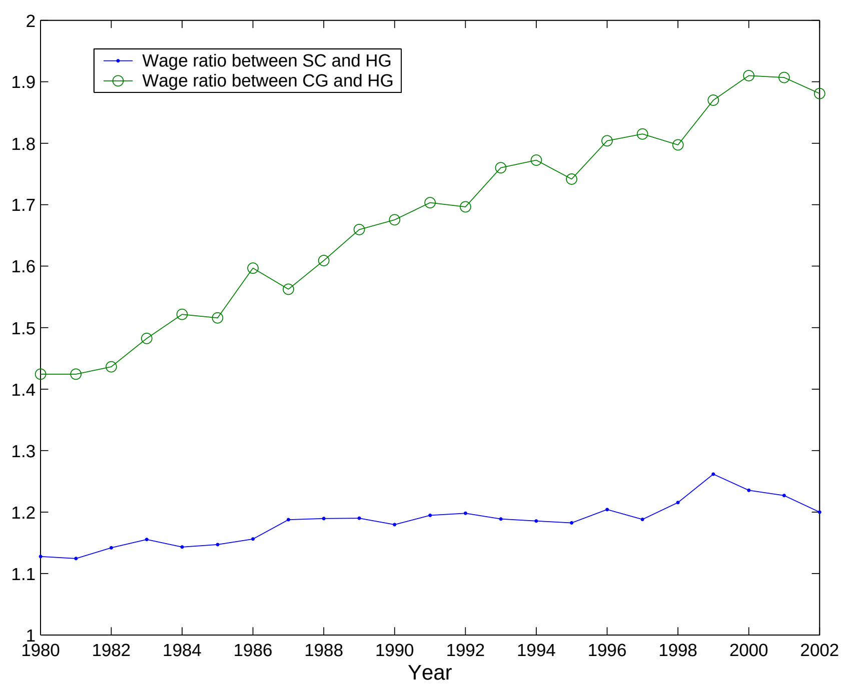


Figure 6: Relative Wages of White Females from CPS 1980-2002

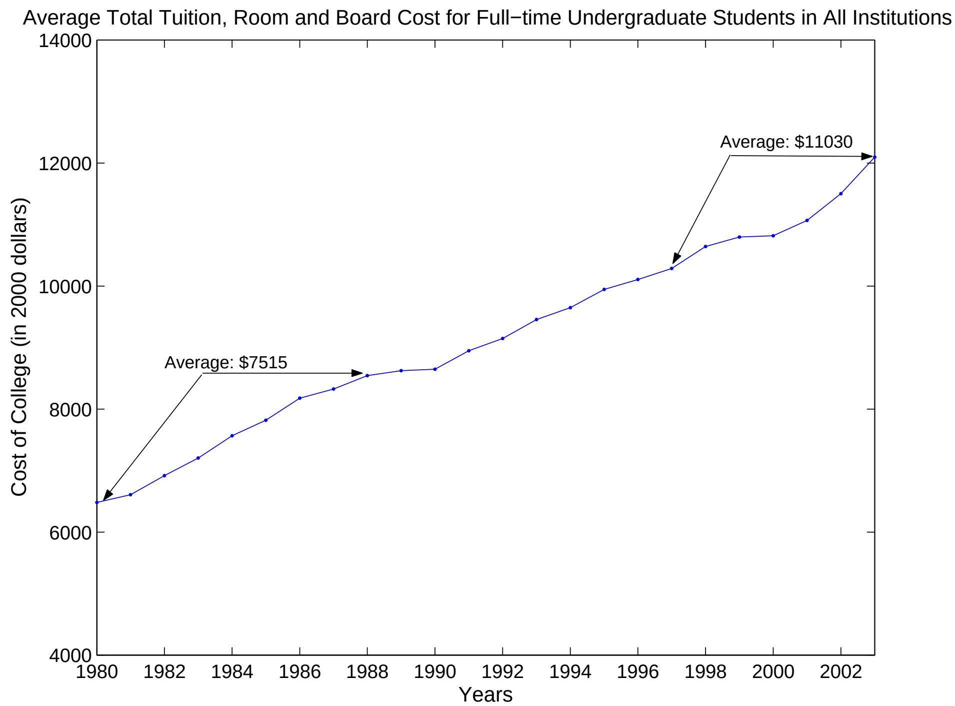


Figure 7: Changes in Direct Cost of College

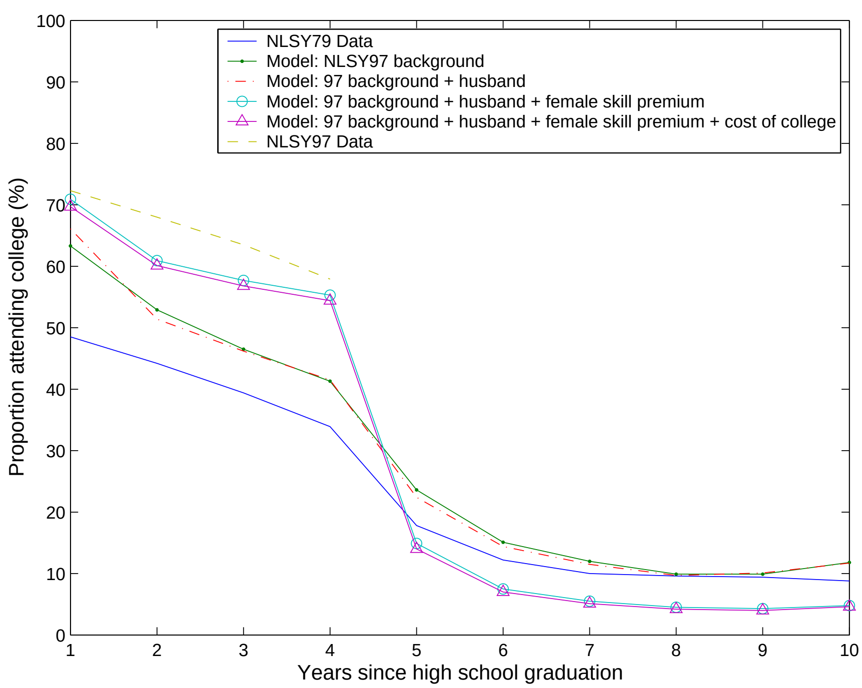


Figure 8: Predictions for NLSY97 Sample

An Anatomy of International Trade: Evidence from French Firms

Jonathan Eaton,* Samuel Kortum,† and Francis Kramarz‡

October 2005

Abstract

We develop an equilibrium model of worldwide competition across a range of goods. Our model encompasses Ricardian and monopolistic competition as special cases. We parameterize the model to gauge its ability to capture the export behavior of French manufacturing firms.

Key words: International trade, exporting, market penetration

*Department of Economics, New York University, and National Bureau of Economic Research (jonathan.eaton@nyu.edu).

†Department of Economics, University of Minnesota, Federal Reserve Bank of Minneapolis, and National Bureau of Economic Research (kortum@econ.umn.edu).

‡CREST-INSEE and CEPR (kramarz@ensae.fr).

1 Introduction

A literature that emerged over the last 10 years has exploited microlevel data to measure various features of the export behavior of individual producers. Bernard and Jensen (1995, 1999), Roberts and Tybout (1997), Clerides, Lach, and Tybout (1998, henceforth CLT), Aw, Chung, and Roberts (1998), and Hallward-Driemeier, Iarossi, and Sokoloff (2002) using data from various countries, have shown that producers who export are typically in the minority and tend to be more productive and much larger, even in terms of their domestic sales; yet they usually export only a small fraction of their output.

All of these characteristics suggest that individual producers face substantial hurdles in entering foreign markets. Several theories have emerged in response to these observations. Bernard, Eaton, Jensen, and Kortum (2003) (Henceforth BEJK) develop a Ricardian model of plant-level export behavior while Melitz (2003) and Chaney (2005) provide models based on monopolistic competition. Essential to either explanation are trade barriers that deter many producers who sell at home from entering foreign markets.

Until recently, our ability to gauge a producer's export activity has been limited to observations on how much it exported. We have been in the dark about how exports broke down into sales in individual destinations. Yet this information is critical in understanding the nature of the barriers that individual producers face in selling abroad. In particular, the existing evidence raises three questions that our analysis here seeks to answer: (1) Is the major hurdle to exporting selling beyond the home market, with broad penetration across foreign markets among producers that do export, (as implicit in Roberts and Tybout, CLT, and Melitz), or do exporters appear to incur such costs market by market (as in BEJK and Chaney)? (2)

In either case, does the cost appear to be fixed (as in CLT and Melitz), or increasing in the amount shipped, as with standard "iceberg" transport costs? (3) What market structure does the evidence favor?

Our work makes use of an extensive source of data that provides some insight into the answers to these questions. The starting point is a comprehensive data set of French firms that has been merged with customs data on the value of each firm's shipments to individual destinations (see, Biscourp and Kramarz, 2002). Focusing on manufacturing firms in 1986, these data reveal enormous heterogeneity across both destinations and across firms in the nature of entry into different markets. Nevertheless, we observe some striking regularities. Looking across firms, the size and productivity advantages of exporters extend very seamlessly into size and productivity advantages of firms that export widely, and to less popular destinations. Looking across destinations, the number of French sellers to a destination increases with overall French market share with an elasticity close to one, while the number increases with market size with an elasticity of around two-thirds (see, Eaton, Kortum, and Kramarz, 2004).

We develop a model of firm competition across a wide number of markets that incorporates a fixed cost of entering an individual market as well as the standard iceberg costs that rise in proportion to the amount shipped. Both are needed to explain the increase in the number of sellers with market size and the dominance of home sales even among exporters. Our model nests the Ricardian framework of BEJK and the monopolistic competition (MC) approach of Melitz and Chaney by introducing the range of possible goods as a parameter of the model. When the range is small relative to the number of active producers, the model implies that multiple producers are competing head to head in different markets of the world, as in BEJK. A large range, however, implies that it is very unlikely that a producer faces a direct competitor anywhere (another firm able to cover the fixed cost of entry there), so monopolistic competition prevails.

We estimate the parameters of the model using data on aggregate production and bilateral trade among France and 112 trade partners as well as moments from our firm level dataset. Our estimation proceeds in two stages.

We first show that, under a simple deterministic formulation of the model with monopolistic competition, several parameter values can be calibrated very directly by observing (1) the relationship between the number of producers selling to at least some given number of destinations and the average sales in France of these producers and (2) average sales per market and the size of the market. The parameterized model fits these relationships tightly and provides ballpark estimates of several parameters close to values delivered by a more sophisticated model. It fails in two dimensions, however. First, it implies a much less heterogeneous sales distribution in any individual market than the data exhibit. Second, it predicts a strict hierarchy of export destinations: That is, a firm selling to the k'th most popular export destination is predicted to sell to the first through $k-1$'st as well. There are substantial deviations from such a hierarchy in the data.

We then estimate the parameters of a richer model that incorporates both an endogenous range of goods (so that monopolistic competition is only a special case) and firm and market specific shocks to demand and to the fixed cost of entry. We estimate the model by simulated method of moments. That is, given a vector of parameters, we use the model to simulate a population of French firms and their export behavior around the world. We then search over parameter values to make the moments of our simulated dataset match key moments of the actual data.

The outline of our paper is as follows. In the next section we describe our data in detail. Section 3 then presents our theoretical framework. In Section 4 we show how a simple version of the model lines up with some systematic features of the data, and delivers some estimates of the parameters. Section 5 describes our simulation approach and the results of our estimation of the parameters by simulated method of moments.

2 The Data

Our analysis uses both aggregate and firm level data. At the firm level we analyze the sales of 234,300 French manufacturing firms in 113 destinations around the world (including France itself). At the aggregate level we use data on bilateral trade flows in manufactures among these 113 countries, including home sales.

Our firm level data are constructed as follows: We merge data from two French administrative sources. The first is a collection of records of the universe of firms subject to the standard tax system. After additions and controls made at INSEE, the data include all balance-sheet variables, employment, industry affiliation, total sales, and a firm identifier (the Siren identifier). Second, French Customs compile all sales of French firms (also indicating their Siren identifier) in over 200 foreign destinations. Biscourp and Kramarz (2002) provide a thorough description of the two sources.

While the data cover all private sector firms, our focus is on a cross section of manufacturing firms from 1986, yielding a sample of 234,300. Since we lack other data (in particular, on domestic production) from many of the smaller destinations, we limit ourselves to 113 destination countries (including France). (Since the entities which we eliminated are very small, they constitute a trivial proportion of France's total export activity.)

As is typically the case, summing across what individual producers report exporting produces a number that is less than what is reported at the aggregate level. In the French case missing exports arise because manufacturing firms sell to nonmanufacturing intermediaries who report the foreign sales, and the connection between producer and destination is lost. Across all destinations, the firm data fail to account for about 20 per cent of total manufacturing exports.[1]

While the raw data themselves are confidential and housed at INSEE, we can construct a rich set of statistics from them. Some of these statistics do not rely on individual export destinations, so can be compared with the analogous statistics from producers located elsewhere (and, in particular, to U.S. producers). Statistics based on individual destinations, however, are to our knowledge unique to these data, providing a new window on the connections between firms and where they sell.

Previous work on the export behavior of individual producers has typically used the plant as the unit of observation. The French data report exports and other features by firm. Obviously differences arise. A firm might own several plants, for example, while a firm might exist that does not own any production unit that corresponds with the definition of a plant. *A priori*, a case can be made for either unit of observation over the other. A firm, for exam-

[1] This figure compares with underreporting of about 40 per cent in the U.S. Census of Manufactures. See Bernard and Jensen (1999) for a discussion.

ple, might own several plants with very diverse characteristics. Hence firm-level observations might mask a great deal of the variation in the plant-level data. But observations at the plant level may fail to pick up inputs provided by the headquarters.

3 Theory

Our theory is about competition across N geographically separated destinations in selling a good j, where there are a continuum of possible goods with measure J. In our quantitative analysis, of course, $N = 113$, while the range of potential goods J is a parameter that we estimate.

3.1 Technology

Our model of technology is adapted from Eaton and Kortum (2002) and BEJK (2003). The most efficient potential producer of good j in country i can produce and amount $z_i(j)$ per unit of input, where $z_i(j)$ is the realization of a random variable Z_i drawn from the distribution:

$$F_i(z) = \Pr[Z_i \leq z] = \exp[-(T_i/J)z^{-\theta}],$$

where $T_i > 0$ and $\theta > 1$ are parameters. The parameter θ, which we treat as common across countries and goods, governs the extent of heterogeneity in efficiency, with larger values of θ implying less heterogeneity. The parameter T_i governs the average level of efficiency in country i. It may appear that the measure of goods J should be simply subsumed into T_i, since the ratio is all that matters for the probability distribution of Z_i. A distinct role for J emerges, however, when we look across all the goods in the economy. The measure of goods that can be produced in country i with efficiency greater than z is $J\left\{1 - \exp[-(T_i/J)z^{-\theta}\right\}$.

The cost of an input unit in i is w_i while it requires shipping $d_{ni} \geq 1$ units of a good from i to deliver one unit in n. We normalize $d_{ii} = 1$ for all i. The unit cost of delivering a unit of good j in n from i is thus

$$c_{ni}(j) = \frac{w_i d_{ni}}{z_i(j)}.$$

The lowest cost version of good j in market n costs:

$$c_n(j) = \min\{c_{n1}(j), c_{n2}(j), ..., c_{nN}(j)\}.$$

3.2 The Distribution of Costs

Our distributional assumptions about Z imply that $c_{ni}(j)$ is the realization of a random variable C_{ni} drawn from the distribution:

$$\Pr[C_{ni} \leq c] = 1 - \exp[-(T_i/J)(w_i d_{ni})^{-\theta} c^{\theta}].$$

while $c_n(j)$ is the realization of a random variable C_i drawn from the distribution:

$$\Pr[C_n \leq c] = 1 - \exp[-(\Phi_n/J)c^{\theta}] \tag{1}$$

where:

$$\Phi_n = \sum_{i=1}^{N} T_i (w_i d_{ni})^{-\theta}.$$

The measure of goods that are potentially supplied to country n at a cost less than c is:

$$\mu_n(c) = J\{1 - \exp[-(\Phi_n/J)c^{\theta}]\}. \tag{2}$$

The probability that country i is the low cost supplier of good j to n is:

$$\pi_{ni} = \frac{T_i (w_i d_{ni})^{-\theta}}{\Phi_n}. \tag{3}$$

In addition to the unit cost, we introduce a fixed cost $E_n(j)$ to a firm of selling good j in market n. We assume that this fixed cost can be decomposed into a country component that applies to all goods and a component that varies across goods j. Hence we can write:

$$E_n(j) = E_n \varepsilon_n(j),$$

where $\varepsilon_n(j)$ is the realization of a random variable ε, which we treat as independent of any producer's efficiency $Z_i(j)$.

3.3 Demand and Market Structure

Our assumptions about demand are very standard. Expenditure on good j in market n, available at price $p_n(j)$ there, is:

$$X_n(j) = \alpha_n(j) \left(\frac{p_n(j)}{P_n} \right)^{1-\sigma} X_n \tag{4}$$

where X_n is total spending and P_n the CES price index:

$$P_n = \left[\int_0^J \alpha_n(j) p_n(j)^{1-\sigma} dj \right]^{1/(1-\sigma)}.$$

where σ is the elasticity of substitution. We restrict $\sigma \in (1, \theta + 1)$.[2]

The term $\alpha_n(j)$ is the realization of a random variable α that is also independent of any producer's efficiency $Z_i(j)$, but may be correlated with ε. We treat goods that are not sold in country n as having an infinite price.

Our assumptions are compatible with a broad range of market structures, at least over certain ranges of parameters. For concreteness we assume here that at most only the lowest

[2] We need $\sigma > 1$ both to get a well defined price for the case in which a producer faces no direct competition and to explain why the range of goods provided in different locations can differ. The restriction that $\sigma < \theta + 1$ is needed to ensure that the price index, derived below, is well defined.

unit cost supplier of good j to market n enters, with the fixed cost of entry deterring entry by others. Hence this supplier, conditional on entry, has a monopoly in market n, so charges a markup over unit cost of:

$$\overline{m} = \frac{\sigma}{\sigma - 1}.$$

Hence its unit price is:

$$p_n(j) = \overline{m}c_n(j)$$

if good j is sold in country n at all.

To summarize our assumptions, a potential producer has three characteristics: (1) the country i of its location, (2) the good j it knows how to make, and (3) its efficiency $z_i(j)$ making good j at location i. In turn, each good $j \in [0, J]$ has $2N$ characteristics: (1) the good-specific component of the entry barrier in each location $\varepsilon_n(j)$ and (2) the good-specific shock to demand in each location $\alpha_n(j)$. Each location i is distinguished by (1) its measure of ideas T_i, (2) its input cost w_i, (3) the component of the entry barrier that is common across goods E_i, and its geography relative to other locations reflected in the geographic barriers d_{ni}. The remaining parameters of the model are the range of goods J and the parameter θ, which governs heterogeneity in technology, and σ, which governs heterogeneity in preferences. We now turn to the determination of equilibrium.

3.4 Equilibrium Entry

The lowest cost supplier, conditional on entry, earns a profit, gross of the fixed cost, of $X_n(j)/\sigma$. Conditional on being the low cost supplier of good j in market n, a producer enters that market if:

$$X_n(j) \geq \sigma E_n \varepsilon_n(j)$$

or if:

$$\eta_n(j) x_n \geq \left(\frac{\overline{m} c_n(j)}{P_n} \right)^{\sigma-1} \tag{5}$$

where:

$$\eta_n(j) = \frac{\alpha_n(j)}{\varepsilon_n(j)}$$

and

$$x_n = \frac{X_n}{\sigma E_n}.$$

Note that η is a positive shock to entry. For any $\eta_n(j) = \eta$, condition (5) for entry determines a cutoff cost $\overline{c}_n(\eta)$ such that only a supplier with $c_n(j) \leq \overline{c}_n(\eta)$ would enter, where

$$\overline{c}_n(\eta) = \eta^{1/(\sigma-1)} \overline{c}_n \tag{6}$$

and:

$$\overline{c}_n = x_n^{1/(\sigma-1)} \frac{P_n}{\overline{m}}.$$

Integrating across the range of costs in any location n, the price index is:

$$\begin{aligned} P_n &= \overline{m} \left\{ E_\eta \left[\int_0^{\overline{c}_n(\eta)} E[\alpha|\eta] c^{1-\sigma} d\mu_n(c) \right] \right\}^{1/(1-\sigma)} \\ &= \overline{m} \left\{ E_\eta \left[E[\alpha|\eta] \int_0^{\eta^{1/(\sigma-1)}\overline{c}_n} c^{\theta-\sigma} \exp[-(\Phi_n/J)c^\theta] \theta \Phi_n dc \right] \right\}^{1/(1-\sigma)} \end{aligned} \tag{7}$$

(since we treat α and η as independent of z, and hence c).

Equation (6) defines a positive relationship between $\overline{c}_n$ and P_n while equation (7) defines a negative one. Together they determine $\overline{c}_n$ and P_n.

To solve for each we define the variable:

$$\widetilde{P}_n = \left(\frac{P_n}{\overline{m}}\right)^{1-\sigma} \Phi_n^{-1/\widetilde{\theta}} \tag{8}$$

where

$$\widetilde{\theta} = \frac{\theta}{\sigma - 1}.$$

The $\widetilde{P}_n$ solves:

$$\widetilde{P}_n = E_\eta \left[E[\alpha|\eta] J^{1-1/\widetilde{\theta}} \Gamma \left(1 - 1/\widetilde{\theta}, \frac{\left(\eta x_n / \widetilde{P}_n\right)^{\widetilde{\theta}}}{J} \right) \right], \tag{9}$$

where $\Gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$ is the incomplete gamma function.[3] Using this new term, the price index can then be written:

$$P_n = \overline{m} \left[\widetilde{P}(x_n)\right]^{-1/(\sigma-1)} \Phi_n^{-1/\theta}$$

while the number of entrants is:

$$J_n = J \left(1 - E_\eta \left[\exp\left\{ - \left(\eta x_n / \widetilde{P}(x_n)\right)^{\widetilde{\theta}} / J \right\} \right] \right). \tag{10}$$

Suppliers to market n have heterogeneous costs, and, from (3) above, a supplier from country i is more likely to sell in country n the larger π_{ni}. But, conditional on entry, suppliers from all countries have the same cost distribution in n and, given the constant markup, have the same distribution of prices and, hence, of sales. An implication is that the probability π_{ni} that a firm from i is the supplier of some particular good j, is also the fraction of spending by

[3] Combining (6) and (7), applying the change of variable $s = \Phi_n c^\theta$, and rearranging gives:

$$\widetilde{P}_n = E_\eta \left[E[\alpha|\eta] \int_0^{\left(\eta x_n / \widetilde{P}_n\right)^{\widetilde{\theta}}} s^{-1/\widetilde{\theta}} \exp(-s/J) ds \right].$$

Applying the change of variable $t = s/J$ we obtain (9).

country n on goods from country i. We can thus relate π_{ni} to data on import shares, that is:

$$\pi_{ni} = \frac{X_{ni}}{X_n} \tag{11}$$

where X_{ni} is n's purchases from i.

Since, conditional on entry, suppliers have the same cost distribution in market n regardless of their origin, the measure of firms from source i in market n, J_{ni}, should equal a fraction π_{ni} of the total number, so that:

$$J_n = \frac{J_{ni}}{\pi_{ni}}.$$

We use this relationship to infer the total number of sellers to a market from the number of French firms selling there and French market share.

3.5 Two Special Cases

Before turning to the general solution we consider two special cases close to those in the existing literature.

3.5.1 Pure Ricardian Competition

Say that $E_n = 0$ and $J = 1$ as in EK (2002). Since there is no entry barrier, the cutoff is infinite while the price index is:

$$P_n = \overline{m}\,[E(\alpha)]^{-1/(\sigma-1)}\,[\Gamma(1 - 1/\widetilde{\theta})]^{-1/(\sigma-1)}\Phi_n^{-1/\theta}$$

which, setting $\overline{m} = 1$ (since they assume perfect competition), reduces to the expression in EK (2002). Only the lowest unit cost supplier of good j to market n sells there.

3.5.2 Monopolistic Competition

Let $J \to \infty$. We then get a price index:

$$P_n = \overline{m}\left(1 - 1/\widetilde{\theta}\right)^{1/\theta} a_1^{-1/\theta} x_n^{-(1-1/\widetilde{\theta})/(\sigma-1)} \Phi_n^{-1/\theta} \tag{12}$$

where:

$$a_1 = \left\{E_\eta\left[E[\alpha|\eta]\eta^{\widetilde{\theta}-1}\right]\right\}.$$

The cutoff is

$$\overline{c}_n = \left(1 - 1/\widetilde{\theta}\right)^{-1/\theta}\left(\frac{\Phi_n}{a_1 x_n}\right)^{-1/\theta}. \tag{13}$$

Taking the measure (2) as $J \to \infty$, the measure of entrants with cost less than or equal to c is Pareto with parameters Φ_n and θ:

$$\lim_{J\to\infty} \mu_n(c) = \Phi_n c^\theta. \tag{14}$$

Taking the limit of (10) as $J \to \infty$ the measure of entrants is:

$$J_n = \left(1 - 1/\widetilde{\theta}\right) x_n E[\eta^{\widetilde{\theta}}]. \tag{15}$$

which rises in proportion to x_n.

If, in addition, we shut down market-specific sales and entry shocks by setting $\alpha_n(j) = \varepsilon_n(j) = 1\ \forall n,\ j$, our formulation is monopolistic competition with potential sellers having a Pareto distribution of efficiencies, as in Chaney (2005). Firms in any source are identical except for their efficiencies z. An implication is that there is a hierarchy of destinations.

Setting $a_1 = 1$ in (13) above, a necessary and sufficient condition for a firm from i to sell in market n is that it have a domestic cost below:

$$\overline{c}_{ni} = \overline{c}_n/d_{ni}$$

For each source i we can rank destinations n according to $\overline{c}_{ni}$, where $\overline{c}_i^{(1)} \geq \overline{c}_i^{(2)} \geq \overline{c}_i^{(3)} \geq ... \geq \overline{c}_i^{(k)} \geq ... \geq \overline{c}_i^{(N)}$. Hence any firm that sells to the k'th ranked market has a domestic cost c below $\overline{c}_i^{(k)}$ which is also below $\overline{c}_i^{(k')}$ for all $k' < k$. Hence it must sell to these markets as well. Hence in this special case each source i should have a hierarchy of destinations, with more efficient firms selling to destinations further down the hierarchy.

This special case yields simple specifications for (1) the distribution of sales in any market, (2) the number of firms entering a market, and (3) the relationship between a firm's sales in any particular market and the number of markets where it sells.

1. **The sales distribution.** To sell in market n a firm must sell at least σE_n to overcome the entry hurdle. The distribution of its sales there is:

$$F_n(x) = 1 - \Pr[X \geq x | X \geq \sigma E_n] = 1 - \Pr\left[C \leq \left(\frac{x}{X_n}\right)^{1/(\sigma-1)} \middle| C \leq \left(\frac{\sigma E_n}{X_n}\right)^{1/(\sigma-1)}\right]$$

which, from (14), is:

$$F_n(x) = 1 - \left(\frac{x}{\sigma E_n}\right)^{-\widetilde{\theta}} \quad x \geq \sigma E_n. \tag{16}$$

while mean sales are:

$$\overline{x}_n = \frac{\sigma E_n}{1 - 1/\widetilde{\theta}}.$$

That is, the sales distribution is Pareto with slope $\widetilde{\theta}$.

2. **Entry.** From (15), the measure of firms selling in market n is simply:

$$J_n = \left(1 - 1/\widetilde{\theta}\right) x_n \tag{17}$$

3. **Sales in a Market and Number of Markets Served.** Consider the sales of a firm from country i selling in market n. Its sales in that market are drawn from the

distribution $F_n(x)$ above and its cost in market n must be below $\overline{c}_n$. If the firm sells in markets that are less popular than n, its cost in market n must be lower still, implying higher sales in n. Denote by $J_{ni}^{(k)}$ the measure of firms from i selling in n that also sell in at least k less popular markets than n. This measure is decreasing in k. From (16) above, to sell in at least k less popular markets, sales in n must be at least:

$$\underline{x}_{ni}^{(k)} = \sigma E_n \left(\frac{J_{ni}^{(k)}}{J_{ni}^{(0)}} \right)^{-1/\widetilde{\theta}}$$

while the mean sales in market n of firms from i selling to k less popular destinations than n is:

$$\overline{x}_{ni}^{(k)} = \frac{\sigma E_n}{1 - 1/\widetilde{\theta}} \left(\frac{J_{ni}^{(k)}}{J_{ni}^{(0)}} \right)^{-1/\widetilde{\theta}} \tag{18}$$

The model delivers a precise relationship between a firm's sales in any given market and the number of less popular markets it sells in.

3.6 Application to French Firms

Our particular focus is on the model's implications for observations on French firms and their export activity. Furthermore, we assume that French firms are observed only if they sell in the French market. Our model does not impose this last requirement, so we will interpret our data on French firms as being a truncated sample.

If a French firm producing good j were to enter market n, it would sell:

$$X_n^*(j) = \alpha_n(j) \left(\frac{\overline{m} c_{nF}(j)}{P_n} \right)^{1-\sigma} X_n \tag{19}$$

where $c_{nF} = w_F d_{nF} / z_F(j)$. To enter the market it has to overcome two distinct hurdles. First, its operating profits need to overcome the cost of entry, meaning that:

$$X_n^*(j) \geq \sigma E_n \varepsilon_n(j), \tag{20}$$

what we call the entry hurdle. Second, it must be the lowest cost supplier of good j to market n, meaning that:

$$c_{nF}(j) < \widetilde{c}_n(j) = \min_{i \neq F}\{c_{ni}(j)\}, \tag{21}$$

what we call the competition hurdle.

It is important to remember that we treat the $\eta_n(j)$ as applying to all potential sellers of good j in market n, regardless of source. Hence if a French firm passes the entry hurdle in destination n so does any other seller with a lower unit cost in that market. Hence the entry hurdle never protects a French firm from a lower unit cost competitor.

4 Quantification I: Calibrating Monopolistic Competition

The special case of the model with monopolistic competition and no shocks to sales or the entry barrier, is particularly simple to calibrate.

A strict implication, as discussed, is a market hierarchy. In fact, every firm in our sample sells in France, so that this first element of the hierarchy is not violated. But 48 percent of the over 35 thousand firms that export in our sample don't sell in Belgium, the most popular foreign destination. Looking at the top seven destinations, 73 percent of exporters violate the hierarchy by skipping more popular destinations. But looking at this figure another way, 27 percent of firms sell to a string of destinations that satisfies the market hierarchy. Such strings constitute only 7 of 128 ($= 2^7$), or 5.5 percent, of possible configurations of sales to 7 markets.[4]

Moreover, organizing firms according to the least popular market they serve and according to the number of markets they serve gives very similar results. Figure 1 graphs the number of firms selling to k or more markets against the number of firms selling to the k'th most popular market. The relationship suggests a rough one-to-one correspondence. Furthermore, as we discuss below, selling in less popular markets has very similar implications for sales in France than selling in many markets. Hence, while the strict implication of market hierarchies is violated, we think that there are enough features of the data consistent with this implication that this simple version of the model is worth exploring further to see what it has to say about parameters of interest. We go on to examine the three relationships discussed above, in reverse order.

4.1 Sales in France by Exporters to Multiple Destinations

Figure 2a plots average sales in France of French firms selling to k or more markets against the number of firms selling to k or more markets. It is the observational analogue of (18) above. Note that the relationship is tight, and approximately linear on a logarithmic scale, as the theory implies. Moreover, its slope is -2/3, suggesting a value of $\widetilde{\theta}$ of 1.5. Hence monopolistic competition combined with an assumption that efficiencies have a Pareto distribution fits the the relationship between French firms' sales in France and the number of export destinations

[4] The most popular foreign destinations, in order, are Belgium, West Germany, Switzerland, Italy, United Kingdom, the Netherlands, and USA. With its prediction of a hierarchy of destinations, the basic model cannot explain how, for example, we could ever observe a French firm selling in Italy but not in Belgium.

that they serve.

Returning to the issue of hierarchies, Figure 2b plots sales in France of French firms selling to the k'th most popular market against the frequency of firms selling there. Note that the relationship is very similar. French firms that sell to unpopular markets sell systematically more in France, just as French firms that sell to many markets sell more in France.

4.2 Entry and the Price Index

Under monopolistic competition, as well as for a wide range of other market structures, the number of French firms selling to a destination, divided by French market share, provides an estimate of the total number of firms selling there. We thus use (17) to infer σE_n across our 113 destinations, using J_{nF}/π_{nF} as a proxy for J_n. That is, we calculate:

$$\sigma E_n = (1 - 1/\widetilde{\theta})\overline{x}_{nF}$$

where $\overline{x}_{nF} = X_{nF}/J_{nF}$ is mean sales of French firms in market n (using our estimate of $\widetilde{\theta} = 1.5$).

Figure 3 plots our estimate of σE_n against total market size X_n (measured as manufacturing absorption, home production plus imports minus exports) on a logarithmic scale. Note that the relationship is linear, upward sloping, and quite tight. A linear regression of $\ln(\sigma E_n)$ against $\ln X_n$ has an R^2 of .71. The slope is .36 with a standard error of .02.

We can use our estimates of σE_n and $\widetilde{\theta}$, along with data on X_n and conjectures about σ, to infer the contribution of x_n to the price index, using (12) (setting $a_1 = 1$) Specifically we calculate the term:

$$\widetilde{P}_n(\sigma) = x_n^{-(1-1/\widetilde{\theta})/(\sigma-1)} = \left[\frac{X_n}{(1-1/\widetilde{\theta})\overline{x}_{nF}}\right]^{-1/[3(\sigma-1)]} = \left[\frac{J_{nF}}{(1-1/\widetilde{\theta})\pi_{nF}}\right]^{-1/[3(\sigma-1)]}$$

for each destination n using various values of σ. Figure 4 plots this component of the price index against market size X_n on a logarithmic scale. A regression of $\ln \widetilde{P}_n$ against $\ln X_n$ has an R^2 of .89 and yields a regression coefficient of $-.046$ (standard error .002). The implication is that a doubling of market size is associated with a decline in the price index of $-.046$ percent due to increased entry.[5] Size has a modest but notable effect on welfare through increased variety.[6]

4.3 The Sales Distribution

In the simple case of monopolistic competition, the distribution of sales in any market is given by (16), a Pareto distribution with parameter $\widetilde{\theta}$. From above, the relationship between a firm's sales in France and number of markets its serves implies that $\widetilde{\theta} = 1.5$.

Figure 4 plots the average sales distribution of French firms across destinations (distinguishing among markets according to whether France's total exports there are large, medium, or small)..Two things don't fit. First, the relationship is nonlinear at the lower end of the distribution, violating the implication that the sales distribution should be linear in logarithms. Second, the slope at the upper end is too steep, with a slope closer to -1 than -2/3. The

[5] Recalculating $\widetilde{P}_n^e(\sigma)$ for $\sigma = 3$ yields an elasticity of the price index with respect to market size of -11 percent.

[6] Taking $\widetilde{\theta} \rightarrow \infty$ delivers the standard formulation of monopolistic competition with homogeneous firms. Note that in this case market size has a greater effect on welfare through variety as the elasticity of $\widetilde{P}^e(\sigma)$ with respect to size is $-1/(\sigma - 1)$. Heterogeneity in technology attenuates the effect of size on welfare as larger markets attract higher-cost firms on the margin. A point made by Ghironi and Melitz (2004) is that, with technological heterogeneity, the average price of a good sold will be higher in a large market, even though the true price index is lower due to greater variety of goods.

sales distribution is more skewed than what is implied by the value of $\widetilde{\theta}$ inferred from the size advantage in France of prolific exporters implies.

We conclude that the model of monopolistic competition does a good job of picking up the relationship between exports in any given market and the number of markets served. It provides hints about the cost of entry (σE_n) and about the ratio of the heterogeneity parameters ($\widetilde{\theta} = \theta/(\sigma - 1)$). But it does not explain aspects of entry (with its prediction of a strict hierarchy of destinations) and it understates the curvature and heterogeneity of sales in any given market.

5 Quantification II: Simulated Method of Moments

We now turn to the estimation of a more general model to assess its ability to grapple with these feature of the data. We generalize the case above by allowing for destination-specific shocks to entry and to sales, and by treating the range of goods J as a parameter to be estimated.

5.1 Stochastic Specification

In all of the quantitative analysis, it is convenient to isolate the stochastic component of $c_{ni}(j)$ by introducing the variable:

$$u_i(j) = (T_i/J)\,(w_i d_{ni})^{-\theta} c_{ni}(j)^{\theta}.$$

Our assumptions imply that $u_i(j)$ is the realization of a random variable U_i drawn from the unit exponential distribution:

$$\Pr[U_i \leq u] = 1 - \exp(-u). \tag{22}$$

This definition allows us to express $c_{ni}(j)$ as:

$$c_{ni}(j) = (w_i d_{ni})(T_i/J)^{-1/\theta} u_i(j)^{1/\theta}.$$

Invoking expression (3) for the trade share, implies:

$$c_{ni}(j) = (\Phi_n/J)^{-1/\theta} \left(u_i(j)/\pi_{ni}\right)^{1/\theta} \tag{23}$$

We can express our competitiveness hurdles (21) in terms of the $u_i(j)$'s and data on trade shares as:

$$u_F(j) < \widetilde{u}_n(j) = \min_{i \neq F}\{\pi_{nF} u_i(j)/\pi_{ni}\}$$

if the firm is competitive in market n. We use $\widetilde{u}(j)$ to denote the vector of competitiveness hurdles across all markets n. As mentioned above, a necessary condition for a French firm to appear in our data is that the firm sells in France. Thus, it must pass the competitiveness hurdle for the French market, i.e. $u_F(j) < \widetilde{u}_F(j)$.

Aside from the $u_i(j)$'s, our model has the stochastic components $\alpha_n(j)$ and $\eta_n(j)$. We assume that these components have the joint bivariate lognormal distribution:

$$\begin{bmatrix} \ln \alpha \\ \ln \eta \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_a\sigma_h \\ \rho\sigma_a\sigma_h & \sigma_h^2 \end{pmatrix}\right].$$

Under this distributional assumption, the expression (9) for $\widetilde{P}(x_n)$ becomes:

$$\widetilde{P}(x_n) = \exp\left(\frac{\sigma_a^2(1-\rho^2)}{2}\right) J^{1-1/\widetilde{\theta}} E_\eta \left[\eta^{\rho\sigma_a/\sigma_h}\Gamma\left(1 - 1/\widetilde{\theta}, \left(\frac{\eta x_n}{\widetilde{P}(x_n)}\right)^{\widetilde{\theta}} J^{-1}\right)\right].$$

The expression (19) for the latent sales of a French firm in market n (actual sales if it enters that market) can be simplified by exploiting (23) and (8):

$$X_n^*(j) = \alpha_n(j) X_n \left(\frac{J u_F(j)}{\pi_{nF}} \right)^{-1/\widetilde{\theta}} \widetilde{P}(x_n)^{-1}.$$

which is, in logs:

$$\ln X_n^*(j) = \Gamma_n - \widetilde{\theta}^{-1} \ln u_F(j) + \ln \alpha_n(j),$$

where the vector Γ summarizes firm-invariant country-level variables with representative element:

$$\Gamma_n = \ln(X_n/\widetilde{P}(x_n)) + \widetilde{\theta}^{-1} \ln(X_{nF}/X_n) - \widetilde{\theta}^{-1} \ln J. \tag{24}$$

Finally, we can express our entry hurdles (20) as:

$$\ln u_F(j) < \ln \overline{u}_n(j) = \widetilde{\theta} \left[\Gamma_n - \ln (\sigma E_n) + \ln \eta_n(j) \right]$$

if the firm covers the fixed cost of entering market n. Thus, for a French firm to enter market n it must be that $u_F(j) < \widetilde{u}_n(j)$ and $u_F(j) < \overline{u}_n(j)$.

To illustrate the role of the parameter J it is useful to consider the special case of $E_n = 0$. In that case the mean sales of a French firm in market n, conditional on entry, is simply X_n/J. The parameter J is the scale factor between aggregate magnitudes and firm-level magnitudes. In the extreme case of $J \to \infty$, on the other hand, the mean sales of entrants varies in proportion to σE_n. More generally, J also enters the model in a more subtle way. Notice that as J gets larger, the entry hurdle gets increasingly difficult to pass. The competitiveness hurdle, on the other hand, is invariant to J. Thus J parameterizes the relative importance of the two hurdles.

The parameters of the model are σ_a^2, σ_h^2, ρ, $\widetilde{\theta}$, J, and, for each country, σE_n. We do not try to estimate the full set of σE_n's. Based on our results for the simple case of monopolistic competition we specify $\sigma E_n = \gamma X_n^{\phi}$, giving us two new parameters γ and ϕ. The vector Θ of 7 parameters of the model is then:

$$\Theta = \left(\widetilde{\theta} \quad J \quad \phi \quad \gamma \quad \sigma_a^2 \quad \sigma_h^2 \quad \rho \right)'.$$

For a value of Θ we can simulate a dataset of firms competing in each of 113 markets around the world, following the procedure described in the appendix. We can then extract firms located in France from that simulated dataset, and observe their entry and sales in markets around the world. Moments generated by these simulated data can then be compared with the actual data.

5.2 Estimation by Simulated Method of Moments

To estimate Θ we seek a value that generates a simulated dataset that approximates the actual data in the following moments:

1. The number of French firms entering each of 113 destinations.

2. The fraction of simulated firms selling the amount sold by the actual 5th percentile of sales in each country.

3. The fraction of simulated firms selling the amount sold by the actual 75th percentile of sales in each country.

4. The fraction of simulated firms selling the amount sold by the actual 95th percentile of sales in each country.

5. The number of firms selling to b or more markets.

6. Average sales in France of firms selling to b or more markets.

7. Average exports of firms selling to b or more markets.

8. The number of firms selling to subsets of the 7 most popular destinations.

Using the amoeba algorithm we searched over values of Θ to minimize the difference between these moments of our simulated data and the actual data.

5.3 Parameter Estimates

The procedure yielded the following estimates for Θ:

$\widetilde{\theta}$	J	ϕ	γ	σ_α^2	σ_η^2	ρ
1.5	$13 * 10^6$	.30	.0034	1.8	0.95	$-.32$

Note first that our estimation yields the same value of $\widetilde{\theta}$ as that provided by the calibration of the simpler model. Our estimate of the elasticity of entry with respect to market size is only slightly lower. Moreover, the value of J is enormous. Our simulations almost never delivered multiple potential suppliers of the same good, consistent with monopolistic competition. The richer model's predictions about how entry varies with market size, and about sales in France of firms that sell to b or more markets are very similar to those of the simpler model. Figures 5, 6, and 7 compare some other moments of our simulated data with moments of the actual data. Figure 5 reports the actual and simulated number of firms selling to at least b countries, for $b = 1, 2, 4, 8, 16, 32$, and 64. Figures 6 and 7 report sales in France and export sales, respectively, according to this same classification. Figure 8 shows how our simulated data pick up on the number of firms entering into different markets.

What about the dimensions in which the simpler model failed? Remarkably, the richer model, in which country-specific shocks to entry $\eta_n(j)$ can generate deviations from a hierarchy of export markets, delivers a simulated data set of firms in which 27 percent follow the proper hierarchy among the top 7 destinations, the same fraction as in the actual data. Figure 9 plots the distribution of sales in France. The richer model, in which loglinear country-specific shocks to sales lead to a more skewed distribution of sales in the upper tale and introduces curvature in the lower tail.

We conclude that a quite simple model of monopolistic competition, with technological heterogeneity and good and country-specific shocks to entry and to sales, can pick up the basic features of the micro-level data very well.

References

Aw, Bee Yan, Chung, Sukkyun, and Roberts, Mark J. (1998), "Productivity and the Decision to Export: Micro Evidence from Taiwan and South Korea." *NBER Working Paper No. 6558*.

Bernard, Andrew B., Eaton, Jonathan, Jensen, J. Bradford, and Kortum, Samuel (2003), "Plants and Productivity in International Trade." *American Economic Review 93*: *1268-1290*.

Bernard, Andrew B. and Jensen, J. Bradford (1995), "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987." *Brookings Papers on Economic Activity: Microeconomics*: 67-119.

Bernard, Andrew B. and Jensen, J. Bradford (1999) "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics 47*: 1-25.

Biscourp, Pierre and Kramarz, Francis (2002), "Les Firmes Francaises et le Commerce Exterieur: Une Etude Descriptive sur la Periode 1986-1992," mimeo, CREST.

Chaney, Thomas (2005), "Distorted Gravity: Heterogeneous Firms, Market Structure, and the Geography of International Trade." mimeo, University of Chicago.

Clerides, Sofronis, Lach, Saul, and Tybout, James R. (1998), "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco." *Quarterly Journal of Economics 113*: 903-947.

Eaton, Jonathan and Kortum, Samuel. (1999), "International Technology Diffusion: Theory and Measurement." *International Economic Review 40*: 537-570.

Eaton, Jonathan and Kortum, Samuel (2002), "Technology, Geography, and Trade." *Econometrica 70*: 1741-1780.

Eaton, Jonathan, Kortum, Samuel, and Francis Kramarz (2004), "Dissecting Trade: Firms, Industries, and Export Destinations." *American Economic Review, Papers and Proceedings*, 94: 150-154.

Feenstra, Robert C., Lipsey, Robert E. and Bowen, Henry P. (1997), "World Trade Flows, 1970-1992, with Production and Tariff Data." *NBER Working Paper No. 5910.*

Ghironi, Fabio and Marc J. Melitz (2004), "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Boston College Working Paper No. 599.*

Hallward-Driemeier, Mary, Iaross, Giusepe, and Sokoloff, Kenneth (2002), "Exports and Manufacturing Productivity in East Asia: A Comparative Analysis with Firm-Level Data." NBER WP #8894

Helpman, Elhanan, Melitz, Marc J., and Yeaple, Stephen R. (2004), "Export versus FDI with Heterogenous Firms," forthcoming, *American Economic Review.*

Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica 71*: 1695-1725.

OECD (1995), *The OECD STAN Database.*

Roberts, Mark and Tybout, Robert (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review 87*: 545-564.

Sutton, John.(1991), *Sunk Costs and Market Structure.* (Cambridge, MA: The MIT Press).

UNIDO (1999), *Industrial Statistics Database.*

6 Appendix: The Simulation Algorithm

Given a vector of parameters we can simulate the behavior of our sample of French firms. We will describe our algorithm as if we were simulating an arbitrary number French firms, indexed by $j = 1, \ldots S$. But, we can easily scale the results to be comparable to our data on all French firms selling in France. We introduce notation here for the data on French firms, indexed by $j = 1, \ldots, J_{data}$. We define the indicator $S_n(j) = 1$ if we observe firm j selling in destination n, and $S_n(j) = 0$ if we observe no sales. In the destinations n in which we observe sales (where $S_n(j) = 1$), we let $y_n(j)$ be the natural logarithm of the firm's sales (and arbitrarily set $y_n(j) = 0$ when $S_n(j) = 0$). Thus: $y_n(j) = S_n(j) \ln X_n(j)$.

1. Stage 1 does not require any parameter values and uses data only on the world bilateral trade matrix (expressed as shares of the importer's absorption) with representative element $\pi_{ni} = X_{ni}/X_n$. It involves four steps.

   (a) Draw $v_i(j)$'s independently from $U[0,1]$, for $i = 1, \ldots, 113$ and $j = 1, \ldots, S$.

   (b) For $i \neq F$ calculate $S \times 112$ values of:

   $$u_i(j) = -\ln\left[1 - v_i(j)\right]$$

   (c) Use the $u_i(j)$'s and the π_{ni}'s to construct $S \times 113$ competitiveness hurdles:

   $$\widetilde{u}_n(j) = \min_{i \neq F}\{\pi_{nF} u_i(j)/\pi_{ni}\}.$$

   (d) Independently draw $S \times 113$ realizations of $a_n(j)$ and $h_n(j)$ from:

   $$\begin{bmatrix} a_n(j) \\ h_n(j) \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$$

2. Stage 2 requires data for each destination n on X_n and π_{nF} as well as a set of parameters. It involves three steps,

   (a) Fix values for $\widetilde{\theta}, \gamma, \phi, \sigma_a^2, \sigma_h^2, \rho$, and J.

   (b) Calculate $\widetilde{P}(x_n)$ for each destination n as the solution to:

$$\widetilde{P}(x_n) = \exp\left(\frac{\sigma_a^2(1-\rho^2)}{2}\right) J^{1-1/\widetilde{\theta}} E_\eta \left[\eta^{\rho\sigma_a/\sigma_h} \Gamma\left(1 - 1/\widetilde{\theta}, \left(\frac{\eta x_n}{\widetilde{P}(x_n)}\right)^{\widetilde{\theta}} J^{-1}\right)\right]$$

   (c) Calculate:

$$\Gamma_n = \ln(X_n/\widetilde{P}(x_n)) + \widetilde{\theta}^{-1} \ln \pi_{nF} - \widetilde{\theta}^{-1} \ln J.$$

   for each destination n.

3. Stage 3 combines the simulation draws from Stage 1 and the parameter values and destination variables from Stage 2. It involves seven steps.

   (a) Use the draws from 1d and the parameter values from 2a to construct $S \times 113$ realizations for each of $\ln \alpha_n(j)$ and $\ln \eta_n(j)$ as:

$$\begin{bmatrix} \ln \alpha_n(j) \\ \ln \eta_n(j) \end{bmatrix} = \begin{bmatrix} \sigma_a\sqrt{1-\rho^2} & \sigma_a\rho \\ 0 & \sigma_h \end{bmatrix} \begin{bmatrix} a_n(j) \\ h_n(j) \end{bmatrix}$$

   (b) Construct the $S \times 113$ entry hurdles:

$$\overline{u}_n(j) = \exp\left\{\widetilde{\theta}\left[\Gamma_n - \ln(\sigma E_n) + \ln \eta_n(j)\right]\right\}.$$

   (c) Construct S joint hurdles $\overline{u}(j)$ faced by a French firm in its home market:

$$\overline{u}(j) = \min\{\overline{u}_F(j), \widetilde{u}_F(j)\}.$$

   Note that a French firm will sell in France if and only if it passes both the entry hurdle and the competitiveness hurdle there, i.e. iff $u_F(j) \leq \overline{u}(j)$.

(d) Construct S probability weights:

$$p[\overline{u}(j)] = 1 - \exp[-\overline{u}(j)].$$

If we were to construct $u_F(j)$ in a manner parallel to how we constructed all the other $u_i(j)$'s in step 1 b, $p[\overline{u}(j)]$ would be the probability of the French firm selling in France.

(e) We actually construct $u_F(j)$, based on the draw $v_F(j)$ from step 1 a, so that we necessarily obtain a French firm selling in France. To do so, we set:

$$u_F(j) = -\ln\left\{1 - p[\overline{u}(j)]v_F(j)\right\}.$$

Our simulated French firm gets a weight $p[\overline{u}(j)]$ in the sample.

(f) Calculate $S_n(j)$ as determined by the competition and entry hurdles:

$$S_n(j) = \begin{cases} 1 \text{ if } u_F(j) \leq \widetilde{u}_n(j) \text{ and } u_F(j) \leq \overline{u}_n(j) \\ 0 \text{ otherwise.} \end{cases}$$

(g) Wherever $S_n(j) = 1$ calculate log sales as:

$$\ln X_n(j) = \Gamma_n - \widetilde{\theta}^{-1} \ln u_F(j) + \ln \alpha_n(j).$$

Following this procedure we simulate the behavior of S firms. In generating statistical moments from this simulated sample, we need to keep track of two issues. First, when summing across firms, we must apply the sampling weight $p[\overline{u}(j)]$ to firm j. Second, if we want to mimic the scale of the French data, we need to apply a scaling factor of J/S. In this way the choice of S matters only for the variance of the resulting simulated moments.

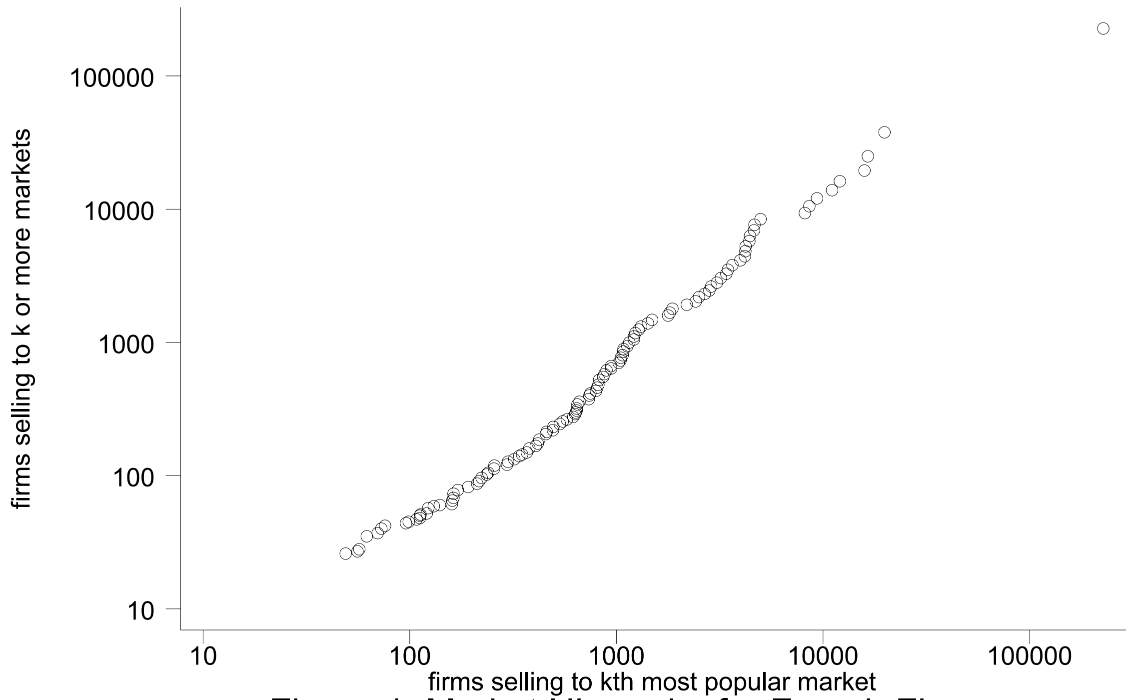


Figure 1: Market Hierarchy for French Firms

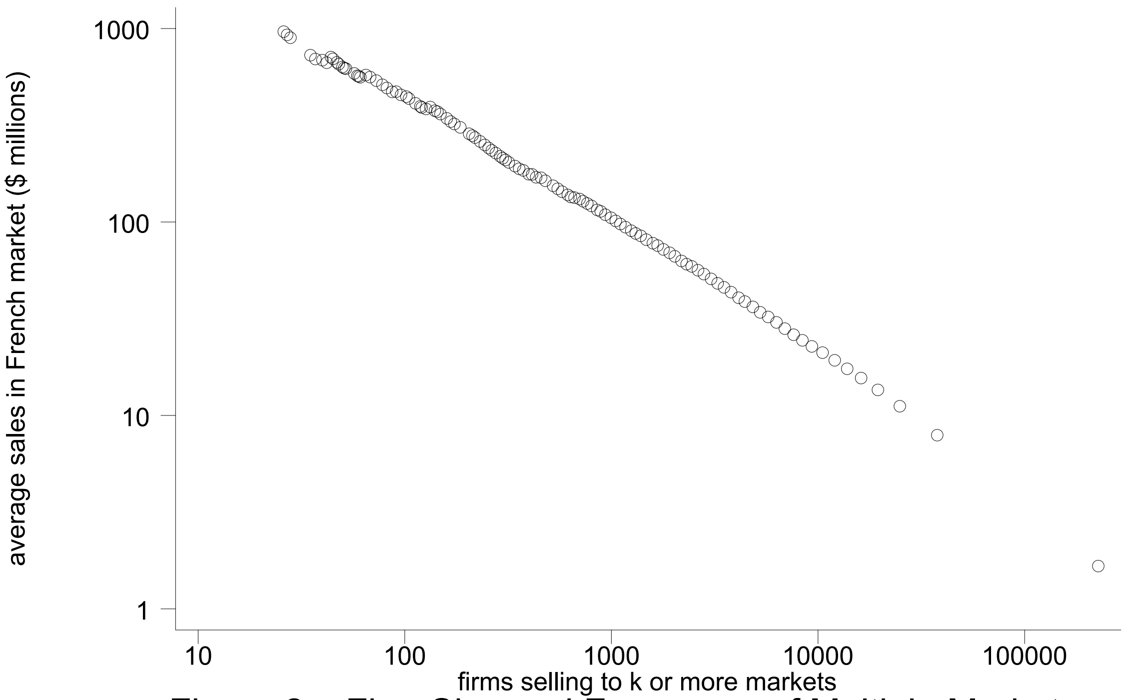


Figure 2a: Firm Size and Frequency of Multiple Markets

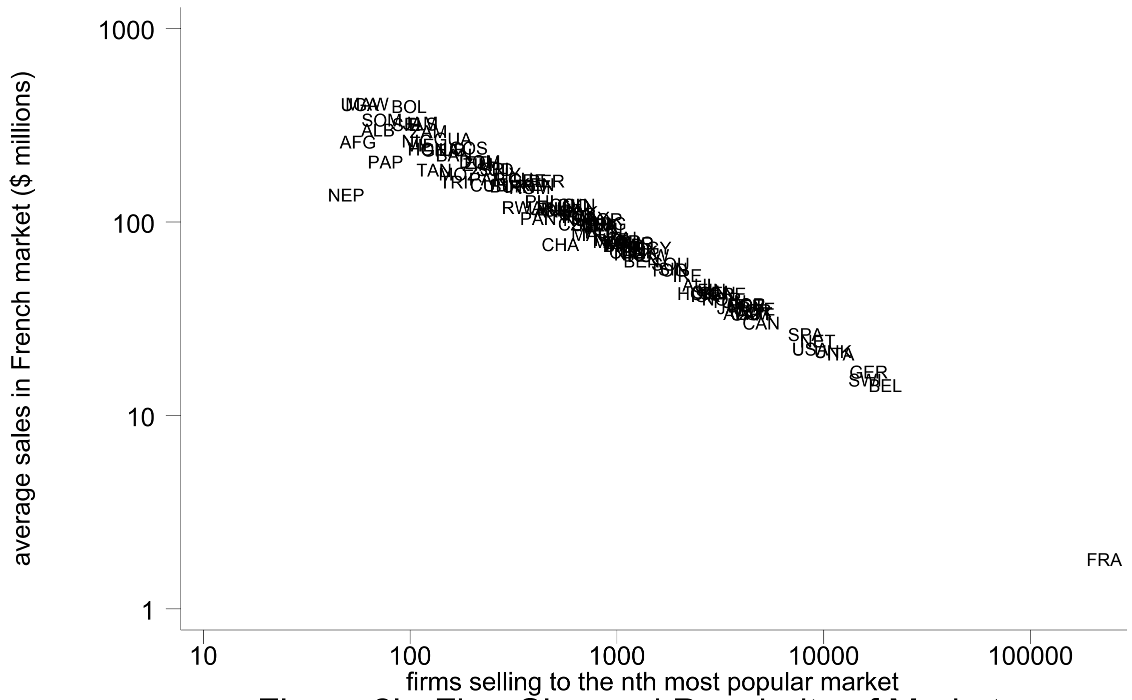


Figure 2b: Firm Size and Popularity of Market

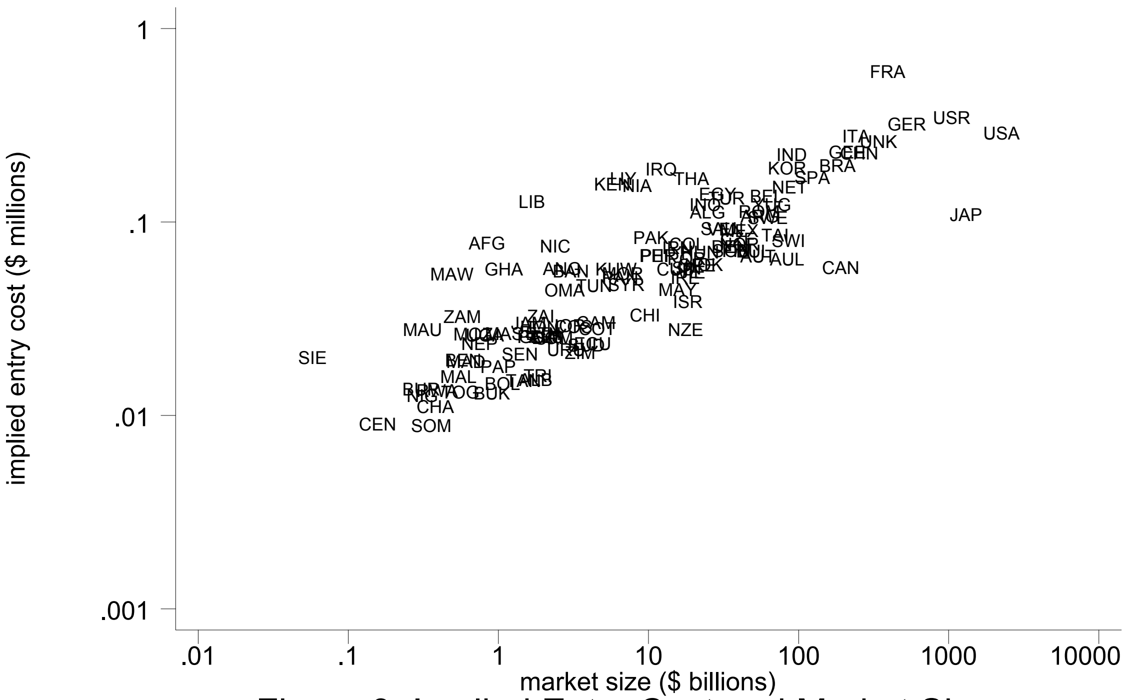


Figure 3: Implied Entry Cost and Market Size

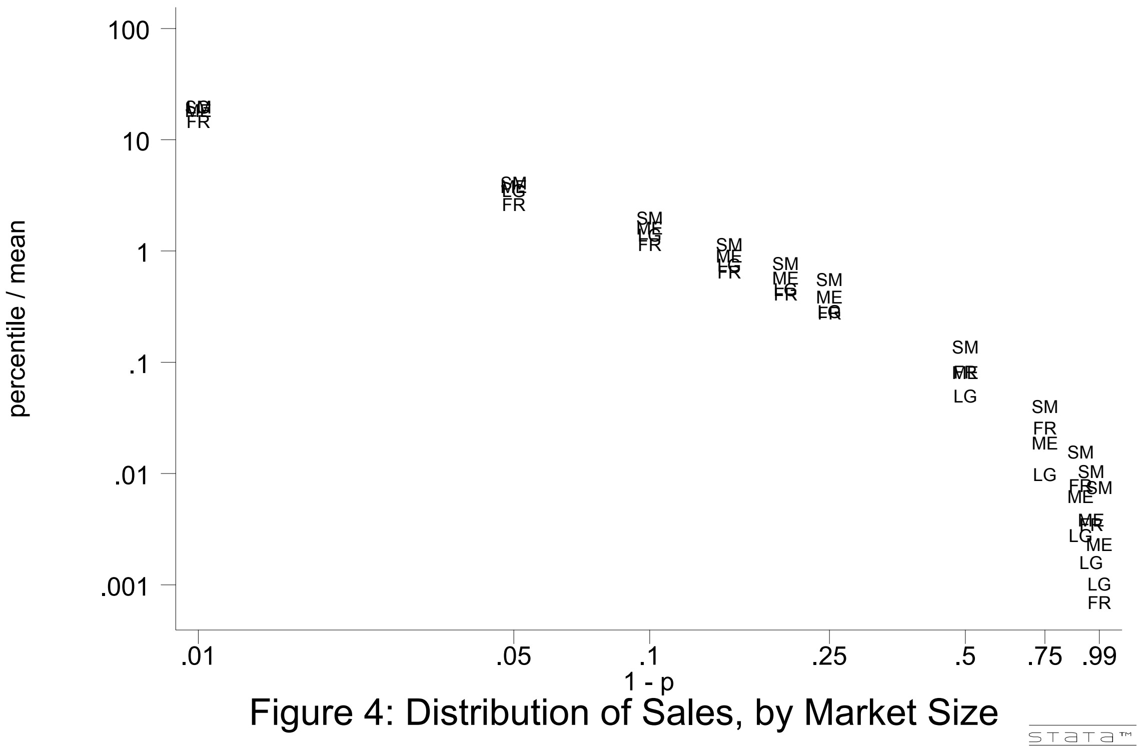


Figure 4: Distribution of Sales, by Market Size

Figure 5: Firms Exporting to B or More Countries

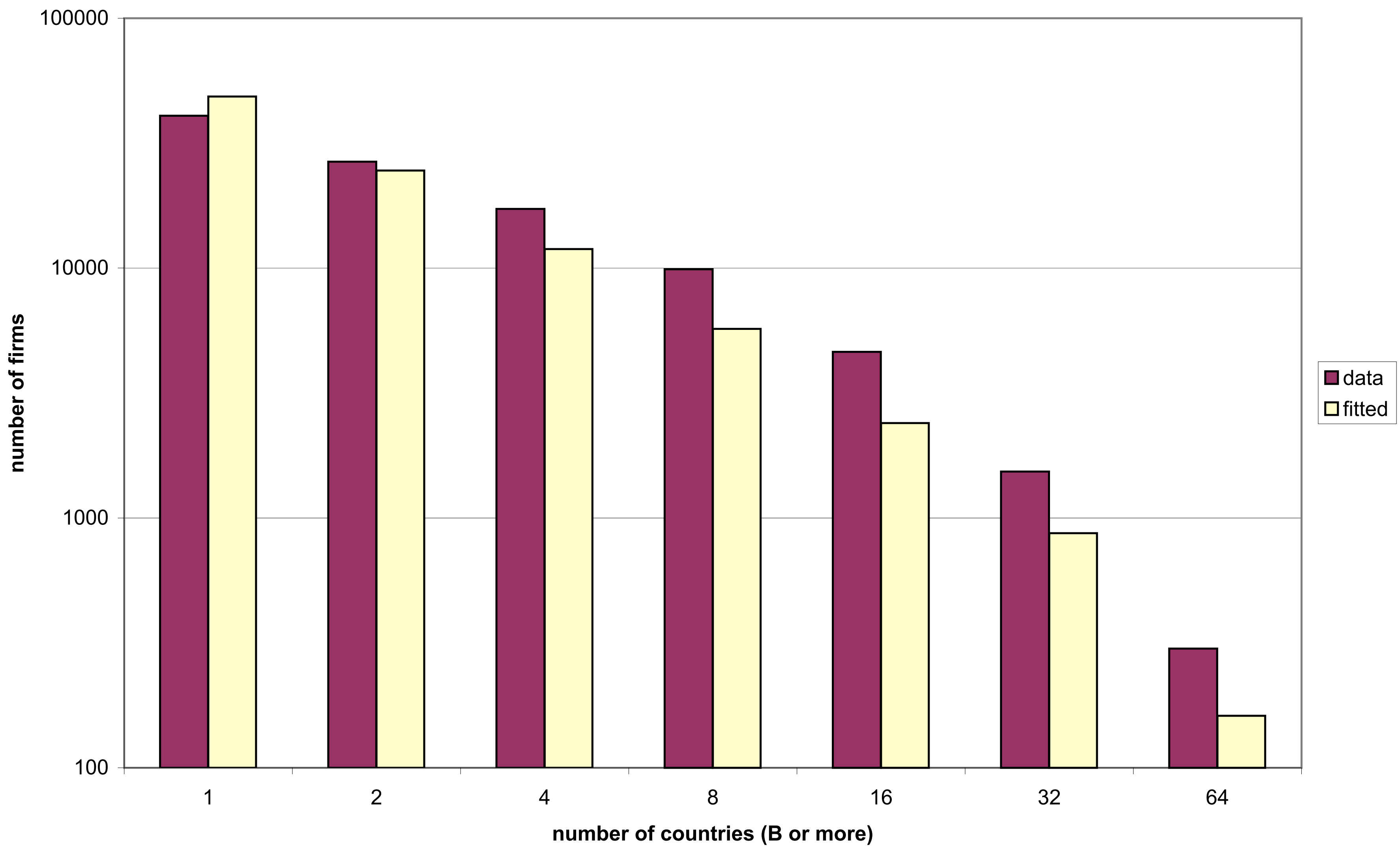


Figure 6: French Sales of Firms Exporting to B or More Countries

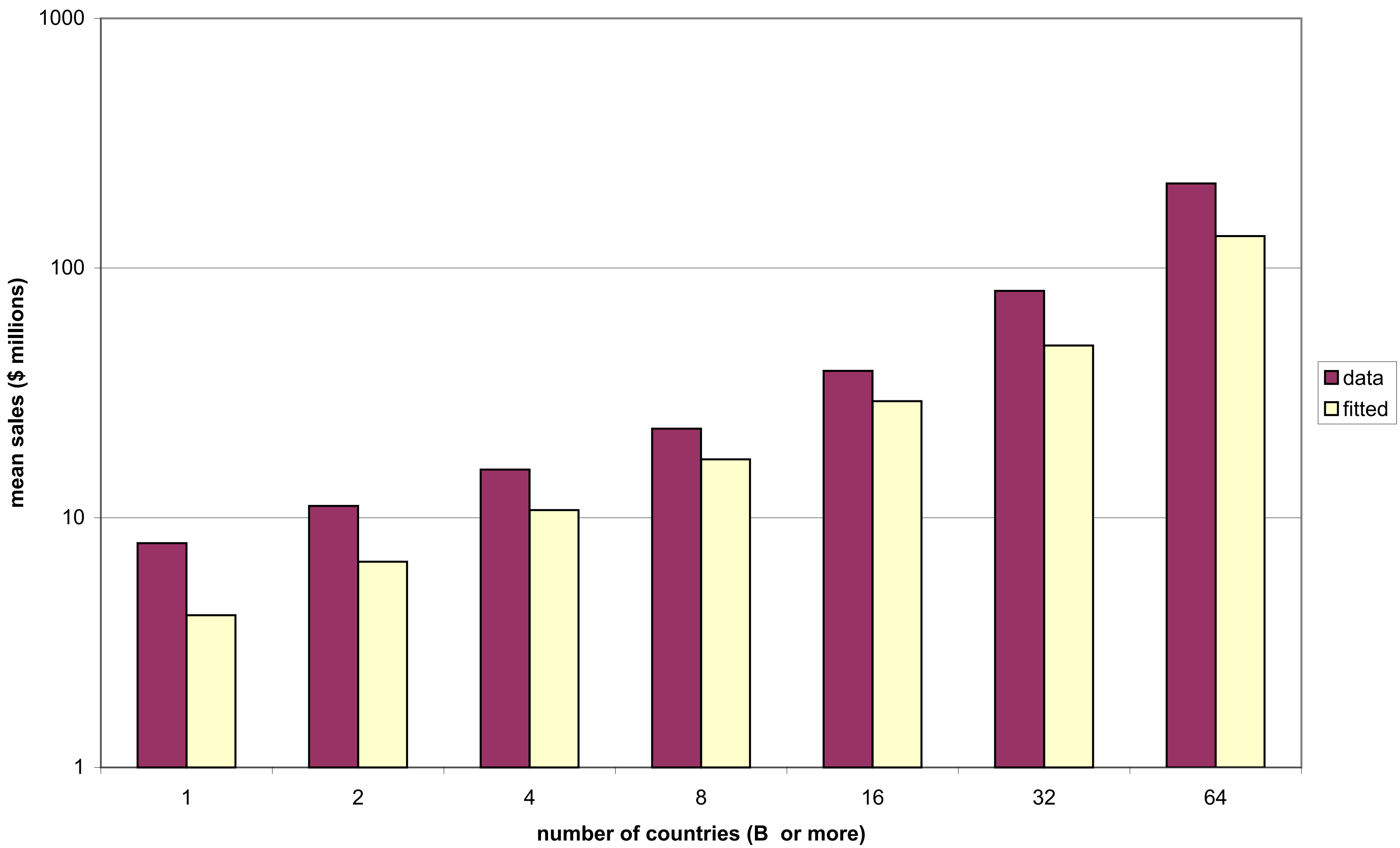


Figure 7: Exports of Firms Selling to B or More Countries

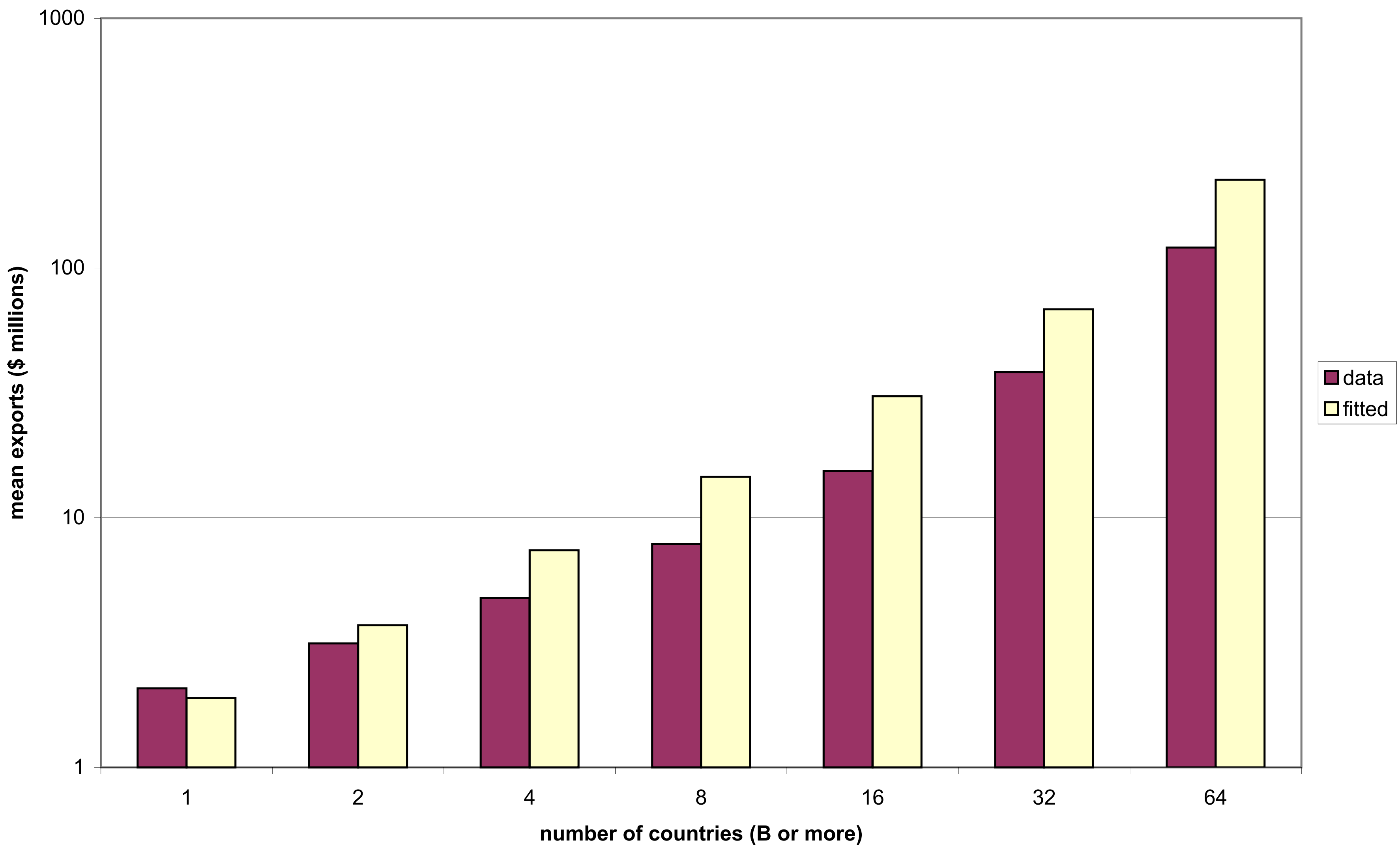


Figure 8: French Firm Entry by Country

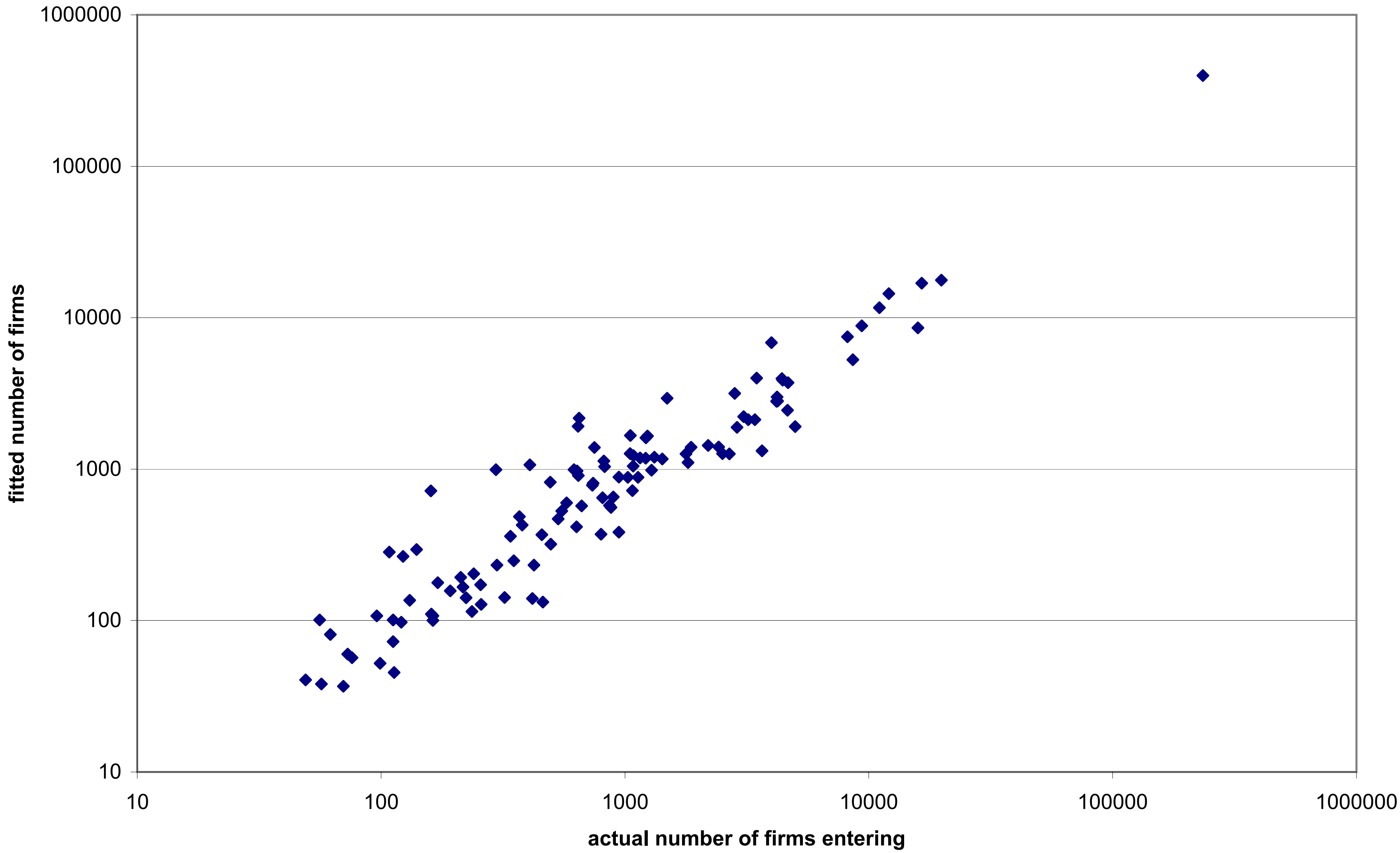
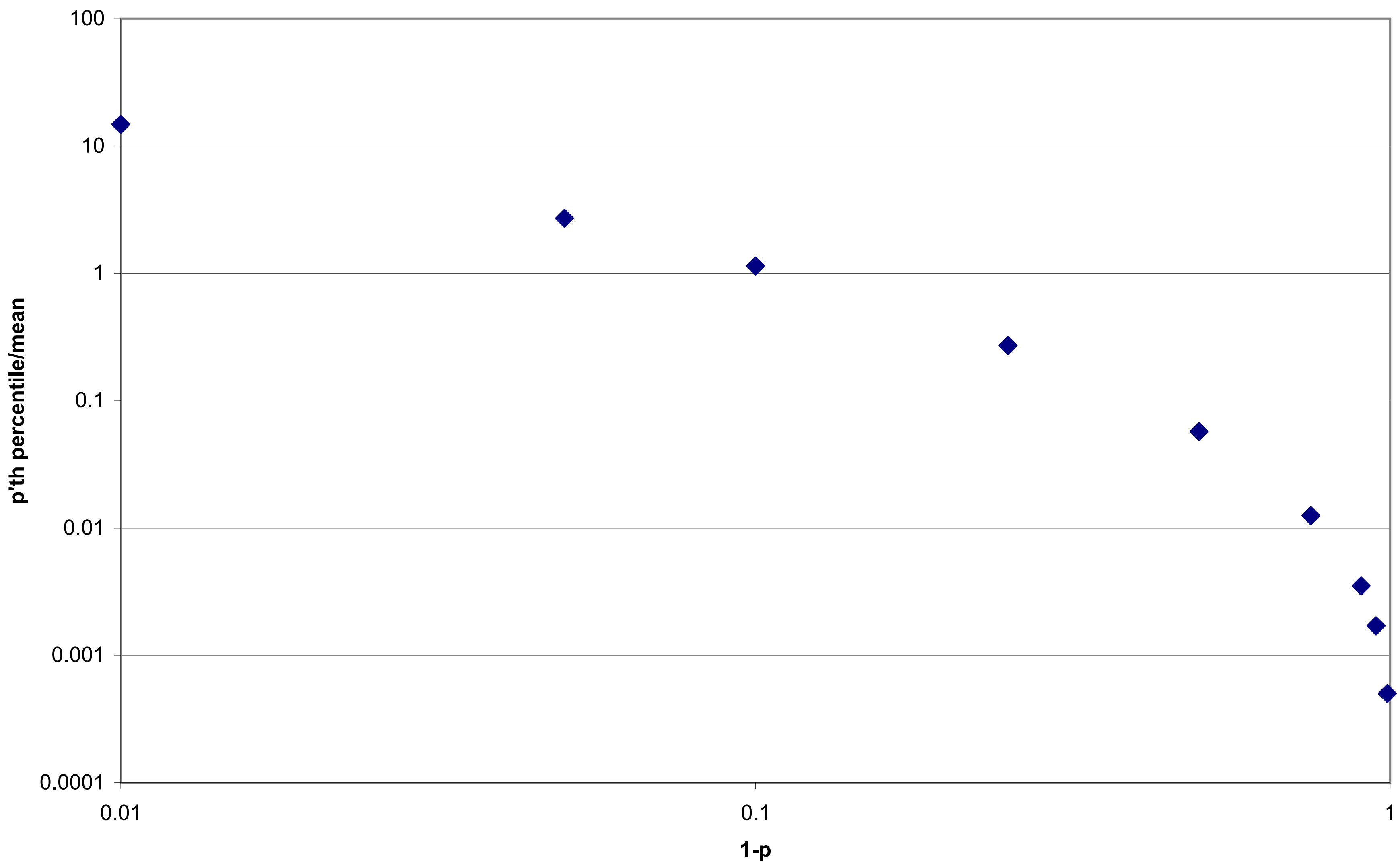


Figure 9: Fitted Sales Distribution in France



An Empirical Model of Growth Through Product Innovation *

Rasmus Lentz
University of Wisconsin-Madison
Boston University and CAM

Dale T. Mortensen
Northwestern University
NBER and IZA

July 17, 2005

Abstract

Productivity dispersion across firms is large and persistent, and worker reallocation among firms is an important source of productivity growth. The purpose of the paper is to estimate the structure of an equilibrium model of growth through innovation that explains these facts. The model is a modified version of the Schumpeterian theory of firm evolution and growth developed by Klette and Kortum (2004). The data set is a panel of Danish firms than includes information on value added, employment, and wages. The model's fit is good and the structural parameter estimates have interesting implications for the aggregate growth rate and the contribution of worker reallocation to it.

Keywords: Labor productivity growth, worker reallocation, firm dynamics, firm panel data estimation.

JEL Classification Numbers: E22, E24, J23, J24, L11, L25, O3, O4.

*This research was supported by a collaborative grant to the authors from the NSF. Funding for data access was provided by the Danish Social Science Research Council through a grant to Bent Jesper Christensen as part of other collaborative research. Rasmus Lentz acknowledges support for this research from Center for Applied Microeconometrics at University of Copenhagen. The authors thank Victor Aguirregabiria, Joseph Altonji, Robert Hall, John Kennan, Samuel Kortum, Jean-Marc Robin, and Rob Shimer for useful comments and suggestions. All remaining errors are ours.

1 Introduction

In their review article, Bartelsman and Doms (2000) draw three lessons from empirical productivity studies based on longitudinal plant and firm data: First, the extent of dispersion in productivity across production units, firms or establishments, is large. Second, the productivity rank of any unit in the distribution is highly persistent. Third, a large fraction of aggregate productivity growth is the consequence of worker reallocation.

Although the explanations for productive firm heterogeneity are not fully understood, economic principles suggest that its presence should induce worker reallocation from less to more productive firms as well as from exiting to entering firms. There is ample evidence that workers do flow from one firm to another frequently. As Davis, Haltiwanger, and Schuh (1996) and others document, job and worker flows are large, persistent, and essentially idiosyncratic in the U.S. Recently, Fallick and Fleischman (2001) and Stewart (2002) find that job to job flows without a spell of unemployment in the U.S. represent at least half of the separations and is growing. In their analysis of Danish matched employer-employee IDA data, Frederiksen and Westergaard-Nielsen (2002) report that the average establishment separation rate over the 1980-95 period was 26%. About two thirds of the outflow represents the movement of workers from one firm to another.

In a companion paper, Lentz and Mortensen (2005) develop a stochastic general equilibrium model in which more profitable firms grow faster and contribute more to the aggregate growth rate through product innovation. The model is a variation on that proposed by Klette and Kortum (2004), which itself builds on the endogenous growth model of Grossman and Helpman (1991). By design, their model is consistent with stylized facts about product innovation and its relationship to the dynamics of firm size evolution. We adopt the approach because it provides an explanation for the fact that there is no correlation between labor force size and labor productivity but a strong positive association between value added and labor productivity in Danish firm data. Furthermore, the model provides a direct link between worker reallocation and productivity growth.

In the model, firms are monopoly suppliers of differentiated intermediate products that serve as inputs in the production of a final consumption good. Better quality products are introduced from time to time as the outcome of R&D investment by both existing firms and new entrants. As new products displace old, the process of creative destruction induces the need to reallocate workers across productive activities. In the version of the model estimated here, product quality differs across firms. In our earlier paper, we established the existence of a general equilibrium solution to the model. In this one, we use the equilibrium relationships implied by the model and information drawn from a Danish panel of firms to estimate the model's parameters.

Providing a good fit to data, the model is estimated on among other moments the relationship between firm size and firm growth which is slightly negative in the data. The model satisfies a theoretical version of Gibrat's law, but nevertheless replicates the negative relationship between size and growth found in data. The model is also estimated to fit a standard growth decomposition which suggests a large growth contribution from reallocation and while the model does in fact imply a large reallocation contribution, the reduced form decomposition is largely explained through measurement error and the fundamental sources of productivity growth are only loosely reflected in the reduced form decomposition.

Given the parameter estimates obtained, we explore the model's quantitative implications for productivity growth and its sources. The model implies an annual rate of overall productivity growth equal to 2.3%. We find that the reallocation of workers from less to more productive surviving firms accounts for 70% of productivity growth in equilibrium.

2 Danish Firm Data

Danish firm data provide information on productivity dispersion and the relationships among productivity, employment, and sales. The available data set is an annual panel of privately owned firms for the years 1992-1997 drawn from the Danish Business Statistics Register. The sample of approximately 4,900 firms is restricted to those with 20 or more employees. The sample does not include entrants.[1] The variables observed in each year include value added (Y), the total wage bill (W), and full-time equivalent employment (N). In this paper we use these relationships to motivate the theoretical model studied. Both Y and W are measured in Danish Kroner (DKK) while N is a body count.

[1]The full panel of roughly 6,700 firms contains some entry, but due to the sampling procedure, the entrant population suffers from significant selection bias. We have chosen not to rely on the entrant population for identification of the model.

Non-parametric estimates of the distributions of two alternative measures of a firm's labor productivity are illustrated in Figure 1. The first measure of firm productivity is value added per worker (Y/N) while the second is valued added per unit of quality adjusted employment (Y/N^*). Standard labor productivity misrepresents cross firm productivity differences to the extent that labor quality differs across firms. However, if more productive workers are compensated with higher pay, as would be true in a competitive labor market, one can use a wage weighted index of employment to correct for this source of cross firm differences in productive efficiency. Formally, the constructed quality adjusted employment of firm j is defined as $N_j^* = W_j/w$ where

$$w = \frac{\sum_j W_j}{\sum_j N_j} \tag{1}$$

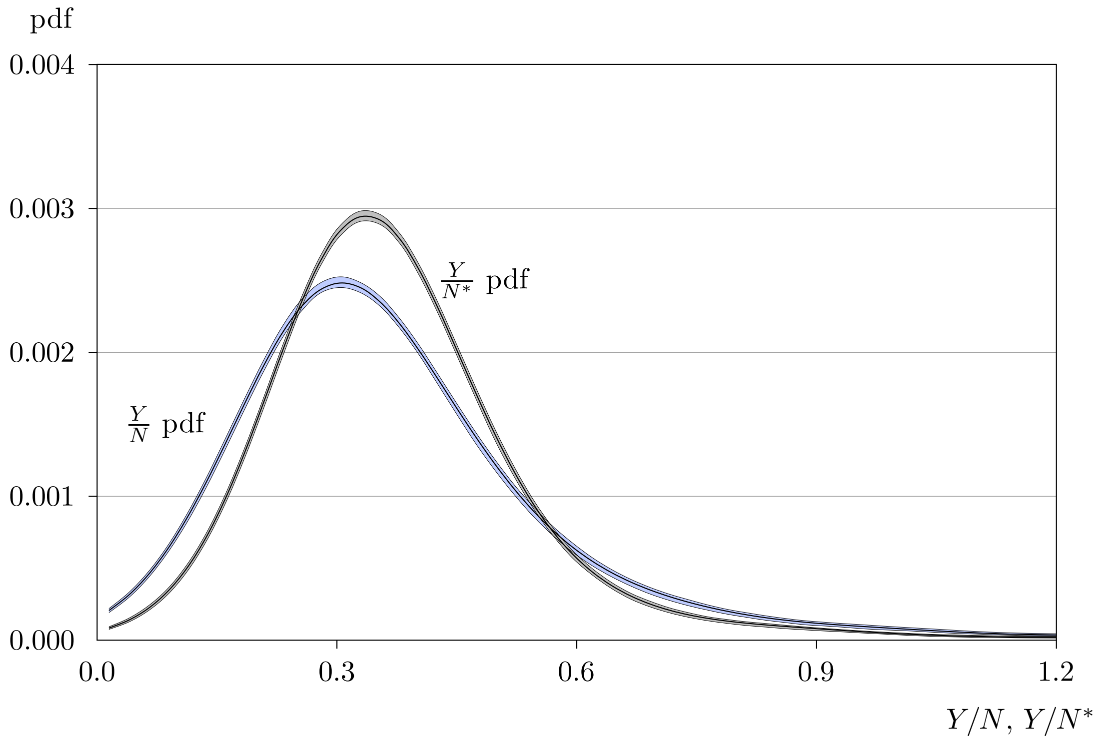
is the average wage paid per worker in the market.[2] Although correcting for wage differences across firms in this manner does reduce the spread and skew of the implied productivity distribution somewhat, both distributions have high variance and skew and are essentially the same general shape.

Both distributions are consistent with those found in other data sets. For example, productivity distributions are significantly dispersed and skewed to the right. In the case of the adjusted measure of productivity, the 5$^{\text{th}}$ percentile is roughly half the mode while the 95$^{\text{th}}$ percentile is approximately twice as large are the mode. The range between the two represents a four fold difference in value added per worker across firms. These facts are similar to those reported by Bartelsman and Doms (2000) for the U.S.

There are many potential explanations for cross firm productivity differentials. A comparison of the two distributions represented in Figure 1 suggests that differences in the quality of labor inputs does not seem to be the essential one. The process of technology diffusion is a well documented. Total factor productivity differences across firms can be expected as a consequence of slow diffusion of new techniques. If technical improvements are either factor neutral or capital augmenting, then one would expect that more productive firms would acquire more labor and capital. The implied consequence would seem to be a positive relationship between labor force size and labor

[2] In the case, where a firm is observed over several periods, the implicit identification of the firm's labor force quality is taken as an average over the time dimension to address issues of measurement error. The alternative approach of identifying a quality measure for each year has no significant impact on the moments of the data set.

Figure 1: Productivity Distributions.



Note: The shaded areas represent 90% bootstrap confidence intervals. Value added (Y) measured in 1 million DKK. N is the firm's labor force head count and N^* is the quality adjusted labor force size.

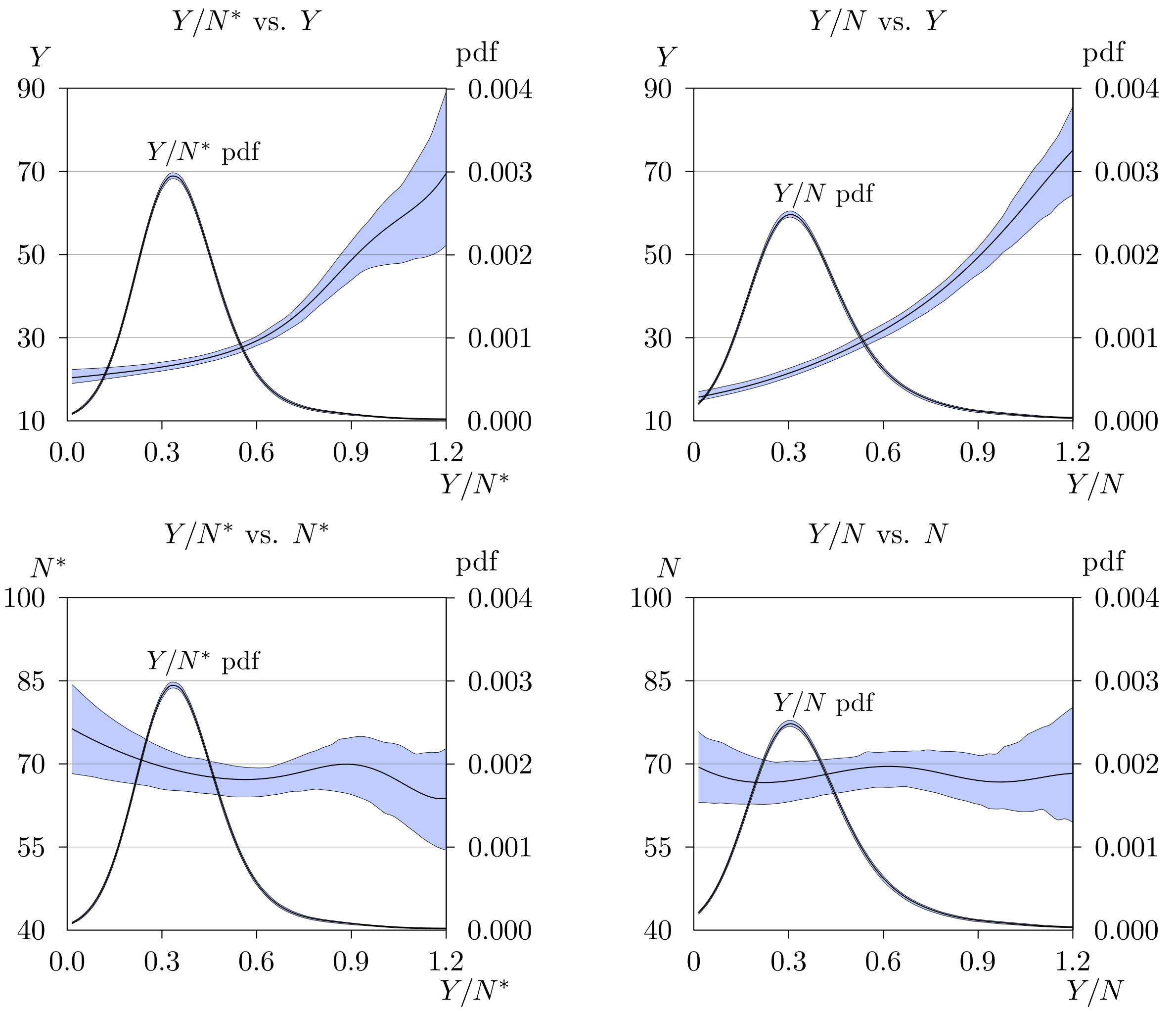
Table 1: Productivity – Size Correlations

	Employment (N)	Adjusted Employment (N*)	Value Added (Y)
Y/N	0.0017	0.0911	0.3138
Y/N*	-0.0095	-0.0176	0.1981

productivity. Interestingly, there is no correlation between the two in Danish data.

The correlations between the two measures of labor productivity with the two employment measures and sales as reflected in value added are reported in Table 1. As documented in the table, the correlation between labor force size and productivity using either the raw employment measure or the adjusted one is zero. However, note the strong positive associate between value added and both measures of labor productivity. Non-parametric regressions of value added and employment on the two productivity measures are illustrated in Figure 2. The top and bottom curves in the figures represent a 90% confidence interval for the relationship. The positive relationships between value added and both measure of labor productivity are highly significant.

Figure 2: Firm Size–Productivity Relationships.



Note: The shaded areas represent 90% bootstrap confidence intervals. Value added (Y) measured in 1 million DKK. N is the firm's labor force head count and N^* is the quality adjusted labor force size.

The theory developed in this paper is in part motivated by these observations. Specifically, it is a theory that postulates labor saving technical progress of a specific form. Hence, the apparent fact that more productive firms produce more with roughly the same labor input per unit of value added is consistent with the model.

3 An Equilibrium Model of Creative Destruction

As is well known, firms come is an amazing range of shapes and sizes. This fact cannot be ignored in any analysis of the relationship between firm size and productivity. Furthermore, an adequate theory must account for entry, exit and firm evolution in order to explain the size distributions observed. Klette and Kortum (2004) construct a stochastic model of firm product innovation and growth that is consistent with stylized facts regarding the firm size evolution and distribution. The model also has the property that technical progress is labor saving. For these reasons, we pursue their approach in this paper.

Although Klette and Kortum allow for productive heterogeneity, firm productivity and growth are unrelated because costs and benefits of growth are both proportional to firm productivity in their model. Allowing for a positive relationship between firm growth and productivity is necessary for consistency with the relationships found in the Danish firm data studied in this paper.

3.1 Preferences and Technology

Intertemporal utility of the representative household at time t is given by

$$U_t = \int_t^\infty \ln C_s e^{-\rho(s-t)} ds \tag{2}$$

where $\ln C_t$ denotes the instantaneous utility of the single consumption good at date t and ρ represents the pure rate of time discount. Each household is free to borrow or lend at interest rate r_t. Nominal household expenditure at date t is $E_t = P_t C_t$. Optimal consumption expenditure must solve the differential equation $\dot{E}/E = r_t - \rho$. Following Grossman and Helpman (1991), we choose the numeraire so that $E_t = 1$ for all t without loss of generality, which implies $r_t = r = \rho$ for all t. Note that this choice of the numeraire also implies that the price of the consumption good expressed in terms of the numeraire, P_t, falls over time at a rate equal to the rate of growth in consumption.

The quantity of the consumption produced is determined by the quantity and quality of the economy's intermediate inputs. Specifically, there is a unit continuum of inputs and consumption is determined by the production function

$$\ln C_t = \int_0^1 \ln(A_t(j)x_t(j))dj = \ln A_t + \int_0^1 \ln x_t(j)dj \tag{3}$$

where $x_t(j)$ is the quantity of input $j \in [0,1]$ at time t, $A_t(j)$ is the quality or productivity of input j at time t, and A_t represent aggregate productivity. The level of productivity of each input and aggregate productivity are determined by the number of technical improvements made in the past. Specifically,

$$A_t(j) = \prod_{i=1}^{J_t(j)} q_i(j) \text{ and } \ln A_t \equiv \int_0^1 \ln A_t(j)dj. \tag{4}$$

where $J_t(j)$ is the number of innovations made in input j up to date t and $q_i(j) > 1$ denotes the quantitative improvement (step size) in productivity attributable to the i^{th} innovation in product j. Innovations arrive at rate δ which is endogenous but the same for all intermediate products.

The model is constructed so that a steady state growth path exists with the following properties: Consumption output grows at a constant rate while the quantities of intermediate products and the endogenous innovation frequency are stationary and identical across all intermediate goods. As a consequence of the law of large numbers, the assumption that the number of innovations to date is Poisson with arrival frequency δ for all intermediate goods implies

$$\begin{aligned} \ln C_t &= \ln A_t + \int_0^1 \ln x(j)dj = \int_0^1 \sum_{i=1}^{J_t(j)} \ln q_i(j)dj + \int_0^1 \ln x(j)dj \\ &= E\ln(q)\delta t + \int_0^1 \ln x(j)dj. \end{aligned} \tag{5}$$

where $EJ_t(j) = \delta t$ for all j is the expected number of innovations per intermediate product and $E\ln(q) \equiv \int_0^1 \frac{1}{J_t(j)} \sum_{i=1}^{J_t(j)} \ln q_i(j)dj$ is the expected quality step size. In other words, consumption grows at the rate of growth in productivity which is the product of the creative-destruction rate and the expected log of the size of an improvement in productivity induced by each new innovation.

3.2 The Value of a Firm

Each individual firm is the monopoly supplier of the products it created in the past that have survived to the present. The price charged for each is limited by the ability of suppliers of previous versions to provide a substitute. In Nash-Bertrand equilibrium, any innovator takes over the market for its good type by setting the price just below that at which consumers are indifferent between the higher quality product supplied by the innovator and an alternative supplied by the last provider. The price charged is the product of relative quality and the previous producer's marginal cost of production. Given the symmetry of demands for the different good types and the assumption that future quality improvements are independent of the type of good, one can drop the good subscript without confusion. Given stationary of quantities along the equilibrium growth path, the time subscript can be dropped as well.

Labor is the only factor in the production of intermediate inputs. Labor productivity is the same across all inputs and is set equal to unity. Hence, $p = qw$ is the price in terms of the numeraire of every intermediate good as well as the value of labor productivity where w, the wage, represents the marginal cost of production of the previous supplier and $q > 1$ is the step up in quality of the innovation. As total expenditure is normalized at unity and there is a unit measure of product types, it follows that total revenue per product type is also unity given the specification of preferences and technology, i.e., $px = 1$. Hence, product output and employment are both equal to

$$x = \frac{1}{p} = \frac{1}{wq}. \tag{6}$$

and the gross profit associated with supplying the good is

$$1 > \pi = px - wx = 1 - \frac{1}{q} > 0. \tag{7}$$

The labor saving nature of improvements in intermediate input quality is implicit in the fact that labor demand is decreasing in q.

The model of quality improvements can equally well be viewed as a model of efficiency improvements, that is, a reduction of the amount of labor that is required to produce a unit of output. This is easily seen by re-interpreting the argument above in terms of quality units of output. Given that one unit of labor produces of unit of output, an increase in product quality of a unit of output is analogous to a reduction of the amount of labor that is required to produce a quality unit. In terms of quality units, the price is ever decreasing, demand for quality units of a product is ever increasing, and the amount of labor engaged in production in a given industry remains stable. In the short run, labor demand does fluctuate depending on the exact realization of the current lead that the industry leader has to the nearest follower - the greater the lead, the lower the demand.

Following Klette and Kortum (2004), the discrete number of products supplied by a firm, denoted as k, is defined on the integers and its value evolves over time as a birth-death process reflecting product creation and destruction. In their interpretation, k reflects the firm's past successes in the product innovation process as well as current firm size. New products are generated by R&D investment. The firm's R&D investment flow generates new product arrivals at frequency γk. The total R&D investment cost is $wc(\gamma)k$ where $c(\gamma)k$ represents the labor input required in the research and development process. The function $c(\gamma)$ is assumed to be strictly increasing and convex. According to the authors, the implied assumption that the total cost of R&D investment is linearly homogenous in the new product arrival rate and the number of existing product, "captures the idea that a firm's knowledge capital facilitates innovation." In any case, this cost structure is needed to obtain firm growth rates that are independent of size as typically observed in the data.

The market for any current product supplied by the firm is destroyed by the creation of a new version by some other firm, which occurs at the rate δ. Below we refer to γ as the firm's creation rate and to δ as the common destruction rate faced by all firms.[3]As product gross profit and product quality are one-to-one, the profits earned on each product reflects a firm's current labor productivity. The firm chooses the creation rate γ to maximize the expected present value of its future net profit flow.

Firms differ with the respect to the quality of their products. Hence, each type is characterized by profitability, π, as defined in equation (7). The value of the firm of type π that currently markets k products is

$$rV_k(\pi) = \max_{\gamma \geq 0} \big\{ [\pi - wc(\gamma)]\, k + \gamma k [V_{k+1}(\pi) - V_k(\pi)] + \delta k [V_{k-1}(\pi) - V_k(\pi)] \big\}. \tag{8}$$

The first term on the right side is current gross profit flow accruing to the firms product portfolio less current expenditure on R&D. The second term is the expected capital gain associated with the arrival of a new product line. Finally, the last term represents the expected capital loss associated with the possibility that one among the existing product lines will be destroyed.

The unique solution to (8) is proportional to the number of product lines. Formally,

$$V_k(\pi) = k \max_{\gamma \geq 0} \left\{ \frac{\pi - wc(\gamma)}{r + \delta - \gamma} \right\} \tag{9}$$

as one can verify by substitution. Consequently, any positive optimal choice of the product creation rate for a type π firm must satisfy the first order condition

$$wc'(\gamma(\pi)) = V_{k+1} - V_k = \max_{\gamma \geq 0} \left\{ \frac{\pi - wc(\gamma)}{r + \delta - \gamma} \right\}. \tag{10}$$

[3]These are in fact the continuous time job creation and job destruction rates respectively as defined in Davis, Haltiwanger, and Schuh (1996).

Hence, the second order condition, $c''(\gamma) > 0$, and the fact that the marginal value of a product line is increasing in π imply that the a firm's creation rate increases with profitability.

3.3 Firm Entry and Labor Market Clearing

The entry of a new firm requires innovation. Suppose that there are a constant measure m of potential entrants, identical ex ante. The rate at which any one of them generates a new product is γ_0 and the total cost is $wc(\gamma_0)$ where the cost function is the same as that faced by an incumbent. The firm's type is unknown ex ante but is realized immediately after entry. Since the expected return to innovation is $E_\pi\{V_1\}$ and the aggregate entry rate is $\eta = m\gamma_0$, the entry rate satisfies the following free entry condition

$$wc'\left(\frac{\eta}{m}\right) = \int_\pi V_1(\pi)\phi(\pi)d\pi = \int_\pi \max_{\gamma\geq 0}\left\{\frac{\pi - wc(\gamma)}{r+\delta-\gamma}\right\}\phi(\pi)d\pi \tag{11}$$

where $\phi(\pi)$ is the density entrant types. Of course, the second equality follows from equation (9).

There is a fixed measure of available workers, denoted by L, seeking employment at any positive wage. In equilibrium, these are allocated across production and R&D activities, those performed by both incumbent firms and potential entrants. Since the number of workers employed for production purposes per product of quality q is $x = 1/wq = (1-\pi)/w$ from equations (6) and (7), the total number demanded for production activity by firms of type π with k products is $L_x(k,\pi) = k(1-\pi)/w > 0$. The number of R&D workers employed by incumbent firms of type π with k products is $L_R(k,\pi) = kc(\gamma(\pi))$. Because each potential entrant innovates at frequency η/m, the aggregate number of worker engaged by all m in R&D is $L_E = mc(\eta/m)$. Hence, the equilibrium wage satisfies the labor market clearing condition

$$\begin{aligned} L &= \int_\pi \sum_{k=1}^{\infty}\left[L_x(k,\pi) + L_R(k,\pi)\right]M_k(\pi)d\pi + L_E \\ &= \int_\pi \left(\frac{1-\pi}{w} + c(\gamma(\pi))\right)\sum_{k=1}^{\infty} kM_k(\pi)d\pi + mc\left(\frac{\eta}{m}\right) \end{aligned} \tag{12}$$

where $M_k(\pi)$ represents the mass of firms of type π that supply k products.

3.4 The Steady State Distribution of Firm Size

Once a firm enters, its size as reflected in the number of product lines supplied evolves as a birth-death process. As the set of firms with k products at a point in time must either have had k

products already and neither lost nor gained another, have had $k-1$ and innovated, or have had $k+1$ and lost one to destruction over any sufficiently short time period, the equality of the flows into and out of the set of firms of type π with $k>1$ products requires

$$\gamma(\pi)(k-1)M_{k-1}(\pi)+\delta(k+1)M_{k+1}(\pi)=(\gamma(\pi)+\delta)kM_k(\pi)$$

for every π where $M_k(\pi)$ is the steady state mass of firms of type π that supply k products. Because an incumbent dies when its last product is destroyed by assumption but entrants flow into the set of firms with a single product at rate η,

$$\phi(\pi)\eta+2\delta M_2(\pi)=(\gamma(\pi)+\delta)M_1(\pi)$$

where $\phi(\pi)$ is the fraction of the new entrants that realize profit π. Births must equal deaths in steady state and only firms with one product are subject to death risk. Therefore, $\phi(\pi)\eta=\delta M_1(\pi)$ and

$$M_k(\pi)=\frac{k-1}{k}\gamma(\pi)M_{k-1}=\frac{\eta\phi(\pi)}{\delta k}\left(\frac{\gamma(\pi)}{\delta}\right)^{k-1} \tag{13}$$

by induction.

The size distribution of firms conditional on type can be derived using equation (13). Specifically, the total firm mass of type π is

$$\begin{aligned} M(\pi) &= \sum_{k=1}^{\infty}M_k(\pi)=\frac{\phi(\pi)\eta}{\delta}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\gamma(\pi)}{\delta}\right)^{k-1} \\ &= \frac{\eta}{\delta}\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)\frac{\delta\phi(\pi)}{\gamma(\pi)}. \end{aligned} \tag{14}$$

where convergence requires that the aggregate rate of creative destruction exceed the creation rate of every incumbent type, i.e., $\delta>\gamma(\pi)\ \forall\pi$. Hence, the fraction of type π firm with k product is

$$\frac{M_k(\pi)}{M(\pi)}=\frac{\frac{1}{k}\left(\frac{\gamma(\pi)}{\delta}\right)^k}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)}. \tag{15}$$

This is the logarithmic distribution with parameter $\gamma(\pi)/\delta$.[4] Consistent with the observations on firm size distributions, that implied by the model is highly skewed to the right.

[4]This result is in Klette and Kortum (2004). We include the derivation here simply for completeness.

By equation (15), the mean of the firm size distribution conditional on product profitability is

$$E\left[k|\pi\right] = \sum_{k=1}^{\infty} \frac{kM_k(\pi)}{M(\pi)} = \frac{\frac{\gamma(\pi)}{\delta-\gamma(\pi)}}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)}, \tag{16}$$

As the product creation rate increases with profitability, expected size does also. Formally, because $(1+a)\ln(1+a) > a > 0$ for all positive values of a, the expected number of products is increasing in firm profitability,

$$\frac{\partial E[k|\pi]}{\partial \pi} = \left(\frac{(1+a(\pi))\ln(1+a(\pi)) - a(\pi)}{(1+a(\pi))\ln^2(1+a(\pi))}\right)\frac{\delta\gamma'(\pi)}{(\delta-\gamma(\pi))^2} > 0 \tag{17}$$

where $a(\pi) = \frac{\gamma(\pi)}{\delta-\gamma(\pi)}$.

Although more profitable firms supply more products, total expected employment, $nE[k]$ where $n = (1-\pi)/w + c(\gamma(\pi))$, need not increase with π in general and decreases with π if innovation is not related to profitability because innovation is labor saving. Hence, the hypothesis that firms with the ability to create products of better quality grow faster is consistent with dispersion in labor productivity and the correlations between value added, labor force size, and labor productivity observed in Danish data reported above.

Finally, the rate of creative-destruction is the sum of the entry rate and the aggregate creation rates of all the incumbents given that the total mass of products is fixed. Because the new product arrival rate of a firm of type π with k products is $\gamma(\pi)k$ and the measure of such firms is $M_k(\pi)$,

$$\delta = \eta + \int_{\pi} \gamma(\pi) \sum_{k=1}^{\infty} kM_k(\pi)d\pi. \tag{18}$$

3.5 Equilibrium

Definition A *steady state market equilibrium* is a triple composed of a labor market clearing wage w, entry rate η, and creative destruction rate δ together with an optimal creation rate $\gamma(\pi)$ and a steady state size distribution $M_k(\pi)$ for each type that satisfy equations (11), (12), (10), (13), and (18) provided that $\gamma(\pi) < \delta$, for every π in the support of the entry distribution.

Proposition If the cost of innovation, $c(\gamma)$, is strictly convex and $c'(0) = c(0) = 0$, then a steady state market equilibrium with positive entry exists. In the case of a single firm type, there is only one.

Proof. See Lentz and Mortensen (2005).[5] ∎

4 Estimation

If product quality is a permanent firm characteristic, then differences in firm profitability are associated with differences in the product creation rates chosen by firms. Specifically, more profitable firms grow faster, are more likely to survive in the future, and supply a larger number of products on average. Hence, a positive cross firm correlation between current gross profit per product and sales volume should exist. Furthermore, worker reallocation from slow growing firms that supply products of lesser quality to more profitable fast growing firms will be an important sources of aggregate productivity growth. On the other hand, if product quality were iid across innovations and firms, all firms grow at the same rate even though persistent differences in profitability exist as a consequence of different realizations of product quality histories.

In this section, we demonstrate that firm specific differences in profitability are required to explain Danish interfirm relationships between value added, employment, and wages paid. In the process of fitting the model to the data, we also obtain estimates of the investment cost of innovation function that all firms face as well as the sampling distribution of firm productivity at entry.

4.1 Danish Firm Data

If more productive firm's grow faster in the sense that $\gamma'(\pi) > 0$, then (17) implies that more productive firms also supply more products and sell more on average. However, because production employment per product decreases with productivity, total expected employment, $nE[k]$ where $n = (1-\pi)/w + c(\gamma(\pi))$, need not increase with π in general and decreases with π when growth is independent of a firm's past product quality realizations. These implications of the theory can be tested directly.

The model is estimated on an unbalanced panel of 4,872 firms drawn from the Danish firm panel described in Section 2. The panel is constructed by selecting all existing firms in 1992 with more than 20 workers and following them through time, while all firms that enter the sample in the

[5] Although the cost of entry is linear in the paper cited while the cost is convex here, the principal argument holds in this case as well.

subsequent years are excluded. In the estimation, the observed 1992 cross-section will be interpreted to reflect steady state whereas the following years generally do not reflect steady state since survival probabilities vary across firm types. Specifically, due to selection the observed cross-sections from 1993 to 1997 will have an increasing over-representation of high creation rate firm types relative to steady state. Entry in the original data set suffers from strong selection bias and the sampling choice to leave out entry altogether is consequently partly driven by data limitations but is also useful in identifying dynamic features of the model. Table 2 presents a number of data moments with standard deviations in parenthesis. The standard deviations are obtained by bootstrapping. Unless otherwise stated, nominal amounts are in 1,000 DKK.

The dynamic moments relating to firm growth rates ($\Delta Y/Y$) include firm death, so specifically an exiting firm will contribute to the statistic with a -1 observation. Should one exclude firm deaths from the growth statistic, one will obtain a more negative correlation between firm size and growth due to the strong negative correlation between firm size and the firm exit hazard rate.

In addition to the moments in Table 2, the model will also be estimated against a standard reduced form labor productivity growth decomposition. We use the preferred formulation in Foster, Haltiwanger, and Krizan (2001) which is taken from Baily, Bartelsman, and Haltiwanger (1996). The decomposition takes the form,

$$\begin{aligned}\Delta P_t \;=\; & \sum_{e\in C} s_{et-1}\Delta p_{et} + \sum_{e\in C}\left(p_{et-1}-P_{t-1}\right)\Delta s_{et} + \sum_{e\in C}\Delta p_{et}\Delta s_{et} + \sum_{e\in N}\left(p_{et}-P_{t-1}\right)s_{et} - \\ & \sum_{e\in X}\left(p_{et-1}-P_{t-1}\right)s_{et-1}, \qquad (19)\end{aligned}$$

where $P_t = \sum_e s_{et}p_{et}$, $p_{et} = Y_{et}/N_{et}$, and $s_{et} = N_{et}/N_t$. Thus, (19) will be used to decompose growth in value added per worker into 5 components in the order stated on the right hand side; within, between, a cross component, and entry and exit. The within component is interpreted to capture growth in the productivity measure due to productivity improvements by incumbents, the between component is designed to capture productivity growth from reallocation of labor from less to more productive firms. The cross component captures a covariance between share of input and productivity growth and the last two terms capture the growth contribution from entrants and exits. The decomposition shares in the data are shown in Table 3. As mentioned, the sample in this paper does not include entry, so there is no entry share in the decomposition and the

Table 2: Data Moments (std dev in parenthesis)

	1992	1997		1992	1997
Survivors	4,872.00	3,628.00	$Cor\left[\frac{Y}{N^*}, \frac{Y_{+1}}{N^*_{+1}}\right]$	0.476	0.550
	-	(32.13)		(0.088)	(0.091)
$E[Y]$	26,277.26	31,860.85	$Cor\left[\frac{Y}{N^*}, \Delta\frac{Y}{N^*}\right]$	-0.227	-0.193
	(747.00)	(1,031.25)		(0.103)	(0.057)
$Med[Y]$	13,471.00	16,432.10	$Cor[Y, W]$	0.852	0.857
	(211.35)	(329.77)		(0.035)	(0.045)
$E[W]$	13,294.48	15,705.09	$Cor\left[\frac{Y}{N^*}, Y\right]$	0.198	0.143
	(457.47)	(609.60)		(0.036)	(0.038)
$Med[W]$	7,229.70	8,670.28	$Cor\left[\frac{Y}{N^*}, N^*\right]$	-0.018	-0.026
	(92.75)	(154.90)		(0.013)	(0.011)
$Std[Y]$	52,798.52	64,129.07	$E[\Delta Y/Y]$	-0.029	
	(5,663.63)	(7,742.51)		(0.008)	
$Std[W]$	30,616.94	35,560.60	$Std[\Delta Y/Y]$	0.550	
	(6,751.09)	(8,138.66)		(0.067)	
$E\left[\frac{Y}{N^*}\right]$	384.40	432.12	$Cor[\Delta Y/Y, Y]$	-0.061	
	(2.91)	(5.10)		(0.012)	
$Std\left[\frac{Y}{N^*}\right]$	205.09	305.35			
	(19.63)	(42.50)			

decomposition shares in Table 3. Consequently, the decomposition cannot be directly related to the results in Foster, Haltiwanger, and Krizan (2001), although a full decomposition is performed on the estimated model in section 4.5.2.

The decomposition provides additional information on dynamics in the data and is therefore valuable for identification purposes. But it is also a useful method of directly relating the model to a standard reduced form measure of sources of productivity growth. In section 5, we determine the labor productivity growth rate and the structural decomposition for the estimated model. Labor productivity growth in the model is only loosely related to growth in value added per worker and consequently, there is no reason to expect the decomposition in equation (19) to coincide with the structural decomposition in section 5.

Table 3: Y/N 1992 to 1997 Growth Decomposition. Std Dev in parentheses.

	Growth Shares
Within	1.015 (0.146)
Between	0.453 (0.112)
Cross	-0.551 (0.196)
Exit	0.084 (0.066)

4.2 Model Estimator

An observation in the panel is given by $\psi_{it} = (Y_{it}, W_{it}, N_{it}^*)$, where Y_{it} is real value added, W_{it} the real wage sum, and N_{it}^* quality adjusted labor force size of firm i in year t. Let ψ_i be defined by, $\psi_i = \left(\psi_{i1,\ldots,}\psi_{iT}\right)$ and finally, $\psi = (\psi_1, \ldots, \psi_I)$

Simulated minimum distance estimators, as described in for example Gourieroux, Monfort, and Renault (1993), Hall and Rust (2003), and Alvarez, Browning, and Ejrnæs (2001), are computed as follows: First, define a vector of auxiliary data parameters, $\Gamma(\psi)$. The vector consists of all the items in Table 2 except the number of survivors in 1992 and three of the moments in table 3. Thus, $\Gamma(\psi)$ has length 33. Second, $\psi^s(\omega)$ is simulated from the model for a given set of model parameters ω. The model simulation is initialized by assuming that the economy is in steady state in the first year and consequently that firm observations are distributed according to the ω-implied steady state distribution. Alternatively, one can initialize the simulation according to the observed data in the first year, $(\psi_{11}, \ldots, \psi_{1I})$. The assumption that the economy is initially in steady state provides additional identification in that $(\psi_{11}, \ldots, \psi_{1I})$ can be compared to the model-implied steady state distribution $(\psi_{11}^s(\omega), \ldots, \psi_{1I}^s(\omega))$. The simulated auxiliary parameters are then given by,

$$\Gamma^s(\omega) = \frac{1}{S}\sum_{s=1}^{S}\Gamma(\psi^s(\omega)),$$

where S is the number of simulation repetitions.

The estimator is then the choice of parameters that minimizes the weighted distance between the data and simulated auxiliary parameters,

$$\hat{\omega} = \arg\min_{\omega\in\Omega}\left(\Gamma^s(\omega) - \Gamma(\psi)\right)' A^{-1}\left(\Gamma^s(\omega) - \Gamma(\psi)\right), \tag{20}$$

where A is some positive definite matrix. If A is the identity matrix, $\hat{\omega}$ is the equally weighted minimum distance estimator (EWMD). If A is the covariance matrix of the data moments $\Gamma(\psi)$, $\hat{\omega}$ is the optimal minimum distance estimator (OMD). The OMD estimator is asymptotically more efficient than the EWMD estimator. However, Altonji and Segal (1996) show that the estimate of A as the second moment matrix of $\Gamma(\cdot)$ may suffer from serious small sample bias. Horowitz (1998) suggests a bootstrap estimator of A. The estimation in this paper adopts Horowitz's bootstrap estimator of the covariance matrix A.

In addition to the $\hat{\omega}$ estimator, the analysis also presents a bootstrap estimator as in Horowitz (1998). In each bootstrap repetition, a new set of data auxiliary parameters $\Gamma\left(\psi^b\right)$ is produced, where ψ^b is the bootstrap data in the b^{th} bootstrap repetition. ψ^b is found by randomly selecting observations ψ_i from the original data with replacement. Thus, the sampling is random across firms but is done by block over the time dimension (if a particular firm i is selected, the entire time series for this firm is included in the sample). For the b^{th} repetition, an estimator ω^b, is found by minimizing the weighted distance between the re-centered bootstrap data auxiliary parameters $\left[\Gamma\left(\psi^b\right) - \Gamma(\psi)\right]$ and the re-centered simulated auxiliary parameters $\left[\Gamma^s\left(\omega^b\right) - \Gamma^s(\hat{\omega})\right]$,

$$\omega^b = \arg\min_{\omega\in\Omega}\Big(\left[\Gamma^s(\omega) - \Gamma^s(\hat{\omega})\right] - \left[\Gamma(\psi^b) - \Gamma(\psi)\right]\Big)' A^{-1}\Big(\left[\Gamma^s(\omega) - \Gamma^s(\hat{\omega})\right] - \left[\Gamma(\psi^b) - \Gamma(\psi)\right]\Big).$$

In each bootstrap repetition, a different seed is used to generate random numbers for the determination of $\Gamma^s(\omega)$. Hence, the bootstrap estimator of $V(\hat{\omega})$ captures both data variation and variation from the model simulation.

The bootstrap estimator of the structural parameters is then the simple average of all the ω^b estimators,

$$\hat{\omega}^{bs} = \frac{1}{B}\sum_{b=1}^{B}\omega^b, \tag{21}$$

where B is the total number of bootstrap repetitions. In the estimation below, $B = 500$ and $S = 10$.

4.3 Model Simulation

To fit the data, the model simulation produces time paths for value added (Y), the wage sum (W), and labor force size (N) for I firms. The estimation introduces a stochastic demand realization for each of a firm's products, $\tilde{Z}$. Thus, the demand for product j is given by $x_j = \tilde{Z}/p_j$. The random variable, $\tilde{Z}$, is iid across products and is assumed to follow a log-normal distribution,

$$\tilde{Z} = \exp\left(\tilde{\xi}\sigma_z + \mu_z\right) \text{ where } \tilde{\xi} \sim N(0,1). \tag{22}$$

Denote the expected value of $\widetilde{Z}$ by $E\left[\tilde{Z}\right] = Z$. Given the formulation of the firm's problem, the innovation rate is affected by the $\tilde{Z}$ distribution only through the expectation of $\tilde{Z}$ and not by any of the higher order moments.

To properly capture the labor share in the data, a capital cost $\kappa \equiv K/Z$ is added to the model where K is the capital associated with the production of a given product and κ is the capital cost relative to average product expenditure. This modifies the pricing of the intermediary goods. Now, providing an intermediary good at price p yields expected operational profits, $Z(1 - w/p - \kappa)$. Thus, the price of intermediary good j is, $p = qw/(1-\kappa)$ since consumers are exactly indifferent between buying from the quality leader at this price and the from the immediate follower at price $p = w/(1-\kappa)$, which is as low as the follower is willing to go. The inclusion of a non-labor cost then modifies the definition of production profits, π, as defined in (7). The more general definition that allows for non-labor cost is given by,

$$\pi = (1-\kappa)(1-q^{-1}), \tag{23}$$

which is identical to (7) if $\kappa = 0$.

The quality of each new innovation (and thereby the profit associated with it) is a stochastic realization drawn from a distribution which is contingent on the firm's type. Specifically, the profit of any particular innovation is assumed to satisfy

$$\widetilde{\pi} = (1-\kappa)(1-\widetilde{q}^{-1}), \text{ where } \widetilde{q} = 1 + \exp\left(\xi\sigma_{\widetilde{\pi}} + \mu_{\widetilde{\pi}}(\pi)\right) \text{ and } \xi \sim N(0,1). \tag{24}$$

where the mean $E[\widetilde{\pi}|\pi] = \pi$ represents the firm's profitability type, the determinant of its creation rate. Each firm's type is itself a random variable realized after entry. We assume that the steady state distribution profit distribution, denoted as $p(\pi)$, is characterized by

$$\pi = (1-\kappa)\left(1-q^{-1}\right), \text{ where } q = 1+\exp\left(z\sigma_\pi + \mu_\pi\right) \text{ and } z \sim N\left(0,1\right) \tag{25}$$

where in both (24) and (25) $N(0,1)$ represents the standard normal distribution. For future reference,

$$p(\pi) \equiv \frac{M(\pi)}{\int_\pi M(\pi) d\pi} \tag{26}$$

where $M(\pi)$, the steady state mass of firms of type, is given by equation (14).

Denote by $\Pi^k = (\pi_1, \ldots, \pi_k)$ the quality realizations of a firm's k products. Similarly let $Z^k = (Z_1, \ldots, Z_k)$ be the demand realizations of the firm's k products. The value added of a type π firm with k products characterized by $\left(\Pi^k, Z^k\right)$ is given by,

$$Y_k\left(\Pi^k, Z^k, \pi\right) = \sum_{i=1}^{k} Z_i, \tag{27}$$

where each product demand realization Z_i is drawn according to (22). The wage bill is given by,

$$W_k\left(\Pi^k, Z^k, \pi\right) = \sum_{i=1}^{k} Z_i\big(1-\kappa-\pi_i\big) + kZw\tilde{c}\big(\gamma(\pi)\big), \tag{28}$$

where $\widehat{c}\left(\gamma\right) = c\left(\gamma\right)/Z$.

The estimation allows for measurement error in both value added and the wage bill. The measurement error is introduced as a simple log-additive process,

$$\begin{aligned} \ln \tilde{Y}_k\left(\Pi^k, Z^k, \pi\right) &= \ln Y_k\left(\Pi^k, Z^k, \pi\right) + \xi_Y \\ \ln \tilde{W}_k\left(\Pi^k, Z^k, \pi\right) &= \ln W_k\left(\Pi^k, Z^k, \pi\right) + \xi_W, \end{aligned}$$

where $\xi_Y \sim N\left(0, \sigma_Y^2\right)$ and $\xi_W \sim N\left(0, \sigma_W^2\right)$. The estimation is performed on the quality adjusted labor force size. Consequently, the wage bill measurement error is assumed to carry through to the labor force size, $\tilde{N}_k\left(\Pi^k, Z^k, \pi\right) = \tilde{W}_k\left(\Pi^k, Z^k, \pi\right)/w$ since by construction, $N_i^* w = W_i$ for all firms in the data.

Lentz and Mortensen (2005) analyze the firm's creation rate choice in the general case where product quality is a stochastic process. Because the value of the next product is linear in profit and the profit realizations across products are iid for each firm, the optimal choice of creation rate for a firm of type π solves,

$$\gamma(\pi) = \arg\max_{\gamma} \frac{E[\widetilde{\pi}] - w\widehat{c}(\gamma)}{r + \delta - \gamma} = \arg\max_{\gamma} \frac{\pi - w\widehat{c}(\gamma)}{r + \delta - \gamma} \tag{29}$$

as in the deterministic case sketch above. Specify the cost function as $\widehat{c}(\gamma) = c_0\gamma^{1+c_1}$. Then, the first order condition for the optimal creation rate choice is,

$$w(1 + c_1) c_0\gamma^{c_1}(r + \delta - \gamma) = \pi - wc_0\gamma^{1+c_1}. \tag{30}$$

Equations (27) and (28) provide the foundation for the model simulation. It then remains to simulate product paths for all firms. The simulation is initialized by the assumption of steady state. By (15), the steady state product size distribution conditional on survival is given by,

$$\Pr(k^* = k|\pi) = \frac{M_k(\pi)}{M(\pi)} = \frac{\frac{1}{k}\left(\frac{\gamma(\pi)}{\delta}\right)^k}{\ln\left(\frac{\delta}{\delta - \gamma(\pi)}\right)}. \tag{31}$$

First, a firm's type, π, is determined according to (25). Then, the initial product size of a firm (k_1) is determined following (31).

With a given initial product size, simulation of the subsequent time path requires knowledge of the transition probability function $\Pr(k_2 = k|k_1, \pi)$. Denote by $p_{\pi,n}(t)$ the probability of a type π firm having product size n at time t. As shown in Klette and Kortum (2004), $p_{\pi,n}(t)$ evolves according to the ordinary differential equation system,

$$\begin{aligned} \dot{p}_{\pi,n}(t) &= (n-1)\gamma(\pi)p_{\pi,n-1}(t) + (n+1)\delta p_{\pi,n+1}(t) - (\delta + \gamma(\pi))p_{\pi,n}(t), \ \forall n \geq 1 \\ \dot{p}_{\pi,0}(t) &= \delta p_{\pi,1}(t). \end{aligned} \tag{32}$$

Hence, with the initial condition,

$$p_{\pi,n}(0) = \begin{cases} 1 \text{ if } n = k_1 \\ 0 \text{ otherwise.} \end{cases} \tag{33}$$

one can determine $\Pr(k_2 = k|k_1, \pi)$ by solving the differential equation system in (32) for $p_{\pi,k}(1)$. Solving for $p_{\pi,k}(1)$ involves setting an upper reflective barrier to bound the differential equation system. It has been set sufficiently high so as to avoid biasing the transition probabilities. Based on the transition probabilities $\Pr(k_{t+1} = k|k_t, \pi)$ one can then iteratively simulate product size paths for each firm. The procedure correctly captures the evolution of k_t but it does not identify the exact evolution of $\left(\Pi^{k_t}, Z^{k_t}\right)$. The evolution of $\left(\Pi^{k_t}, Z^{k_t}\right)$ is assumed to follow the net change in products.[6]

Finally, the simulation allows for an exogenous growth factor in both value added and the wage bill, denoted as $\widehat{g}$, that is independent of the endogenous quality improvements produced by incumbents and entrants.

4.4 Identification

The set of model parameters to be identified (ω) is given by,

$$\omega = \left\{c_0, c_1, \delta, \kappa, Z, \sigma_z^2, \sigma_{\widetilde{\pi}}^2, \sigma_\pi^2, \mu_\pi, \widehat{g}, \sigma_Y^2, \sigma_W^2\right\} \in \Omega,$$

where Ω is the feasible set of model parameters choices. The interest rate will be set at $r = .05$. The wage w is immediately identified as the average worker wage in the sample $w = 190.24$. Experimentation with non-parametric identification of the firm type distribution has been performed with a distribution with 4 support points. Because the results showed little sensitivity in the remaining model parameters to this alternative specification, we report only those parameters obtained given the assumed parametric distribution of types.

4.5 Estimation Results

The model parameter estimates are given in Table 4. Table 5 produces a comparison of the data moments and the simulated moments associated with the model parameter estimates.

The estimated model does well in fitting the labor productivity distribution and the correlations between productivity and firm size. These relationships are also shown in Figure 3. Notice that the model has not been fitted to the higher order moments of these relationships but fits them quite well nonetheless.

The estimation implies a significant level of firm type heterogeneity. In Table 4, it is expressed via the distribution of the firm's expected quality improvement of an innovation. The type dis-

[6] Suppose firm i is simulated to lose one product in a given year. In this case, $\left(\Pi^{k_{it}}, Z^{k_{it}}\right)$ is updated by randomly eliminating one element from it. This assumes that the net loss of one product took place by the gross destruction of one product and zero gross creation. This is the most likely event by which the firm loses one product. However, the net loss could also come about by the gross destruction of two products and gross creation of one product during the year. In this case, $\left(\Pi^{k_{it}}, Z^{k_{it}}\right)$ should be updated by randomly eliminating two elements and adding one. There are in principle an infinite number of ways that the firm can loose one product over the year. The estimation consequently over-estimates the persistency of $\left(\Pi^{k_{it}}, Z^{k_{it}}\right)$. The bias will go to zero as the period length is reduced, though.

Table 4: Model Parameter Estimates

	Point Estimate	Bootstrap Estimator	Std Deviation
c_0	595.2774	598.6864	50.8455
c_1	4.4186	4.4224	0.0785
κ	0.4420	0.4403	0.0055
Z	17,024.5242	17,053.0482	428.3154
δ	0.0794	0.0791	0.0029
σ_z^2	0.9138	0.8612	0.0487
$\sigma_{\tilde{\pi}}^2$	2.3317	2.0939	0.3692
μ_π	-4.6093	-4.7304	0.3004
σ_π^2	5.9086	6.8098	0.7993
$\widehat{g}$	0.0163	0.0166	0.0012
σ_Y^2	0.0114	0.0093	0.0047
σ_W^2	0.0283	0.0293	0.0039
Inferred Estimates			
η	0.0456	0.0448	0.0018
m	1.2370	1.2116	0.0973
M	0.7174	0.7069	0.0209
L	44.8899	44.6207	1.2804
$\bar{\gamma}$	0.0338	0.0344	0.0029
Entry q-distribution			
10^{th} percentile	1.0004	1.0003	0.0001
Median	1.0072	1.0064	0.0018
90^{th} percentile	1.1327	1.1402	0.0261
Steady state q-distribution			
10^{th} percentile	1.0004	1.0003	0.0001
Median	1.0099	1.0091	0.0025
90^{th} percentile	1.2206	1.2410	0.0462

tribution at entry is such that the median firm expects to produce a 0.72% quality improvement upon discovering an innovation. The 90^{th} percentile firm expects a 13.27% improvement. The heterogeneity in creation rates across types is reflected in the steady state distribution where the high type firms are over-represented relative to the entry distribution to the point where the 90^{th} percentile firm in steady state expects a 22.06% quality improvement when it innovates. We have experimented with more flexible choices of type distributions and have found the current choice to be non-restrictive.

Given the steady state equilibrium definition, one can infer the overall entry rate, η, and the measure of potential entrant, m.[7] The implied values of these parameters are also reported in Table 4. The average incumbent creation rate, $\bar{\gamma}$, is simply the difference between the entry rate

[7] The formulas used to make the calculations are presented in the appendix.

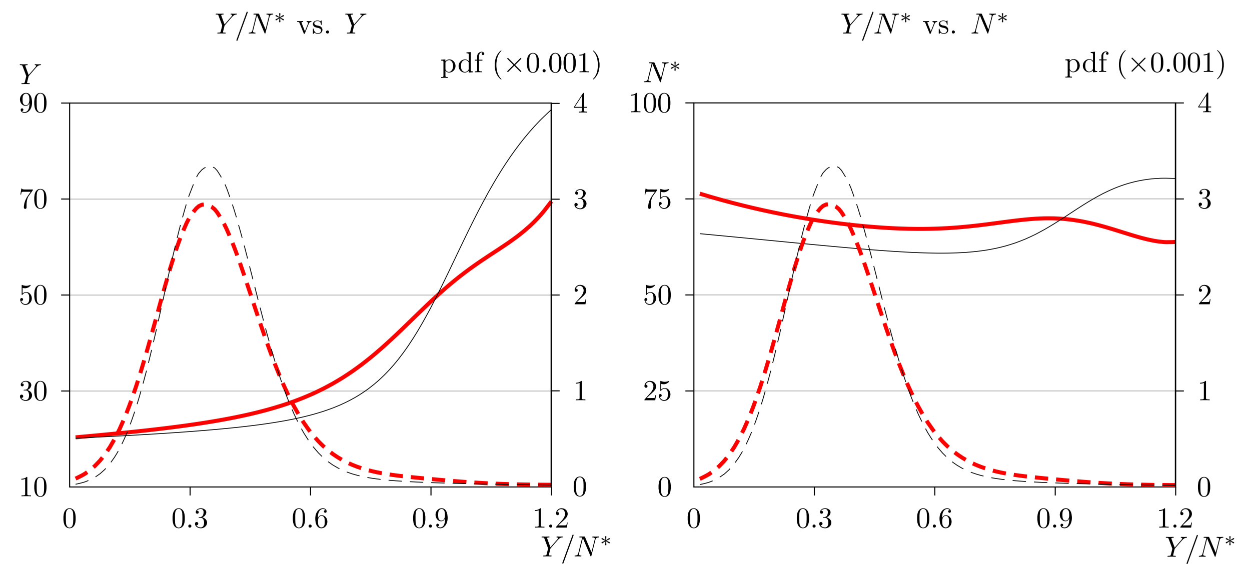
Table 5: Model Fit

	Data		Simulated Model	
	1992	1997	1992	1997
Survivors	4,872.000	3,628.000	4,872.000	3,594.300
$E[Y]$	26,277.262	31,860.851	23,832.346	28,088.419
$Med[Y]$	13,471.000	16,432.098	13,536.529	15,718.961
$E[W]$	13,294.479	15,705.087	11,976.439	13,868.172
$Med[W]$	7,229.704	8,670.279	7,146.571	8,234.080
$Std[Y]$	52,798.524	64,129.072	39,536.429	46,974.187
$Std[W]$	30,616.944	35,560.602	15,439.476	17,810.208
$E\left[\frac{Y}{N^*}\right]$	384.401	432.118	384.046	421.007
$Std\left[\frac{Y}{N^*}\right]$	205.095	305.348	199.185	216.350
$Cor\left[\frac{Y}{N^*}, \frac{Y_{+1}}{N^*_{+1}}\right]$	0.476	0.550	0.798	0.793
$Cor\left[\frac{Y}{N^*}, \Delta\frac{Y}{N^*}\right]$	-0.227	-0.193	-0.295	-0.312
$Cor[Y, W]$	0.852	0.857	0.855	0.853
$Cor\left[\frac{Y}{N^*}, Y\right]$	0.198	0.143	0.207	0.230
$Cor\left[\frac{Y}{N^*}, N^*\right]$	-0.018	-0.026	-0.021	-0.011
$E[\Delta Y/Y]$	-0.029	–	0.011	–
$Std[\Delta Y/Y]$	0.550	–	0.844	–
$Cor[\Delta Y/Y, Y]$	-0.061	–	-0.029	–
Growth decomp.				
– Within	1.015	–	0.939	–
– Between	0.453	–	0.350	–
– Cross	-0.551	–	-0.429	–
– Exit	0.084	–	0.140	–

and the destruction rate. It is seen that the estimates imply that more than half of all innovation comes from entrants. Given the estimated steady state distribution of firms, $p(\pi)$, and the other parameters of the model, one can also infer the ex ante type distribution, $\phi(\pi)$. The cdf's of the two distributions are shown in Figure 4 along with the incumbent creation rate choice conditional on firm type. It is clear from the figure that the higher quality type firms choose higher creation rates and consequently grow faster. Therefore, those with better products will make up a larger fraction of firms in steady state relative to their shares at entry. The consequences of this fact for aggregate growth are explored more fully below.

The estimation is performed given the assumption that the true firm population of interest coincides with the size censoring in the data. That is, the estimation does not correct for size censoring bias. While this is obviously a strong assumption, it reasonable assumption that the large number of very small firms in the economy are qualitatively different from those in this

Figure 3: Firm Productivity and Size, 1992 (Data and Simulation).



Note: Observed relationships drawn in bold pen and estimated relationships drawn in thin pen. Value added measured in 1 million DKK.

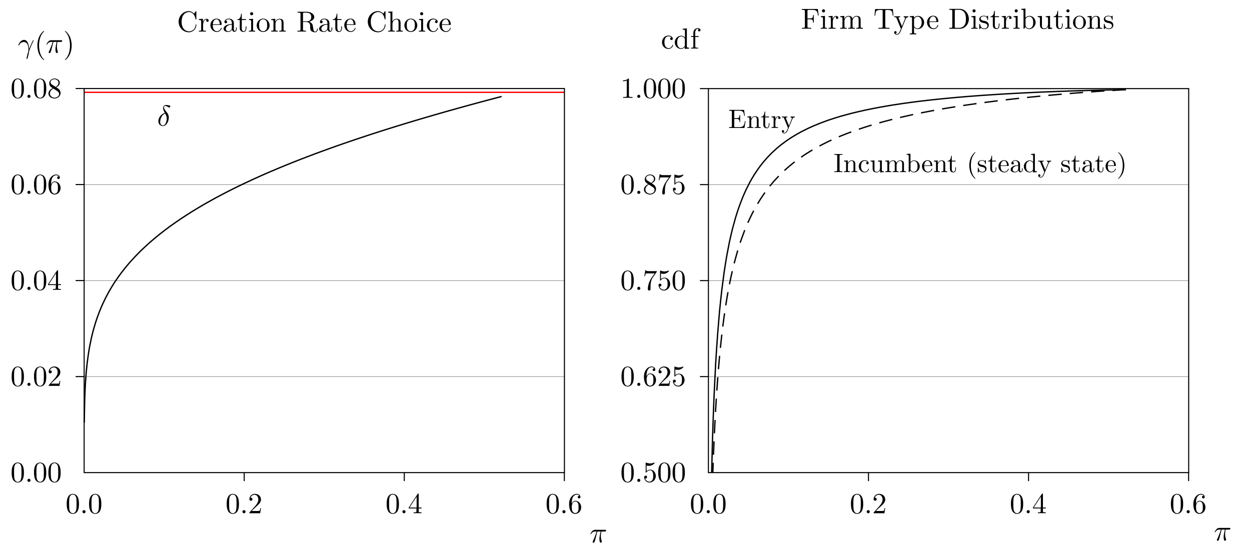
analysis and are not just firms with fewer products.

The estimation explicitly includes a number of dynamic moments. In addition, it should be noted that since the estimation is performed on cross-section moments not just in 1992 but also in 1997 and because of the specific sampling procedure in the data, the estimation implicitly address dynamic features of the model. The trends in the moments over time are in part interpreted as a result of systematic selection bias due to creation rate heterogeneity across types.

Size Distributions The model captures the medians of the Y and W distributions, but underestimate the mean and the variance. Thus, the model is not quite capturing the heaviness of the right-tail of the size distributions. This can likely be remedied by a more flexible choice of demand and supply shock processes.

The dispersion estimate is a result of a combination of the stochastic nature of the birth-death process of products, the demand shock process, and to a lesser extend the measurement error processes. Model simulation without measurement error ($\sigma_Y^2 = \sigma_W^2 = 0$) yields a reduction in the 1992 value added standard deviation estimate from $39,536.43$ to $37,697.31$. A model simulation

Figure 4: Creation Rate Choice and Firm Type Distributions.



Note: Type distribution at entry drawn in solid pen. The steady state type distribution drawn in dashed line.

with no demand shocks ($\sigma_Z^2 = 0$) yields a reduction in the 1992 value added standard deviation estimate from $39,536.43$ to $28,838.20$.

Productivity–Size Correlations Type heterogeneity and supply side shocks, σ_π^2 and $\sigma_{\tilde{\pi}}^2$ respectively, play an important role in explaining the productivity – size correlations. Type heterogeneity provides the foundation for a positive correlation between productivity and output size through a greater product creation rate for higher productivity type firms. The overall heterogeneity in product quality realizations both through type heterogeneity and random quality realizations within types explains the difference between the productivity – input size correlation and the productivity – output size correlation. Together σ_π^2 and $\sigma_{\tilde{\pi}}^2$ are chosen to get the exact levels of the correlations right. Measurement error has the potential of explaining these correlations as well. The estimation allows for both input and output measurement error which are estimated at fairly moderate amounts. If the model is simulated without the measurement error ($\sigma_Y^2 = \sigma_W^2 = 0$), the 1992 size–productivity correlations change to $corr\,(Y/N, Y) = 0.210$ and $corr\,(Y/N, N) = 0.0190$. Thus, measurement error is estimated to have virtually no impact on these moments in the data. Rather, these moments are explained to be a result of the labor saving innovation process at the heart of the model.

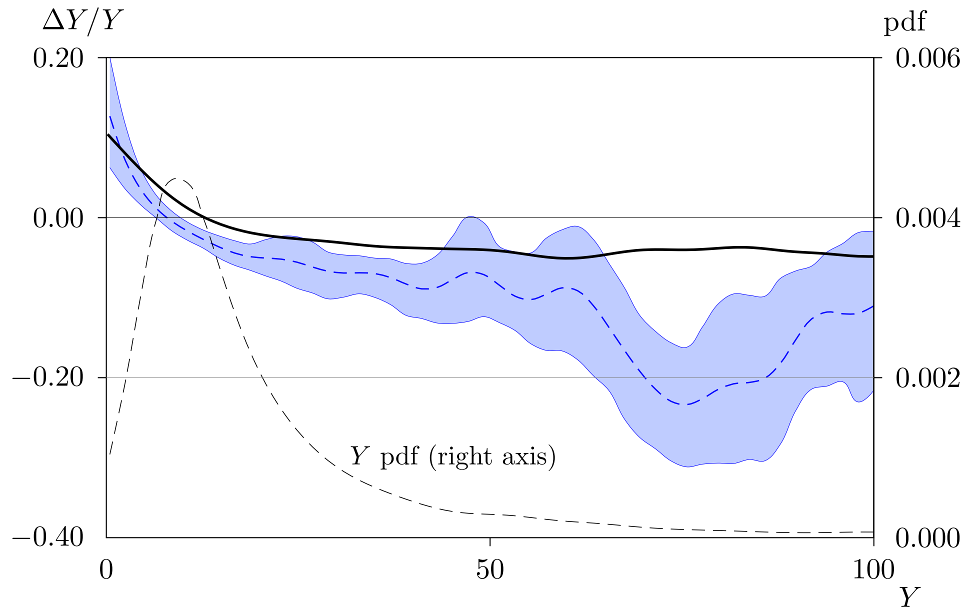
Right-Shift of Size Distributions Notice that the model successfully captures the right shift of the Y and W distributions of survivors from 1992 to 1997. There are three effects that contribute to the right shift: Generally, since the sampling eliminates the flow in of entrants, the model predicts a general decrease in mass of firms of all product sizes and types, $M_k(\pi)$, since all firms face an overall negative product growth rate. However, since entrants are assumed to flow in from the lower end of the size distribution, the reduction in mass is relatively stronger at the lower end and consequently the size distribution of survivors will begin to place relatively more weight on the upper end as time passes. Thus, the model predicts that the use of an unbalanced panel that excludes entry will itself produce a right shift of the distributions since entrants are assumed to enter as small firms. Second, the positive exogenous growth estimate directly predicts a right shift of the Y and W distributions. The third effect comes from type heterogeneity. In steady state, larger firms will over-represent high type firms with high creation rates and small firms will over-represent low type firms with low creation rates. Thus, smaller firms face greater net product destruction than large firms. In the absence of entry, the negative correlation between size and net product destruction rate will in isolation produce a right shift of the Y and W distributions over time. Hence, this effect is also a consequence of the use of an unbalanced panel that excludes entry, but is separate from the first explanation which is not a result of destruction rate heterogeneity.

Value Added per Worker Distribution The distribution of firm labor productivity Y/N is explained primarily through type heterogeneity, the capital share, the structural noise processes, and measurement error. The mean level of value added per worker is closely linked to the estimate of κ. The dispersion in Y/N across firms is explained primarily and in roughly equal parts through type heterogeneity and the positive estimate of $\sigma_{\tilde{\pi}}^2$ - supply side shocks. Measurement error adds to the dispersion measure, but to a smaller extend. Simulation without measurement error ($\sigma_Y^2 = \sigma_W^2 = 0$) yields a reduction in the 1992 Y/N standard deviation measure from 199.19 to 174.62. To an even lesser extend dispersion in Y/N is also affected by the positive estimate of σ_Z^2, that is, demand side shocks because the size of the R&D department is unaffected by particular demand realizations for a firm's products. In the absence of the R&D department, demand side shocks cannot affect labor productivity because an increase in Z realizations will increase value added and manufacturing labor demand by the same fraction. However, since the demand for R&D labor is unaffected by an increase in overall demand, a positive demand shock will result in an increase in the overall labor productivity measure, Y/N. Demand side shocks turn out to be a secondary source of labor productivity dispersion, though. Simulating the model with $\sigma_Z^2 = 0$ yields a reduction in the 1992 Y/N standard deviation measure from 199.19 to 194.03.

The right shift of the value added per worker distribution from 1992 to 1997 is explained as a combination of the exogenous growth estimate and the selection effect in that more productive firms have lower exit hazard rates. However, given the relatively low estimate of overall creative destruction, the primary effect is from the exogenous growth estimate.

Value Added per Worker Persistence and Mean Reversion The persistence in firm labor productivity $cor\left(\frac{Y}{N}, \frac{Y_{+1}}{N_{+1}}\right)$ can be explained directly through σ_π^2, σ_Z^2, the magnitudes of the creation and destruction rates $\gamma(\pi)$ and δ, and measurement error. The estimate of the relatively low level of overall creation and destruction implies that both the supply and the demand shock processes are fairly permanent and they turn out to contribute very little in the explanation of the persistence and mean reversion of value added per worker. Thus it is left to the transitory nature of the measurement error processes to explain the exact persistence and mean reversion of the value added per worker measures. Simulating the model without measurement error ($\sigma_Y^2 = \sigma_W^2 = 0$) results in 1992 persistence and mean reversion moments of $cor\left(\frac{Y}{N}, \frac{Y_{+1}}{N_{+1}}\right) \approx .97$ and $cor\left(\frac{Y}{N}, \Delta\frac{Y}{N}\right) \approx -.015$. So, without the measurement error, the model implies a high level of value added per worker persistence, which is ultimately reduced by the measurement error components. It is important to note that transitory demand shocks have much the same impact as the measurement error components along this dimension. One can speculate that the introduction of an additional demand noise component of a more transitory nature will result in a lower measurement error noise estimate.

Figure 5: Kernel Regression of Firm Growth Rate and Size (1992).



Note: Estimated model drawn in solid line. Data drawn in dashed line. Value added measured in 1 million DKK. Shaded area represents 90% confidence bounds. Value added distribution from data drawn on right axis.

4.5.1 Growth Rate and Size

Beginning with Gibrat (1931), much emphasis has been placed on the relationship between firm growth and firm size. Gibrat's law is interpreted to imply that a firm's growth rate is size independent and a large literature has followed testing the validity of this law. See Sutton (1997) for a survey of the literature. No real consensus seems to exist, but at least on the study of continuing establishments, a number of researchers have found a negative relationship between firm size and growth rate. For a recent example, see Rossi-Hansberg and Wright (2005). One can make the argument that Gibrat's law should not necessarily hold at the establishment level and that one must include firm death in order to correct for survivor bias. Certainly, if the underlying discussion is about issues of decreasing returns to scale in production, it is more likely to be relevant at the establishment level than at the firm level. However, as can be seen from Figure 5, in the current sample of firms where the growth rate – size regression includes firm exits, one still obtains a negative relationship.

Table 6: Firm Size and Growth Moments. Estimate and Counterfactuals

	Data	Point Estimate	$\sigma_Y^2 = 0$ $\sigma_W^2 = 0$	$\sigma_Z^2 = 0$ $\sigma_Y^2 = 0$ $\sigma_W^2 = 0$
$E[\Delta Y/Y]$	-0.029	0.011	-0.001	-0.034
$Std[\Delta Y/Y]$	0.550	0.844	0.083	0.300
$Cor[\Delta Y/Y, Y]$	-0.061	-0.029	-0.022	0.016

At a theoretical level, the model satisfies Gibrat's law; A firm's net innovation rate is size independent. But two opposing effects will impact the unconditional size–growth relationship: First, due to selection, larger firms will tend to over-represent higher creation rate types and in isolation the selection effect will make for a positive relationship between size and and the unconditional firm growth rate. Second, the mean reversion in demand shocks, measurement error, and to a smaller extend in supply shocks introduces an opposite effect: The group of small firms today will tend to over-represent firms with negative demand and measurement error shocks. Chances are that the demand realization of the next innovation will reverse the fortunes of these firms and they will experience relatively large growth rates. On a period-by-period basis, the same is true for the measurement error processes that are assumed to be iid over time. Large firms have many products and experience less overall demand variance. The demand shock and measurement error effects dominate in the estimated model as can be seen in Figure 5.[8] Note that the growth statistics include firm death. If firm deaths are excluded and the statistic is calculated only on survivors, the survival bias will steepen the negative relationship between firm size and firm growth both for the data and for the model since the model reproduces the higher exit hazard rate for small firms that is also found in data.

Thus, in our interpretation the model satisfies Gibrat's law by design, but it nevertheless exhibits a negative relationship between observed firm size and growth rate. As shown in Table 6, the model explains the negative relationship found in data through demand fluctuations and measurement error.[9] Gibrat's law may at one level simply be a statement about the observed

[8]Figure 5 uses value added as the firm size measure. Using labor force size as the size measure instead results in a very similar looking figure and no significant change in the correlation between size and growth.

[9]It is important to note that identification of the demand shock and measurement error processes comes from other aspects of the data as well such as dispersion in the size distribution and a number of the dynamic moments. If the Gibrat related moments are excluded from the estimation, the estimated model still exhibits a negative relationship between observed firm size and growth rate.

relationship between firm size and growth, and its validity is in this sense an issue that can be settled through observations such as the one in Figure 5. However, we have interpreted Gibrat's law to be a statement about a more fundamental proportionality between size and the firm's growth process, specifically innovation. In this case, the structural estimation shows that observation of the relationship between firm growth and firm size is not enough to falsify the statement.

4.5.2 Y/N Growth Decomposition

With the introduction of longitudinal micro-level data sets, a large literature has emerged with the focus on firm level determinants of aggregate productivity growth. See Bartelsman and Doms (2000) for a review of the literature. Given the observation of extensive firm level productivity dispersion, one particular area of interest has been the contribution to aggregate productivity growth from resource reallocation. The discussion has been quantified through decompositions such as (19), where productivity has been defined either as value added per worker or firm TFP. In the estimation in this paper, we have used the value added per worker measure. It should be immediately clear that value added per worker is only loosely related to actual productivity growth in our model, so we should at the outset expect some level of divergence between the reduced form decomposition in (19) and the structural decomposition that we present in the following section.

In the estimation and in the data sample, entry is excluded and the decomposition consequently has no value added per worker growth contribution from entry. The first two columns of Table 7, presents the decomposition results from the data and the simulated steady state that excludes entry. The remaining three columns in the table presents the simulated steady state with entry for the actual point estimate and for the two counterfactuals where measurement error noise and demand shocks have been eliminated.

The steady state with entry simulates not only the dynamic evolution of the sample of incumbents, which is the sample that the estimation is based on, it also simulates the entry process implied by the steady state general equilibrium. The entry process is described in section 3.3. For the estimated model, the size of the potential entrant pool is $4,872m/M = 8,400$. At any point in time, each of these entrants will enter according to entry rate $\gamma_0 = \eta/m = 0.0369$. The entry process is simulated to fit the one year observation frequency in the data. Thus, for each entrant who starts the year in the potential entrant pool, we calculate the transition probability that after 1 year the potential entrant has k products, $\Pr(k_e = k|\pi)$, where the type conditioning refers to the firm type realization at entry. The type realization is obviously unknown to the potential entrant prior to entry, but is subsequently of importance in terms of determining the birth-death process of product lines in the remainder of the year after entry. If $k_e > 0$ the firm is registered as an entrant with k_e products and the subsequent life of the entrant is simulated through the incumbent transition probability described in section 4.3.

The type π conditional potential entrant transition probability, $\Pr(k_e = k|\pi)$, is calculated in a similar fashion to the incumbent transition probability as described in section 4.3. However, in this case, the differential equation system that describes the probability that the potential entrant has product size n at time t, takes the form,

$$\begin{aligned}
\dot{p}_e(t) &= -\gamma_0 p_e(t) \\
\dot{p}_{\pi,1}(t) &= \gamma_0 p_e(t) - (\delta + \gamma(\pi)) p_{\pi,1}(t) \\
\dot{p}_{\pi,n}(t) &= (n-1)\gamma(\pi) p_{\pi,n-1}(t) + (n+1)\delta p_{\pi,n+1}(t) - (\delta + \gamma(\pi)) p_{\pi,n}(t), \ \forall n \geq 2 \\
\dot{p}_{\pi,0}(t) &= \delta p_{\pi,1}(t),
\end{aligned}$$

where the notation follows the notation in section 4.3 with the addition that $p_e(t)$ refers to the probability that the potential entrant is still a potential entrant at time t (and obviously has product size 0). Given the initial condition $p_e(0) = 1$, the potential entrant transition probability is found by solving the above differential equation system for $p_e(1)$ and $p_{\pi,k}(1)$. Thus, the probability that the potential entrant will not have entered after one year is $p_e(1) + p_{\pi,0}(1)$. The latter term reflects the event that a firm enters but exits again before the year's end, in which case the firm is not included in the pool of entrants. It is also seen that the discrete observation frequency implies that entry with more than one product is a positive likelihood event.

The decomposition results on the data suggest a significant contribution to productivity growth from reallocation, roughly 45%, which is a bit higher than results in Foster, Haltiwanger, and Krizan (2001), but still within the general range of their results. Part of this could have been interpreted to be a result of a missing entry component. The model does reasonably well in

Table 7: Y/N Growth Decomposition. Estimate and Counterfactuals

	Data	Point Estimate	Steady State with Entry: Point Estimate	Steady State with Entry: $\sigma_Y^2 = 0$, $\sigma_W^2 = 0$	Steady State with Entry: $\sigma_Z^2 = 0$, $\sigma_Y^2 = 0$, $\sigma_W^2 = 0$
Within	1.015	0.939	1.108	0.796	0.820
Between	0.453	0.350	0.301	0.037	0.053
Cross	-0.551	-0.429	-0.612	-0.065	-0.104
Exit	0.084	0.140	0.160	0.161	0.160
Entry	–	–	0.069	0.072	0.072

capturing the decomposition. The third column introduces the model implied steady state entry to the decomposition and does confirm the idea that the somewhat high reallocation contribution could be a result of missing entry observations.

The fourth column in Table 7 shows the model decomposition results without the measurement error. Both the cross-term and reallocation contribution components drop to close to zero magnitude and measurement error is in this case shown to be a very important issue for the form in (19). Obviously, true productivity growth is unaffected by measurement error. We quantify true productivity growth and a structural decomposition in section 5.

Foster, Haltiwanger, and Krizan (2001) note the potential importance of measurement error and present alternative forms that may be less sensitive to measurement error. But it is doubtful that these alternative measures will be better reflections of productivity growth for the structure in this paper given the loose connection between value added per worker and TFP to the actual productivity contribution of a firm. This issue is related to points raised in Klette and Griliches (1996) where unobserved endogenous pricing at the firm level is discussed. It is an interesting issue whether one can obtain a simple reduced form that can approximate the true decomposition for this paper's model.

In terms of identification, the cross-term component turns out to be of particular importance for the input measurement error parameter. If the model is estimated subject to $\sigma_W^2 = 0$, the remaining model parameters change a little towards a bit more estimated type dispersion, but leaves the estimated cross term component close to zero. Allowing for input measurement error results in the fairly good fit of the cross-term component as shown in Table 7. In isolation, the input

Table 8: Data Moments by Industry

	Manufacturing		Wholesale and retail		Construction	
	1992	1997	1992	1997	1992	1997
Survivors	2,051.000	1,536.000	1,584.000	1,189.000	651.000	480.000
$E[Y]$	30,149.460	35,803.473	22,952.920	28,386.719	15,191.354	16,869.550
$Med[Y]$	15,117.552	18,855.682	12,740.250	15,288.949	8,688.501	10,691.434
$E[W]$	15,047.636	17,318.195	10,696.683	12,712.899	9,973.166	10,594.737
$Med[W]$	8,031.273	9,530.273	6,417.403	7,650.565	5,785.053	6,832.554
$Std[Y]$	56,095.672	69,597.651	33,410.862	41,426.484	31,311.623	22,478.083
$Std[W]$	24,673.900	27,168.284	15,365.073	16,809.785	24,545.298	14,195.942
$E\left[\frac{Y}{N^*}\right]$	379.047	422.471	410.234	466.591	305.075	342.273
$Std\left[\frac{Y}{N^*}\right]$	163.214	226.934	171.716	278.613	133.213	174.052
$Cor\left[\frac{Y}{N^*}, \frac{Y_{+1}}{N^*_{+1}}\right]$	0.650	0.728	0.325	0.674	0.428	0.345
$Cor\left[\frac{Y}{N^*}, \Delta\frac{Y}{N^*}\right]$	-0.024	-0.195	-0.195	-0.259	-0.327	-0.560
$Cor[Y, W]$	0.889	0.855	0.922	0.914	0.967	0.922
$Cor\left[\frac{Y}{N^*}, Y\right]$	0.236	0.200	0.252	0.188	0.131	0.174
$Cor\left[\frac{Y}{N^*}, N^*\right]$	0.011	-0.003	-0.028	-0.039	-0.040	-0.093
$E[\Delta Y/Y]$	-0.035	–	-0.042	–	-0.025	–
$Std[\Delta Y/Y]$	0.474	–	0.425	–	0.448	–
$Cor[\Delta Y/Y, Y]$	-0.073	–	-0.090	–	-0.122	–
Growth decomp.						
– Within	0.863	–	1.176	–	0.986	–
– Between	0.365	–	0.618	–	0.635	–
– Cross	-0.297	–	-0.826	–	-0.870	–
– Exit	0.068	–	0.032	–	0.249	–

measurement error implies some Y/N dispersion and the estimation responds by lowering the type dispersion estimate a little to fit the actual Y/N dispersion. It is interesting that the measurement error estimate is very moderate, and has little effect on the remaining model parameter estimates, but it has a very significant impact on the decomposition results.

4.6 Estimation by Industry

It is of course possible that the correlations and other data moments in Table 2 are a result of firm heterogeneity across industries and does not reflect the true picture within more homogenous subgroups of firms. This turns out not to be the case. Data moments by industry reveal the same qualitative picture as in Table 2 for each industry. Table 8 presents data moments for the 3 largest industries (by firm count). All industries show evidence of significant firm productivity dispersion, a roughly zero correlation between productivity and firm input size and a positive

Table 9: Point Estimate by Industry

	Manu-facturing	Wholesale and retail	Construction
c_0	821.1786	639.0215	93.0344
c_1	4.2496	3.7907	3.2079
κ	0.4515	0.5000	0.3163
Z	19,588.1611	17,962.8955	10,717.1519
δ	0.0687	0.0584	0.0704
σ_z^2	0.8291	0.7748	0.5938
$\sigma_{\tilde{\pi}}^2$	1.6388	0.2173	0.0012
μ_π	-6.1584	-7.3303	-5.2908
σ_π^2	8.1522	13.0962	6.1584
$\widehat{g}$	0.0194	0.0211	0.0131
σ_Y^2	0.0140	0.0151	0.0188
σ_W^2	0.0215	0.0194	0.0301
Inferred Estimates			
η	0.0483	0.0465	0.0556
m	1.8490	2.5405	2.8637
M	0.8139	0.8765	0.8777
L	53.0892	45.3600	36.3381
$\bar{\gamma}$	0.0205	0.0120	0.0149
Entry q-distribution			
10^{th} percentile	1.0000	1.0000	1.0002
Median	1.0016	1.0005	1.0042
90^{th} percentile	1.0500	1.0384	1.0869
Steady state q-distribution			
10^{th} percentile	1.0001	1.0000	1.0002
Median	1.0021	1.0007	1.0050
90^{th} percentile	1.0821	1.0677	1.1212

correlation between productivity and firm output size (roughly 0.2). All industries also display a fair amount of productivity persistence and mean reversion. Finally, both the value added and wage bill distributions are characterized by a strong right shift over time across industries.

The estimates by industry are reported in Table 9. The model estimates by industry are not qualitatively different from the full sample estimate but it is worth noting a consistent drop in the estimated type dispersion in the industry estimates. This is likely a result of effectively allowing for more heterogeneity in other model parameters.

5 Reallocation and Growth

If more profitable firms grow faster, then workers move from less to more profitable surviving firms as well as from exiting to entering firms. This selection effect can be demonstrated by noting that more profitable firms are over represented (relative to their fraction at entry) among those that produce more than one product and that this "selection bias" increases with the number of products produced. Namely, by equation (13), the difference between the relative fraction of a given firm type in the surviving population with k products and relative the fraction in its entry cohort,

$$\frac{M_k(\pi')}{M_k(\pi)} - \frac{\phi(\pi')}{\phi(\pi)} = \frac{\phi(\pi')}{\phi(\pi)}\left[\left(\frac{\gamma(\pi')}{\gamma(\pi)}\right)^{k-1} - 1\right], \tag{34}$$

is positive and increasing in k when $\pi' > \pi$.

From equation (5), the equilibrium rate of growth in consumption is

$$\begin{aligned}
\frac{\dot{C}}{C} &= g = \delta E \ln q \\
&= \delta\left(\int_\pi E\left[\ln q(\pi)\right]\frac{\eta\phi(\pi)d\pi}{\delta} + \int_\pi E\left[\ln q(\pi)\right]\frac{\gamma(\pi)\sum_{k=1}^{\infty} kM_k(\pi)d\pi}{\delta}\right) \\
&= \eta\int_\pi E\left[\ln q(\pi)\right]\phi(\pi)d\pi + \int_\pi \gamma(\pi)E\left[\ln q(\pi)\right]\sum_{k=1}^{\infty} kM_k(\pi)d\pi
\end{aligned}$$

where $q(\pi) = (1-\kappa)/(1-\pi-\kappa)$ is the quality of the products of a type π firm and δ is the aggregate rate of creative destruction as defined in equation (18). The decomposition of the rate of productivity growth,

$$\begin{aligned}
g &= \eta\int_\pi E\left[\ln q(\pi)\right]\phi(\pi)d\pi + \int_\pi \gamma(\pi)E\left[\ln q(\pi)\right]\phi(\pi)d\pi \\
&\quad + \int_\pi \gamma(\pi)E\left[\ln q(\pi)\right]\left[\frac{\eta}{\delta-\gamma(\pi)} - 1\right]\phi(\pi)d\pi
\end{aligned} \tag{35}$$

where $\frac{\eta\phi(\pi)}{\delta-\gamma(\pi)} = \sum_{k=1}^{\infty} kM_k(\pi)$ from equation (13), highlights the role of worker reallocation from exiting to entering firms as well as from less to more productive firms as sources of productivity growth. The first term $\eta\int_\pi E\left[\ln q(\pi)\right]\phi(\pi)d\pi$ is the net effect of entry and exit on productivity growth. The second term $\int_\pi \gamma(\pi)E\left[\ln q(\pi)\right]\phi(\pi)d\pi$ is the average contribution of continuing firms to growth were there no selection. Finally, the last term $\int_\pi \gamma(\pi)E\left[\ln q(\pi)\right]\left[\sum_{k=1}^{\infty} kM_k(\pi) - \phi(\pi)\right]d\pi$ represents the contribution of worker reallocation from firms with products of lesser quality to firms that produce higher quality products.

Table 10: Labor Productivity Growth Rate Estimates.

	Point Estimate	Bootstrap Estimate	Std Deviation
Growth rate g	0.0213	0.0232	0.0034
Decomposition shares:			
– Entry	0.1436	0.1301	0.0138
– Continuing	0.1765	0.1662	0.0250
– Reallocation	0.6800	0.7032	0.1120

Since the total measure of products is unity $\left(\int_\pi \sum_{k=1}^\infty kM_k(\pi)d\pi = 1\right)$ and $\phi(\pi)d\pi$ is the fraction of entrants of type π, $\left(\int_\pi \phi(\pi)d\pi = 1\right)$, it follows that

$$0 = \int_\pi \left[\sum_{k=1}^{\infty} kM_k(\pi) - \phi(\pi)\right] d\pi = \int_\pi \left(\frac{\eta}{\delta - \gamma(\pi)} - 1\right) \phi(\pi)d\pi.$$

Hence, the fact that $\gamma(\pi)$ is strictly increasing in π implies that the contribution to growth of the reallocation of workers among continuing firms, the last term in (35), is positive. Equivalently, it is positive because $\gamma(\pi)E\left[\ln q(\pi)\right]$ is strictly increasing in π and the steady state distribution of types stochastically dominates the distribution of types at entry as a consequence of the firm size selection process.

Given the parameter estimates reported in the previous section, the implied aggregate growth rate and its components are those reported in in table 10. These calculations raise several interesting issues. First, they imply an over all growth rate in productivity somewhat higher than the typical estimate. This fact provides indirect support for arguments that the measurement methodologies currently in use fail to fully separate quality improvements from price increases.[10] In addition, the estimates imply that worker reallocation from both exiting to entering firms and among surviving firms account for 13% and 70% respectively of the aggregate rate of growth. These numbers suggest a very important role to both forms of reallocation.

It is seen that the reduced form growth decomposition in (19) discussed in section 4.5.2 is not a useful reflection of the actual structural decomposition as it has been presented in this section. This is in large part because the empirical measure of labor productivity Y/N is not a direct reflection of the productivity contribution of a given firm in the model. This is partly because the product of the labor engaged in innovation is not measured in Y. Furthermore, while there

[10] In the U.S., this argument is fully articulated in Boskin, Dulberger, Gordon, Griliches, and Jorgenson (1996).

will be a monotonic relationship between value added per manufacturing labor and the quality improvement of the innovation, value added per manufacturing labor will not necessarily correctly reflect the exact labor productivity growth contribution. The problem is unfortunately not solved by looking at TFP rather than Y/N since exactly the same problems apply. Thus, the reduced form decomposition in (19) will not be very informative about sources of aggregate productivity growth for a structure like the one in this paper. It is an interesting question whether a simple reduced form measure on standard observable statistics exists that will provide a good approximation of the growth decomposition for the model in this paper.

6 Concluding Remarks

Large and persistent differences in firm productivity and firm size exist. Evidence suggests that the reallocation of workers across firms and establishments is an important source of aggregate economic growth. In a companion paper, Lentz and Mortensen (2005), we explore a variant of the equilibrium Schumpeterian model of firm size evolution developed by Klette and Kortum (2004) that provided insights into these and other empirical regularities. In our version of the model, firms that can develop products of higher quality grow larger at the expense of less profitable firms though a process of creative destruction. Worker reallocation from less to more profitable firms induced by the process contributes to aggregate productivity growth. Furthermore, the model is consistent with the observation that there is no correlation between employment size and labor productivity and a positive correlation between value added and labor productivity observed in Danish firm data.

In this paper, we take the model to the data. Namely, we fit its structure to Danish firm panel data for the 1992–1997 time period. We find that the parameter estimates are sensible and that the model provides a reasonable fit to many of the moments of the joint distribution of size as measured by value added and employment. The model also explains the evolution of the size distribution of firms in the panel over the observation period.

By design, the growth rate of a firm is size independent, but the model fits the negative unconditional firm size growth relationship in data. The model also captures the reduced form growth decomposition form that is standard to the literature, which suggests a strong contribution to growth from reallocation. But the model explains a large part of the fit with a moderate amount of measurement error and the actual determinants of productivity growth in the model are not reflected well by the reduced form decomposition.

Finally, the quantitative model has interesting aggregate implications for the growth process. First, the implied rate of productivity growth, 2.1% per year, is larger than estimates based on standard accounting methods. Second, reallocation of workers from less to more productive surviving firms is shown to account for more than 2/3 of aggregate productivity growth.

A Appendix

In this section, we present the algorithm used to compute the values of model parameters implied by the estimates and the equilibrium and optimal growth rates, all reported in the text. To do so, one must account for the two parameters not explicitly used in the initial presentation of the model, the average demand per product, Z, which was normalized to unity in the model, and the cost of capital per product line, denoted κZ. Hence, profit per product line can be represented as πZ for a firm of type π where

$$\pi = (1-\kappa)(1-q^{-1}) \tag{36}$$

is now profit express as a fraction of value average sales.

Since the parametric form of the steady state distribution of firms over profit, denoted $p(\pi)$ in the text, is specified in the model estimated, one needs to derive its relationship to the initial density of entering firms over profit, $\phi(\pi)$, by inverting the steady state relationship implied by the model. Specifically,

$$p(\pi) = M(\pi)/M$$

where $M(\pi)$ is the steady state mass of firms of type π and $M = \int_\pi M(\pi)d\pi$ is the total mass of firms. Since

$$M(\pi) = \sum_{k=1}^{\infty} M_k(\pi) = \ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)\frac{\eta\phi(\pi)}{\gamma(\pi)}$$

from equation (13), it follows that

$$\eta\phi(\pi) = \frac{\gamma(\pi)M(\pi)}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)} = \frac{\gamma(\pi)p(\pi)M}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)},$$

At this stage, the aggregate entry rate η and the total mass of firms M have yet to be separately identified. But by $\int_\pi \phi(\pi)d\pi = 1$, it follows that,

$$\eta = \eta\int_\pi \phi(\pi)d\pi = M\int_\pi \frac{\gamma(\pi)p(\pi)}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)}d\pi. \tag{37}$$

Consequently, the profit density at entry is

$$\phi(\pi) = \frac{\frac{\gamma(\pi)p(\pi)}{\ln\left(\frac{\delta}{\delta-\gamma(\pi)}\right)}}{\int_x \frac{\gamma(x)p(x)}{\ln\left(\frac{\delta}{\delta-\gamma(x)}\right)}dx}. \tag{38}$$

Equation (15) and the assumption that the measure of products is unity, the steady state measure of continuing firms in the market solves

$$\begin{aligned} 1 &= \int_{\pi} \sum_{k=1}^{\infty} k M_k(\pi) d\pi = \int_{\pi} M(\pi) \sum_{k=1}^{\infty} \frac{k M_k(\pi)}{M(\pi)} d\pi \\ &= \int_{\pi} \frac{\gamma(\pi) M(\pi)}{(\delta - \gamma(\pi)) \ln\left(\frac{\delta}{\delta - \gamma(\pi)}\right)} d\pi = M \int_{\pi} \frac{\gamma(\pi) p(\pi)}{(\delta - \gamma(\pi)) \ln\left(\frac{\delta}{\delta - \gamma(\pi)}\right)} d\pi. \end{aligned} \tag{39}$$

Hence,

$$\eta = \frac{\int_{\pi} \frac{\gamma(\pi) p(\pi)}{\ln\left(\frac{\delta}{\delta - \gamma(\pi)}\right)} d\pi}{\int_{\pi} \frac{\gamma(\pi) p(\pi)}{(\delta - \gamma(\pi)) \ln\left(\frac{\delta}{\delta - \gamma(\pi)}\right)} d\pi}. \tag{40}$$

from by equations (37) and (39).

To solve the planner's problem, one also needs the size of the aggregate labor force, L, and the measure of potential entrants, m. Because one can show that the limit price charged by the current supplier of each product solves $p(1-\kappa) = wq$ when a capital cost exists, the demand for production workers is $Zx(\pi) = 1/p = Z(1-\kappa)/wq = Z(1-\kappa-\pi)/w$ from (36). Hence, equations (12) and (13) imply

$$L = Z\left[\int_{\pi} \left(\frac{1-\kappa-\pi}{w} + \widehat{c}(\gamma(\pi))\right) \frac{\eta \phi(\pi) d\pi}{\delta - \gamma(\pi)} + m\widehat{c}(\eta/m)\right] \tag{41}$$

where, as specified in the text, $\widehat{c}(x) = c_0 x^{1+c_1}$. Finally, one can obtain the value of m by using the fact that the marginal cost of entry must equal the expected marginal cost of innovation by incumbents. Specifically, equations (11) and (10) imply require that m solves

$$\widehat{c}'\left(\frac{\eta}{m}\right) = \int_{\pi} \widehat{c}'\left(\gamma(\pi)\right) \phi(\pi) d\pi \tag{42}$$

Finally, the parametric specification of heterogeneity in product quality is

$$q(z) = 1 + e^{\mu_{\pi} + \sigma_{\pi} z} \tag{43}$$

where z is the standard normal random variable. Hence, one can use the fact that $f(z)dz = p(\pi(z))d\pi(z)$, where $f(z)$ is the standard normal pdf and $\pi(z) = (1-\kappa)(1-q(z)^{-1})$ by (36), to compute all the necessary integrals in the equations above and those that define the components of the growth rate found in the text.

References

Altonji, J. G. and L. M. Segal (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics 14*: 353–366.

Alvarez, Javier, Martin Browning, and Mette Ejrnæs (2001). Modelling income processes with lots of heterogeneity. *Working Paper*.

Baily, Martin Neil, Eric J. Bartelsman, and John Haltiwanger (1996). Downsizing and productivity growth: Myth or reality? *Small Business Economics* 8, no. 4: 259–78.

Bartelsman, Eric J. and Mark Doms (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature* 38, no. 3: 569–594.

Boskin, Michael J., Eller R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson (1996). *Toward a More Accurate Measure of the Cost of Living*. Final Report to the Senate Finance Comittee from the Advisory Commisson to Study the Consumer Price Index. Washington: Senate Finance Committee.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh (1996). *Job Creation and Destruction*. Cambridge and London: MIT Press.

Fallick, Bruce C. and Charles A. Fleischman (2001). The importance of employer-to-employer flows in the u.s. labor market. *Federal Reserve Board Finance and Economics Discussion Paper 2001-18*.

Foster, Lucia, John Haltiwanger, and C. J. Krizan (2001). Aggregate productivity growth: Lessons from microeconomic evidence. In Charles R. Hulten, Edwin R. Dean, and Michael J. Harper (Eds.), *New Developments in Productivity Analysis*, pp. 303–63. Chicago and London: NBER Studies in Income and Wealth, University of Chicago Press.

Frederiksen, A. and N. Westergaard-Nielsen (2002). Where did they go? *Århus School of Business Working Paper*.

Gibrat, Robert (1931). *Les Inégalités Économiques; Applications: Aux Inégalités Des Richesses, À la Concentration Des Entreprises, Aux Populations Des Villes, Aux Statistiques Des*

Familles, Etc., D'une Loi Nouvelle, la Loi de L'effet Proportionnel. Paris: Librairie du Recueil Sirey.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993). Indirect inference. *Journal of Applied Econometrics* 8, no. 0: S85–S118.

Grossman, Gene M. and Elhanan Helpman (1991). *Innovation and Growth in the Global Economy*. Cambridge, Massachusetts and London, England: MIT Press.

Hall, George and John Rust (2003). Simulated minimum distance estimation of a model of optimal commodity price speculation with endogenously sampled prices. *Working Paper*.

Horowitz, Joel L. (1998). Bootstrap methods for covariance structures. *Journal of Human Resources 33*: 39–61.

Klette, Tor Jakob and Zvi Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics* 11, no. 4: 343–361.

Klette, Tor Jakob and Samuel Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy* 112, no. 5: 986–1018.

Lentz, Rasmus and Dale T. Mortensen (2005). Productivity growth and worker reallocation. *International Economic Review* 46, no. 3: 731–751.

Rossi-Hansberg, Esteban and Mark L.J. Wright (2005). Firm size dynamics in the aggregate economy. *NBER Working Paper Series 11261.*

Stewart, Jay (2002). Recent trends in job stability and job security: Evidence from the march CPS. *U.S. BLS Working Paper 356.*

Sutton, John (1997). Gibrat's legacy. *Journal of Economic Literature* 35, no. 1: 40–59.

Career Dynamics Under Uncertainty: Estimating the Value of Firm Experimentation*

Elena Pastorino
University of Iowa

First Draft: June 2004
This Draft: July 2005

Abstract

This paper develops and structurally estimates a dynamic learning model in which a firm can acquire information about a worker's ability by observing his performance on the job. Ability determines both the profitability of a job and the job-dependent distribution of output signals. As a result, the informativeness of performance about individual productivity varies in the job a worker performs. Because of the trade-off between learning and short-run profit maximization, the firm's optimal information acquisition strategy is the solution to an experimentation problem, a multi-armed Bandit problem with dependent and independent arms. Under the firm's optimal employment policy, a worker is assigned to jobs of decreasing degree of informativeness, as measured by the dispersion in posterior beliefs. The purpose of the analysis is to investigate to what extent uncertainty about ability affects the dynamic pattern of a worker's transition across jobs within a firm, i.e., the timing and job characteristics of a career. To this end, the model is structurally estimated using longitudinal data from a single U.S. firm, on the cohorts of managers entering the firm at the lowest managerial level between 1970 and 1979. Estimation results confirm that a theoretically restricted learning model can succeed in fitting the dynamic profile of the probability of retention and promotion at the major job positions within the firm. The estimated model is then used to compute the firm's value of information and to evaluate the effect on this value, the pattern of job assignments, and of turnover of (i) changes in the discount rate, which reflect changes in market interest rates, and (ii) alternative information structures.

Keywords: Retention, Job Assignment, Learning, Experimentation, Dependent Bandit.

JEL Classification: C73, D21, D83, J41, M12, M51, M54.

*I am truly indebted to Kenneth Burdett, Jan Eeckhout, George Mailath, Steven Matthews, Petra Todd and Kenneth Wolpin for their generous advice and encouragement. I benefited from conversations with Antonio Merlo, Nicola Persico, Jesus Fernandez-Villaverde, Iourii Manovskii and Braz Camargo. I am especially grateful to George Baker for kindly providing the data and to Bengt Holmström for his support in the project. I also wish to thank Hector Chade, Leonardo Felli, Andy McLennan, David Levine, Maurizio Mazzocco, Dale Mortensen, John Rust, Edward Schlee, the participants in the 2004 North American Summer Meeting of the Econometric Society (Brown University, June 2004), the 2004-2005 Annual Meeting of the Society for Economic Dynamics (Florence, July 2004 - Budapest, July 2005), the Fourth Villa Mondragone Workshop in Economic Theory and Econometrics (Rome, July 2004) and numerous seminar participants for their suggestions. Correspondence: Department of Economics, University of Iowa, W332 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242-1994. E-mail: elena-pastorino@uiowa.edu.

1 Introduction

A controversial issue in the theory of the firm is the extent to which promotion and compensation are motivated by a firm's need to provide incentives in the face of moral hazard or to sort employees according to their unobserved abilities. While the existing literature is rich with theoretical contributions, both interpretations have received modest empirical attention (see, for instance, the discussion in Baker and Holmström [1995], Gibbons and Waldman [1999a] and Chiappori [2003]). The understanding of firms' internal organization, and its impact on the allocation of workers to jobs, does nonetheless have important implications for workers' productivity growth with tenure and, therefore, for firms' incentives to employ them.

Intuitively, when a worker's ability is imperfectly observed at the time of hiring, the only way for a firm to assess whether the worker is talented for a job is to employ him and observe his performance over time. However, if the profitability of a job depends on the worker's true skill, then, when deciding whether to employ the worker, or which task to make him perform, the firm has to trade-off the benefit of receiving additional information about his ability against the cost of employing a worker who might be unsuited to the firm's needs. The purpose of this paper is to investigate the role that information acquisition on the part of a firm plays in determining: (i) the ordering of tasks into a hierarchy of job positions, and (ii) the change in a worker's task assignment over his career. Specifically, the focus of the analysis is on quantifying the extent to which uncertainty about ability affects the dynamic pattern of a worker's transition across jobs within a firm, i.e., the timing and job characteristics of a career.

This problem is formalized as a learning game between a firm and a worker. For simplicity, the worker's ability can be one of two levels ('high' or 'low') and it is assumed to be unobserved to both the firm and worker. The firm consists of a finite number of jobs, which differ in their profitability and informational content. In particular, ability is more valuable at jobs which contribute more to the firm's profit. Moreover, since the likelihood of observing any given output realization depends both on the worker's skill and on the job he performs, the revenue realized in a period provides information about the worker's true ability.

The fact that the firm can generate different signals about a worker's productivity by assigning him to different jobs implies that, when allocating a worker to a position, the firm faces the same sequential sampling problem of a decision maker who has to choose one among a given set of alternatives, without knowing the distribution of payoffs associated with each. Because of the implied trade-off between learning and short-run profit maximization, the firm's employment problem can be shown to be strategically equivalent to a particular type of experimentation problem, a Bandit problem with dependent arms (the jobs) and independent arms (the outside option the firm collects if it does not employ the worker). In a way, the main distinction between this framework and learning and matching models à la Jovanovic [1979a, 1979b] can be traced to the generalization of the informational content of jobs. While in those models a worker experiments across different jobs of uncertain but independent worth, so that experience on a job is uninformative about the quality of alternative matches, in the following the perspective of interest is the one of a firm choosing a job for a worker of uncertain talent. By allowing for the existence of components of ability that are correlated across jobs, the quality of a job-worker match is dependent across different jobs for the same worker. Thus, performance at one job can in principle provide valuable information about a worker's productivity at alternative jobs within a firm.

Under suitable restrictions, the solution to this problem is essentially unique and can be completely characterized by a sequence of reservation beliefs. The firm's optimal employment policy prescribes that the worker be assigned to more informative jobs, i.e., those that generate greater dispersion in posterior beliefs, when uncertainty about the worker's human capital is highest, and to more profitable but riskier positions, as the firm learns about the worker's true productivity. In particular, due to the benefit of improved information, it is optimal for the firm to allocate a worker to a job at which he has a strict comparative disadvantage early in his career, when the prior distribution on ability is most diffuse. In this framework, however, learning is typically incomplete, in the sense that the firm always faces (in an ex ante sense) the risk of dismissing a high ability worker, after observing a sequence of low revenue realizations sufficiently long to convince it that the worker's talent is low rather than high.

One purpose of the analysis is to assess the extent to which a learning rationale for job transitions inside a firm can account for the pattern of retention and promotion observed in the data. In order to focus on the interpretation of promotion dynamics as a sorting device, the model intentionally abstracts from issues of incentive provision. As mentioned, promotions could also be rationalized as an incentive mechanism, to induce workers to undertake costly unobserved actions in the interest of the firm. In this case, however, estimation of the effect of screening would require isolating the learning component from the incentive one, given that informational asymmetries arise endogenously in a dynamic moral hazard setup. Therefore, as a first approximation to investigate the empirical relevance of the hypothesis that workers are gradually sorted to higher level jobs, according to their perceived ability, the analysis restricts attention to the problem of information acquisition in a pure learning framework.

To this end the model is structurally estimated, by smooth simulated maximum likelihood, using a unique longitudinal dataset from a single U.S. firm in a service industry between 1969 and 1988. The estimation sample consists of the ten years of observations on job assignments, either Level 1, Level 2 or Level 3, and performance ratings for the cohorts of managers entering the firm at the lowest managerial level, Level 1, between 1970 and 1979, with at least sixteen years of education at entry (i.e., college graduates). The estimation results confirm that the model fits successfully the dynamic profile of the probability of separation from the firm and of retention at Level 1, respectively increasing and decreasing in a worker's tenure. It also captures the qualitative and quantitative features of the pattern of assignment to Level 2, decreasing after the second year since entry at the firm, and to Level 3, at first increasing and then decreasing.

The estimated model can also be used to provide a measure of the firm's value of information and of the inefficiency of job assignment and turnover. Intuitively, since the firm can condition its employment decision in any future period on the performance signal observed in the current period, a natural measure of the gross value of information is the maximal expected extra profit that the firm obtains by observing the worker's output in the current period, due to the improved assessment of his ability. This measure specifically captures the firm's own valuation of the variation in posterior beliefs, i.e., 'new' information, generated by all the possible outcome signals to be realized at each job. In this sense, the firm's demand for information can be uncovered as measured by the firm's willingness to pay to acquire it. Because, as explained, acquiring information about a worker's ability is costly in an opportunity cost sense, the option value of this information can also be quantified and the net value of information estimated. The opportunity cost of information is then measured as the one-period profit loss from choosing the assignment (no employment, Levels 1, 2 or 3) which maximizes dynamic rather than static profit.

Given the estimated values of the parameters of the model, a number of counterfactual exercises are performed. The goal is to investigate the impact on the value of information and, through this, on the probability of retaining a high ability worker (i.e., the extent to which learning takes place through employment), of (*i*) changes in the firm's degree of time impatience, which reflect changes in market interest rates, and (*ii*) alternative informational structures. In particular, increased precision of prior information increases the probability of employment of a high ability worker between 1 and 5.4 percent. Compared to the benchmark case, in which parameters are fixed at their estimated values, when Level 1 becomes perfectly informative, i.e., one period of observation of the worker's output at the level perfectly reveals his ability, the value of information to the firm can increase by more than 100 percent. This in turn causes a reduction in the turnover of high ability workers between 30.4 percent, at low tenures, and 3,791.3 percent, at high tenures. The greatest increase in the probability of retention of high ability workers is nevertheless achieved when Level 2 becomes perfectly informative. In this case, the increase in the firm's value of information can be as large as 480.6 percent, down to a minimum of approximately 1 percent at the highest belief values. The corresponding increase in the probability of employment of a high ability worker is between 50.3 and 7,691.9 percent. These results seem to suggest that improved monitoring of workers' performance would be most effective, in terms of the firm's ability to select talented workers, at next-to-entry jobs rather than at entry level positions, given the substantial fraction of exit observed in the data, and predicted by the model, at the intermediate job, Level 2.

The paper is organized as follows. Section 2 introduces the model, Section 3 describes the data and analyzes relevant descriptive statistics. Section 4 presents the solution and the estimation method, while Section 5 contains the estimation results. Section 6 comments on the results of the counterfactual experiments and Section 7 reviews the relevant related literature. Finally, Section 8 briefly concludes and explores directions of further research.

2 A Learning Model

Consider a market populated by firms and workers. Time is discrete and has an infinite horizon, with dates $t = 1, 2, ...$ Firms and workers are infinitely-lived and risk-neutral and share the common discount factor $\delta \in [0, 1)$. In what follows the focus is on a particular firm and a potential employee, under the assumption that the revenue generated by the worker at that firm is independent of any other workers' output.

The worker's true ability at the firm is unknown to both the firm and worker. Nevertheless, they both know that this ability can be described by the parameter θ, which can take on only one of two values, high, $\overline{\theta}$, or low, $\underline{\theta}$, where $\overline{\theta} > \underline{\theta}$. The firm and worker' prior distribution at the beginning of period 1 over the worker's unobserved ability is $\Pr(\overline{\theta}) = \phi_1$ and $\Pr(\underline{\theta}) = 1 - \phi_1$, with $\phi_1 \in (0, 1)$.

If the firm hires the worker in a period, the worker is assigned to one of three tasks, tasks 1, 2 or 3. Suppose the worker is assigned to task k in period t. Focussing on essentials, we assume the revenue generated can be one of two values, $\overline{y}_k$ or $\underline{y}_k$, where $\overline{y}_k > \underline{y}_k$. When the worker's unknown ability is high, revenue is more likely to be high, i.e., $\Pr(\tilde{y}_{kt} = \overline{y}_k | \theta = \overline{\theta}) = \alpha_k$ and $\Pr(\tilde{y}_{kt} = \overline{y}_k | \theta = \underline{\theta}) = \beta_k$, where $1 > \alpha_k > \beta_k > 0$, $k = 1, 2, 3$. In the following we will refer to the task the worker performs in a period equivalently as the job position to which he or she is assigned. We also assume the expected return to the worker outside the match is independent of any knowledge of the worker's ability. Moreover, the worker's ability at the firm is independent of the worker's ability at any other firm.[1]

At the start of any period, the firm proposes employment to the worker at wage w_t. If the worker is hired that period, the firm pays the worker the period wage w_t and then allocates him or her to one of the tasks. All the worker does is to either accept the offer made by the firm or reject it.[2] If the firm does not hire the worker, it obtains the period profit Π and the worker the period income U. At the end of the period, with probability $\xi_k \in (0, 1)$ the match dissolves for exogenous reasons, potentially dependent on the task the worker performed. This matching friction can be interpreted either as the probability that the job position to which the worker is assigned is closed, due to adverse market conditions, or as the (reduced-form) probability of a preference shock that forces the worker to leave the firm. In the model, and in estimation, it is meant to capture all instances of separation which do not depend on the worker's ability, as revealed by his performance on the job.

Since the revenue distribution at each job is completely characterized by the worker's unobserved ability, the actual income generated implies that both the firm and worker can update their beliefs

[1] Ability, here modelled as firm specific, could also be interpreted as general human capital, as long as the worker's employment history at the firm is unobserved by other firms. For a discussion of the wage dynamics which would emerge in presence of outside labor market competition, see the companion paper (Pastorino [2005]). Estimation results contained in the present draft only relate to promotion dynamics, but wage dynamics, in the presence of general human capital, can be accommodated as well. See the discussion in Section 5.

[2] Equivalently, effort can be thought to be verifiable and provided inelastically by the worker, with disutility cost normalized to zero.

about the worker's true talent. Specifically, given the prior ϕ at the beginning of period t that the worker's ability is $\overline{\theta}$, and the fact that the worker is assigned to task k, the updated posterior, after revenue $\overline{y}_k$ or $\underline{y}_k$ is produced, can be respectively computed as

$$\phi_{kh}(\phi) = \frac{\alpha_k \phi}{\alpha_k \phi + \beta_k (1-\phi)} \quad \text{and} \quad \phi_{kl}(\phi) = \frac{(1-\alpha_k)\phi}{(1-\alpha_k)\phi + (1-\beta_k)(1-\phi)}$$

by Bayes' rule. Observe that ϕ_{kh} increases in ϕ and α_k, but decreases in β_k, while ϕ_{kl} increases in ϕ and β_k, but decreases in α_k. In particular, $\alpha_k > \beta_k$, $k = 1, 2, 3$, implies $\phi_{kh}(\phi) > \phi_{kl}(\phi)$, for $\phi \in (0, 1)$. This implies that observing a high revenue improves on the common assessment of the worker's true ability. The worker's objective is to maximize his or her expected discounted lifetime income, whereas the firm's is to maximize expected discounted profit.

In the following, without loss of generality, we will restrict attention to Markov Perfect equilibria (MPE's) of the complete information game played by the firm and the worker, for which ϕ is the state variable.[3] Actions, histories and strategies can be specified in the usual way. From the assumption that the revenue distribution at each task is Bernoulli, it follows that the updated probability at the beginning of period t that the worker is of high ability, from the sequence of revenue realizations at each task, is a sufficient statistic for the firm's and worker' posterior beliefs. Since all MPE's are essentially time-invariant, the subscript t is omitted and the state will be simply denoted by ϕ.[4]

Note that if the worker rejects the firm's offer in a period, the firm obtains a flow payoff of Π, but it does not receive any additional information about the worker's ability. Therefore, if the belief at date t is such that not employing the worker is optimal for the firm, the same choice must be optimal at $t+1$, given that the belief has not changed. Let then $\overline{\Pi} \equiv \Pi/(1-\delta)$ denote the expected discounted profit to the firm if it does not employ the worker. Suppose the firm hires the worker at wage w_k if it employs him or her at task k, when ϕ denotes the common belief about the worker's ability being high. Let the one period expected revenue at task k be denoted by

$$y_k(\phi) \equiv [\alpha_k \phi + \beta_k (1-\phi)]\overline{y}_k + [(1-\alpha_k)\phi + (1-\beta_k)(1-\phi)]\underline{y}_k.$$

In this case the expected return to the firm, from assigning the worker to task $k = 1, 2, 3$, can be expressed as

$$\widetilde{\Pi}_k(w_k, \phi) = y_k(\phi) - w_k(\phi) + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi}$$

[3] As discussed in the companion paper, the game between the firm and the worker admits a representation as a complete information game, with a unique starting node given by the the firm and worker' prior, and a perfectly observed move by Nature in each period, which determines the known transition on the state variable, ϕ. An alternative representation is an incomplete information game where Nature moves first selecting the type of the worker. For the characterization of the equilibrium outcomes of interest, the two representations are equivalent.

[4] For the characterization of equilibria, see the companion paper. Note that the perfection and invariance requirements reduce the equilibrium set to stationary equilibria which are essentially unique in the outcome of interest, i.e., sample paths along which the worker is continuously employed at the firm. It can also be shown that, being the match game between the firm and a worker a Pareto problem, regardless of the degree of competition in the outside market, promotion and wage dynamics can be solved for separately. The dynamic profile of retention and job assignment is hitherto characterized, while the dynamics of wages is currently in progress.

where the expectation E_k is taken over the future values of the posterior, $\tilde{\phi}$, conditional on its current period value ϕ and the task k the worker performs in the period, and $\Pi(\cdot)$ denotes the firm's maximal value from the problem. Notice that the firm and the worker will meet in the following period with probability $1 - \xi_k$.

The wage paid by the firm is relatively simple to derive. Let $\overline{U} \equiv U/(1-\delta)$ denote the worker's expected discounted lifetime income. The worker will accept employment at the firm in a period if and only if $V_k(\phi) \geq \overline{U}$, where $V_k(\phi)$ denotes the worker's expected discounted lifetime income if assigned to task k, when the firm and the worker's belief is ϕ. It can be shown that $V_k(\cdot)$ is strictly increasing in the wage paid, for any $k = 1, 2, 3$ and ϕ. As a consequence, the firm will maximize its expected return if it hires the worker at wage z such that $V_k(z) = \overline{U}$. In particular, in equilibrium the worker is paid U in any period of employment.[5] Hence, for $k = 1, 2, 3$, $\Pi_k(\phi) = \max_w \widetilde{\Pi}_k(w_k, \phi) = \Pi(U, \phi)$, where

$$\begin{aligned}\Pi_k(\phi) &= y_k(\phi) - U + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} \\ &= p_k(\phi)[\overline{y}_k - U + \delta(1-\xi_k)\Pi(\phi_{kh}(\phi))] \\ &\quad +(1-p_k(\phi))[\underline{y}_k - U + \delta(1-\xi_k)\Pi(\phi_{kl}(\phi))] + \delta\xi_k\overline{\Pi}\end{aligned}$$

and $p_k(\phi) \equiv \alpha_k\phi + \beta_k(1-\phi)$ is the probability that high revenue realizes when the worker performs task k.

If the firm hires the worker in a period, given belief ϕ, it maximizes its expected return by assigning the worker to task j, where $\Pi_j(\phi) \geq \Pi_k(\phi)$, $j, k = 1, 2, 3$. Further, the firm employs the worker if and only if $\Pi_j(\phi) \geq \overline{\Pi}$. In particular, the firm's value function $\Pi(\cdot)$ satisfies the following Bellman equation,

$$\begin{aligned}\Pi(\phi) = \max\{&\overline{\Pi}, y_1(\phi) - U + \delta(1-\xi_1)E_1[\Pi(\tilde{\phi})|\phi] + \delta\xi_1\overline{\Pi}, \\ &y_2(\phi) - U + \delta(1-\xi_2)E_2[\Pi(\tilde{\phi})|\phi] + \delta\xi_2\overline{\Pi}, \\ &y_3(\phi) - U + \delta(1-\xi_3)E_3[\Pi(\tilde{\phi})|\phi] + \delta\xi_3\overline{\Pi}\}.\end{aligned}$$

The difference between the firm's expected discounted profit from assigning the worker to task k and to task k', $k, k' = 1, 2, 3$ and $k' \neq k$, can be expressed as

$$\begin{aligned}\Pi_k(\phi) - \Pi_{k'}(\phi) &= y_k(\phi) - y_{k'}(\phi) \\ &\quad +\delta\{(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] - (1-\xi_{k'})E_{k'}[\Pi(\tilde{\phi})|\phi] + (\xi_k - \xi_{k'})\overline{\Pi}\}.\end{aligned} \tag{1}$$

The sign of the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$ therefore depends on the magnitude of the difference in the one period expected revenue, the first term in (1), and in the expected continuation profit, the second term in (1), between the two tasks k and k'. In fact, at any state the return to the firm from task k can be decomposed in the expected revenue produced by the worker in the period and in the

[5]This is an immediate consequence of the fact that the worker's best response consists in accepting any wage offer at least equal to U and rejecting any other offer.

expected continuation value, which depends on the additional information about the worker's ability conveyed by the revenue realized. Since the firm's value function is convex in the posterior belief, as is proved below, this information is of value as long as there is uncertainty about the worker's true ability, i.e., $E_k[\Pi(\tilde{\phi})|\phi] \geq \Pi(\phi)$.

In general, the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$ depends on the particular configuration of parameter values. For instance, suppose $\alpha_1 = 0.900$, $\alpha_2 = 0.828$, $\alpha_3 = 0.999$, $\beta_1 = 0.069$, $\beta_2 = 0.000$, $\beta_3 = 0.274$, $\overline{y}_1 = 114.982$, $\overline{y}_2 = 1,601.966$, $\overline{y}_3 = 4,453.495$, $\underline{y}_1 = -1,998.849$, $\underline{y}_2 = -5,119.335$, $\underline{y}_3 = -369,085.007$, $\xi_1 = \xi_2 = \xi_3 = 0.000$ and $\delta = 0.95$. At all belief values between $\phi = 0.033$ and $\phi = 0.420$, where task 1 is more profitable than task 2 (i.e., $\Pi_1(\phi) - \Pi_2(\phi) \geq 0$), $y_1(\phi) > y_2(\phi)$ holds true but $E_1[\Pi(\tilde{\phi})|\phi] - E_2[\Pi(\tilde{\phi})|\phi] < 0$. On the other hand, if $\delta = 0.99999$, between $\phi = 0.341$ and $\phi = 0.420$ there exist values of ϕ for which task 2 is more profitable than task 1 (i.e., $\Pi_2(\phi) - \Pi_1(\phi) \geq 0$), so that the positive difference $E_2[\Pi(\tilde{\phi})|\phi] - E_1[\Pi(\tilde{\phi})|\phi]$ offsets the negative difference $y_2(\phi) - y_1(\phi)$. Moreover, when $\alpha_k = \alpha_{k'}$, $\beta_k = \beta_{k'}$ and $\xi_k = \xi_{k'}$, $k, k' = 1, 2, 3$ and $k' \neq k$, it follows that $E_k\Pi(\phi) = E_{k'}\Pi(\phi)$, since the distribution of the updated posterior is the same at tasks k and k'.[6] Hence, to make further progress, additional restrictions have to be imposed.

The main assumption we formulate on the profitability of the three tasks is the following:

$$\begin{aligned} &(\mathrm{A1}): \quad y_3(\overline{\theta}) > y_2(\overline{\theta}) > y_1(\overline{\theta}), \;\; y_1(\underline{\theta}) > y_2(\underline{\theta}) > y_3(\underline{\theta}) \\ &(\mathrm{A2}): \quad y_3(\overline{\theta}) > \Pi + U > y_1(\underline{\theta}) \end{aligned}$$

where $y_k(\theta) \equiv E(y_k|\theta)$ is the one period expected revenue to the firm from assigning the worker to task k in period t, conditional on his or her ability being θ.[7] Assumption (A1) is meant to capture the feature that the impact of ability on expected revenue is greater at potentially more profitable tasks. This restriction also implies that task y_2 entails the risk of greater output destruction than task 1, if the worker assigned to it is not of high ability. Similarly, task 3 is 'riskier' than task 2 in output terms.[8] As for (A2), the assumption $y_3(\overline{\theta}) > \Pi + U$ ensures that employment can be profitable for the firm, while the restriction $\Pi + U > y_1(\underline{\theta})$ implies that the firm might find optimal not to hire the worker than to employ him or her at any task. In particular, the firm would never hire a worker of low ability, if it could perfectly observe θ. The first result can then be proved.

Proposition 1. *The firm's value function* $\Pi(\cdot)$ *is well-defined, continuous and convex. Under (A1) and (A2), it is also increasing.*

Proof: See Appendix A. □

Intuitively, characterizing the firm's optimal retention and task assignment policy requires comparing the maximal expected profit that the firm could obtain from assigning the worker to each of

---

[6]See Subsection 4.1 for a description of the numerical solution method.

[7]These restrictions will not be imposed in the estimation of the model. See the discussion in Section 5.

[8]Observe that $y_k(\overline{\theta}) > y_k(\underline{\theta})$, $k = 1, 2, 3$, is, instead, a consequence of the fact that $\alpha_k > \beta_k$ implies that the revenue distribution at task k, when the worker is of high ability, first-order stochastically dominates the revenue distribution at the same task, when he is of low ability.

the three tasks. As discussed, the sign of the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$, $k, k' = 1, 2, 3$ and $k' \neq k$, depends in turn on the difference in the expected one period revenue and in the expected continuation profit from tasks k and k'. In particular, even if the firm's value function was strictly convex, the difference between the expected discounted profit from tasks k and k' could be non monotonic.[9] By assumptions (A1) and (A2), however, the difference in the one period revenue from any two tasks is strictly monotonic in ϕ. Namely, the difference $y_k(\phi) - y_{k'}(\phi)$, $k > k'$, is strictly increasing. Then, in the static case, the unit interval can be partitioned in regions where task k is unambiguously preferred to task k' or viceversa. This observation suggests that a set of sufficient conditions for a characterization of the firm's employment policy can be identified by guaranteing that a global monotonicity condition holds for the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$.

Under some conditions, it can be shown that the single-crossing property of the static revenues $y_k(\phi)$ and $y_{k'}(\phi)$, $k \neq k'$, implied by (A1) and (A2), translates into an analogous single-crossing property of the dynamic profits $\Pi_k(\phi)$ and $\Pi_{k'}(\phi)$. Specifically, let $\phi_{0,1}$ be the cut-off belief value which makes the firm indifferent between not hiring the worker and employing him at task 1 in the static case, when $\delta = 0$. Similarly, let $\phi_{k,k+1}$, $k = 1, 2$, be the cut-off belief which makes the firm indifferent between tasks k and $k + 1$ when $\delta = 0$. Condition $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$ guarantees that the belief values for which the firm is indifferent, respectively, between not employing the worker and allocating him to task 1 ($\phi_{0,1}$), between tasks 1 and 2 ($\phi_{1,2}$) or between tasks 2 and 3 ($\phi_{2,3}$) can be ordered. Then, the firm's policy in the static problem consists in assigning the worker to task 1 if $\phi \in [\phi_{0,1}, \phi_{1,2})$, to task 2 if $\phi \in [\phi_{1,2}, \phi_{2,3})$, to task 3 if $\phi \in [\phi_{2,3}, 1]$ and not employing him or her altogether otherwise.[10] As for the comparison of the expected continuation values, whenever the distribution of the updated posterior at task k is a mean-preserving spread of the corresponding distribution at task $k + 1$, i.e., task k is more informative about ability than task $k + 1$, it follows $E_k\Pi(\cdot) \geq E_{k+1}\Pi(\cdot)$, with $\Pi(\cdot)$ increasing and convex. This is a consequence of the fact that, being the firm's uncertain about the worker's true worth, it values dispersion is posterior beliefs. Then, more informative tasks, which cause a greater spread in the distribution of the updated posterior, are those which are more profitable when the prior distribution is most diffuse.

The restriction $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$ also implies that there might exist a range of belief values for which the worker is assigned to task 2 in equilibrium, and task 2 is preferred to task 3, even in the dynamic case, if (i) there exists an interval of beliefs for which task 2 is statically more profitable than task 3, past the static cut-off $\phi_{1,2}$, and (ii) task 2 is more informative than task 3. The reason is that, due to the greater informativeness of task 2 as compared to task 3, the threshold belief which makes the firm indifferent between tasks 2 and 3 in the dynamic case, ϕ_2^*, is typically greater than

[9] Strict convexity can be shown to hold if the expected one period revenue at each task is strictly convex in ϕ. One way would be to assume that the firm incurs a one period stochastic cost of supervision, as a fraction of the revenue produced, in monitoring the worker's performance at any task and that this cost depends on the worker's true ability.

[10] Observe that we assumed that, whenever indifferent, the firm assigns the worker to the task at which the impact of ability on expected revenue is highest. No employment is meant to indicate all those instances in which the firm offers a wage strictly smaller than U.

$\phi_{2,3}$. However, $y_3(\overline{\theta}) > \Pi + U$ implies that when ϕ is sufficiently close to 1, task 3 is the dominant choice for the firm. Then, only if ϕ_2^* is smaller than ϕ_3^*, the cut-off belief for which the firm is indifferent between tasks 2 and 3 in the dynamic case, the firm benefits from assigning the worker to task 2, when $\delta > 0$.

Define $\overline{\phi}$ to be the belief which makes the firm indifferent between tasks 1 and 2, whenever task 1 is perfectly informative about ability, while task 2 does not provide any information about the worker's true skill. Given the trade-off between the additional payoff generated at task 2, if the worker's assessed ability is sufficiently high, and the greater informativeness of task 1, $\overline{\phi}$ is indeed an upper bound on the range of beliefs for which the firm might prefer assigning the worker to task 1 rather than to task 2 in the static case. It follows

$$\overline{\phi} \equiv \frac{y_1(\underline{\theta}) - y_2(\underline{\theta}) + \frac{\delta(1-\xi_2)\Pi}{1-\delta(1-\xi_2)}}{\frac{y_2(\overline{\theta})-y_2(\underline{\theta})}{1-\delta(1-\xi_2)} - y_1(\overline{\theta}) + y_1(\underline{\theta}) - \frac{\delta(1-\xi_1)(y_3(\overline{\theta})-U-\Pi)}{1-\delta(1-\xi_3)}}.$$

Note that $\overline{\phi} \in (0,1)$ as long as ξ_k, $k = 1,2,3$, is sufficiently small. Let also $k(\overline{\phi}) \equiv \overline{\phi}/(1-\overline{\phi})$. The formal characterization result of the firm's employment policy is contained in the following Proposition.

Proposition 2. *Let (A1) and (A2) hold. Suppose $y_1(\overline{\theta}) > \Pi$, $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, $\alpha_1 \geq \alpha_2 \geq \alpha_3$ and $\beta_3 \geq \beta_2 \geq \beta_1$. Then, there exists $\{\underline{\xi}_k, \overline{\xi}_k\}_{k=1}^3$, with $0 < \underline{\xi}_k < \overline{\xi}_k < 1$, such that $\xi_k \in (\underline{\xi}_k, \overline{\xi}_k)$, $k = 1,2,3$, $\xi_3 \geq \xi_2 \geq \xi_1$ and $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$. In this case, $0 < \phi_1^* < \phi_2^* < \phi_3^* < 1$ exist such that in any MPE the firm's essentially unique employment policy consists in not employing the worker if $\phi \in [0, \phi_1^*)$, assigning him or her to task 1 if $\phi \in [\phi_1^*, \phi_2^*)$, to task 2 if $\phi \in [\phi_2^*, \phi_3^*)$ and to task 3 if $\phi \in [\phi_3^*, 1]$. Moreover, $\phi_1^* < \phi_{0,1}$, $\phi_2^* > \phi_{1,2}$ and $\phi_3^* > \phi_{2,3}$.*

Proof: See Appendix A. □

The set of conditions listed in the Proposition guarantee that the same qualitative features of the optimal policy in the static case carry over to the dynamic case.[11] In particular, the firm's optimal employment policy is again an interval belief strategy, with increasing cut-offs determined by the points of indifference between the alternative-specific values Π_k's, i.e., the expected discounted profit to the firm from assigning the worker to task k. Notice that, modulo the way indifference is solved, the firm's employment policy is also uniquely determined, given that, from single-crossing, the differences $\Pi_1(\phi) - \Pi_2(\phi)$ and $\Pi_2(\phi) - \Pi_3(\phi)$ are strictly decreasing in ϕ, so that the cut-offs ϕ_k^*, $k = 1,2,3$, are unique. Also, the result that $\phi_1^* < \phi_{0,1}$ and $\phi_2^* > \phi_{1,2}$ implies that the worker is assigned to task 1 for belief values for which, in the static case, respectively, either employment would not be profitable or task 2 would be more profitable than task 1. Similarly, from $\phi_3^* > \phi_{2,3}$ it follows that task 2 is allocated to the worker over a belief range for which, in the static case, the firm would make the worker perform task 3 rather than task 2. This distortion in the dynamic cut-offs, with

[11] Notice that the restrictions on ξ_k, $k = 1,2,3$, would reduce to $\xi_3 \geq \xi_2 \geq \xi_1$ if the firm's outside option, Π, was zero.

respect to the static threshold beliefs, implies that it is optimal for the firm to distort the pattern of static comparative advantage, to generate information about the worker's ability when the worker's true worth is uncertain.

Finally, the conditions under which the characterization result in Proposition 2 holds have also implications for the probability of retention of a high ability worker. As expected, the possibility for the firm to experiment on the worker's ability at tasks which are more informative than task 3 reduces the probability of inefficient turnover of high ability workers. The following result can then be proved.

Proposition 3. *Let (A1) and (A2) hold. Suppose* $y_1(\overline{\theta}) > \Pi$, $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, $\alpha_1 \geq \alpha_2 \geq \alpha_3$ *and* $\beta_3 \geq \beta_2 \geq \beta_1$*. Then, there exists* $\{\underline{\xi}_k, \overline{\xi}_k\}_{k=1}^{3}$*, with* $0 < \underline{\xi}_k < \overline{\xi}_k < 1$*, such that* $\xi_k \in (\underline{\xi}_k, \overline{\xi}_k)$, $k = 1, 2, 3$, $\xi_3 \geq \xi_2 \geq \xi_1$ *and* $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$*. In equilibrium in the long run only a high ability worker is retained by the firm and employed at task 3. Moreover, the probability of permanent retention of a high ability worker, strictly smaller than one, is higher when at least tasks* 1 *or* 2 *and task* 3 *are assigned in equilibrium than when only task* 3 *is.*

Proof: See Appendix A. □

The proposition implies that the assignment of tasks 1 and 2 in equilibrium has merely a screening purpose, and it is optimal as long as there is uncertainty about the worker's ability. As characteristic of experimentation problems, also, limiting learning is incomplete. The firm, in an ex ante sense, always faces the risk of observing a sequence of low output realizations sufficiently long to convince it that the worker is actually of low ability, even if his or her true ability is high. One of the purposes of the empirical analysis is indeed to assess the extent to which changes in the informational structure can improve on the firm's capacity to identify high ability workers, by observing their performance at different jobs. This in turn requires investigating the effect on the profitability of employment of changes in the firm's valuation of information on ability. Measurement and estimation of the value of experimentation are discussed in more detail in Section 6.

3 Data

3.1 Sample and Variable Definitions

The data consist of personnel records for all management employees of a medium-sized U.S. firm in a service industry between 1969 and 1988. As described in Baker, Gibbs and Holmström [1994a] (BGH), these records include information on every managerial employee in the firm as of December 31 of each year. Each record consists of an employee ID number, the employee's year of entry, age, education, job title and level, cost center code (i.e., the six-digit code of the organizational unit defined for measuring costs, revenues or profits), salary, salary grade (available from 1979 to 1988), bonus and a job performance rating (from 1, lowest, to 5, highest). In total the data contain 74,071 observations on managerial employees at the firm over the sample years. Salary, title and performance rating are year-end values. It is unclear though when, during the year, pay or title changes occurred or performance ratings were attributed, so these variables may not be exactly concurrent. In the empirical analysis we assume, consistently with the model, that title changes occurred after performance ratings were recorded. However, titles were not coded for some new hires in the last years. Specifically, missing data are significant in 1987 and 1988, in which approximately 10 percent of employees and half of new hires do not have title data.

The size of entry cohorts into managerial positions at the firm grew significantly during the sample period. The entry cohort in 1970 was 230 individuals, while by 1988 it was 1175. BGH report that management constituted about 20 percent of total employment each year. The average age of employees entering managerial positions was 33 with a standard deviation of 8 years; the range was from a minimum of 20 to a maximum of 71 years. The average number of years of education was 15.6 with a standard deviation of 2.4 years; the range was from a minimum of 12 to a maximum of 23 years. Both age and education show little variation across cohorts.[12] As for exit, for the sample of entrants at the firm between 1970 and 1979, 10.9 percent left the firm after one year, while 20.4 percent left after two years and 57.7 percent after nine years.[13]

BGH aggregated job titles into levels according to the timing and frequency of transitions of employees across titles. Specifically, as explained in BGH, in the original data there were 276 different titles, but 14 titles, each representing at least 0.5 percent of employee-years, comprised about 90 percent of the observations and 93 percent of those in which the title was coded. In order to fill the job ladder to the top of the firm's hierarchy, BGH added the top title of Chairman-CEO, together with the only two titles observed in moves from the fourteen major titles to the position of Chairman. Transition matrices were then constructed to analyze movements of employees between these seventeen titles, both for individual years and over the whole sample.

Eight job levels were constructed. Level 1 consists of the three titles which employed almost only new hires. Most moves from Level 1 within the firm were to six other titles, identified as Level 2. Moves from Level 2 were almost exclusively to three other job titles, categorized as Level 3. This process was continued until the original seventeen titles were assigned to 8 job levels, with Chairman-CEO at Level 8. After major titles were assigned, less common titles were assigned to levels based on moves between them and titles already assigned.

The hierarchy which emerges from this level structure consists of two parts, Levels 1-4 and Levels 5-8, with Levels 1-4 containing 97.6 percent of employees, each of approximately the same size. Specifically, over the sample period 16,981 employees are at Level 1, 17,725 at Level 2, 17,253 at Level 3, 13,892 at Level 4. The corresponding figures at Levels 5-8 are 1,194, 373, 56 and 20. It is commonly interpreted that upper level jobs correspond more to general management,

[12]The composition of entrants across job titles did not change markedly, though there was a relative increase in lower level entry during the years 1976-1985. BGH report that the proportion of minorities and women increased steadily. Our data, though, do not include information on sex or race.

[13]As noted by BGH, patterns are similar for later entrants, even if the average career length becomes shorter over time.

while lower level jobs depend more on specialized functional knowledge and require performing less complex tasks. For instance, as described by BGH, at Levels 1-4 about 60 percent of the jobs correspond to specific 'line' (revenue-generating) business units, positions with direct contact with customers or creating and selling products, while approximately 35 percent are 'staff' or 'overhead' positions, in areas such as Accounting, Finance or Human Resources. At Levels 5-6, these two percentages decrease, respectively, to 45 and 25 percent, while general management descriptions such as 'General Administration' or 'Planning' increase to about 30 percent. At Levels 7-8 all jobs are of this form and they entail managing large groups, coordinating across business units and strategic planning, responsibilities which possibly rely more on firm-specific rather than general skills. These observations suggest that the task content of higher level jobs is consistent with our assumption that human capital is most valuable at those jobs.

Over the twenty year sample, the firm has been remarkably stable in the composition of titles and levels. Even as firm size has tripled, the fraction of people at each level has changed very little. After 1984, some new titles were created, but only two are of significant size, representing respectively only 0.6 percent and 0.9 percent of employees (see Table 1 in BGH).

In the data, it is not possible to distinguish whether new entrants into managerial positions in any given year are also new hires at the firm. For instance, a worker could have been promoted from a clerical to a management position. Because a promotion in this case entails a major shift in job tasks, as argued in BGH, and a change from hourly to salaried employment, new promotees into managerial positions are likely to be treated similarly to outside hires. In estimation we focus on the individuals who entered managerial positions between 1970 and 1979 at Level 1. Each entrant cohort is followed for 10 years. This restriction reduces the original sample of 16,133 individuals to 2,714 individuals. The estimation sample is further restricted to the 1,552 individuals with 16 or more years of education at entry.

Performance ratings were coded in the data from 1 (best) to 5 (worst). Ratings of 3, 4 and 5 comprise only a small fraction of all ratings. Ratings 2 to 5 where therefore combined into a single rating, leading to a binary classification of 1 (high rating) and 0 (low rating) as in the model.[14]

3.2 Descriptive Statistics

The model has implications for the ex ante probability that in each period a worker assigned at entry to Level 1 will remain at Level 1 or will be assigned to Level 2 or 3, or will leave the firm. Table 1 shows the proportion of employees at each level, as well as the proportion who separated for each

[14]Ratings of 1 and 2 represent 80.5 percent of all the ratings observed in the original sample (28,398 employee-years have missing rating information, where only 4,703 individuals having no missing rating information in any period) and 89.5 percent of the ratings in the sample used in estimation, in which, by construction, no rating information is missing (see Table B1 in Appendix B). To preserve the informativeness of observed performance about employees' productivity in a year, a rating of 1 has been treated a success, while a rating of $2, 3, 4$ and 5 as a failure. See Appendix B for a comparison of the fraction of ratings 1 through 5 is the original sample and in the estimation sample.

year since entry over a ten year period, for the sample of employees entering the firm between 1970 and 1979 at Level 1 with at least 16 years of education and no level information missing.[15]

As noted, at entry all employees are at Level 1. In the second period, 39.6 percent of employees who entered the firm are assigned to Level 2 and 13.3 percent leave. In the third period, only 17.6 percent of the individuals assigned to Level 1 in the first period remain at Level 1, while 47.6 percent are at Level 2 and 9.3 percent at Level 3. The fraction of employees in Level 1 and 2 jobs rapidly decreases with tenure at the firm, while the proportion of employees assigned to Level 3 increases until the fifth year after entry and then decreases. The proportion of workers who have left the firm is substantial in each year. By the last period of observation, 66.8 percent of the individuals hired at Level 1 have left the firm.

Table 1. Distribution of Employees Across Levels (16 or More Years of Education at Entry, Missing Ratings - 1,552 Employees)

Years Since Entry	**Level 1**	**Level 2**	**Level 3**	**Exit**	**Total**
0	1.000	0.000	0.000	0.000	1.000
1	0.471	0.396	0.000	0.133	1.000
2	0.176	0.476	0.093	0.256	1.000
3	0.079	0.309	0.261	0.352	1.000
4	0.047	0.193	0.327	0.434	1.000
5	0.028	0.128	0.352	0.492	1.000
6	0.020	0.085	0.351	0.544	1.000
7	0.016	0.066	0.329	0.588	1.000
8	0.011	0.050	0.311	0.628	1.000
9	0.008	0.037	0.286	0.668	1.000

The hazard rates of employment termination and promotion are displayed in Table 2, stratified by tenure at each level. At Level 1 the separation hazard is approximately constant over time at about 0.1. The hazard rate for promotions to Level 2 increases in the second year of tenure in Level 1 (from 0.396 to 0.486), it follows slightly in year 3 to 0.436, it decreases to about 0.3 in years 4 and 5 and then decreases to about 0.15 in years 6-8. At Level 2, similarly, the hazard rate of separation shows little variation over the sample periods compared to the hazard rate of promotion to Level 3, which

[15]The corresponding statistics for the current estimation sample of 502 individuals, with at least 16 years of education at entry and no level or performance rating missing, are reported in Appendix B. However, for this sub-sample only 22 employees are observed at Level 4 (24 in the sample of 698 individuals which include all education groups) and none at Levels 5 through 8. Observations at Level 4 were therefore added to Level 3. For the total of 1,552 managerial employees with at least 16 years of education at entry, over the first ten years there are only 1,359 observations on employees at Level 4 (11.4 percent of all observations), 15 on employees at Level 5 and 8 on employees at Level 6. Observations on Level 4 to 6 have similarly been added to observations on Level 3.

at first increases, between the first and the second year of tenure, and then decreases, between the second and the sixth year of tenure. At Level 3 the separation hazard is roughly constant at about 0.1, but the significance of this pattern is limited by the small number of observations available.[16]

Table 2. Hazard Rates of Exit and Promotion by Level (16 or More Years of Education at Entry, Missing Ratings - 1,552 Employees)

Years at Level	Level 1 to Exit	Level 1 to Level 2	Level 2 to Exit	Level 2 to Level 3	Level 3 to Exit
1	0.133	0.396	0.155	0.221	0.097
2	0.130	0.486	0.172	0.556	0.140
3	0.106	0.436	0.202	0.471	0.099
4	0.107	0.298	0.088	0.294	0.110
5	0.083	0.306	0.095	0.286	0.101
6	0.114	0.182	0.077	0.077	0.075
7	0.065	0.129	0.000	0.273	0.108
8	0.160	0.160	0.125	0.000	-
9	0.000	0.235	-	-	-

Table 3. Fraction of High Ratings Among Employees at Level and Promoted Into a Level (16 or More Years of Education at Entry, No Rating Missing - 502 Employees)

Years Since Entry	Level 1	Promoted to Level 2	Level 2	Promoted to Level 3	Level 3
0	0.510	0.518	-	-	-
1	0.362	0.433	0.567	0.810	-
2	0.200	0.211	0.388	0.516	0.810
3	0.118	0.250	0.155	0.273	0.500
4	0.143	0.000	0.243	0.333	0.400
5	0.000	0.000	0.154	0.000	0.208
6	0.000	-	0.167	-	0.118
7	-	-	0.000	-	0.111
8	-	-	0.000	-	0.250
9	-	-	-	-	1.000

Table 3 displays, for each year since entry, the proportion of employees at Levels 1, 2 and 3 who receive a rating of 1 (high), as well as the proportion of employees at each level who receive

[16] The number of individuals employed at Level 3 in period 3 is 144, in period 3 and 4 is 129, in periods 3 to 5 is 111, in periods 3 to 6 is 100, in periods 3 to 7 is 89 and in periods 3 to 8 is 80. At high tenures, the number of retained managers at Level 3 reduces to 74, in periods 3 to 9, and to 66, in periods 3 to 10.

a high rating and are assigned to the next level in the following period, i.e., the fraction of high rating among promoted employees.[17] The empty entries in the first row are due to the fact that all employees are assigned Level 1 when hired. The empty entries for workers at Level 1 promoted to Level 2 are a consequence of the fact that no employee was promoted to Level 2 after the fifth year since entry. Analogously, the empty entries for promoted workers from Level 2 to Level 3 are due to the fact that no employee was promoted to Level 3 after the fifth year since entry.

In each period the proportion of employees at Levels 1 or 2 receiving a high rating decreases over time and is significantly smaller than the fraction of promoted workers with a high rating. Moreover, the frequency of high ratings is larger among employees promoted earlier from either Level 1 to Level 2 or from Level 2 to Level 3 than among employees promoted after longer tenures. At Level 3 as well, a part from periods 8, 9 and 10, the proportion of employees receiving a high rating decreases over time.

3.3 Evidence from the Data and Predictions of the Model

As implied by the model, the probability of employment at the firm at any level is decreasing over time, because an increasing number of high performance realizations must occur for the firm to be willing to retain a worker. This is reflected in the data by the fact that the fraction of individuals employed at Levels 1, 2 or 3 eventually decreases over time. Moreover, the probability of being assigned to Levels 2 and 3 increases only at low levels of tenure, suggesting that employees are sorted at Level 1 according to their perceived ability, before being allocated to higher levels. The intuition from the model behind these patterns is that, at Level 2, the decrease in the probability of employment is due to the fact that good performing employees are eventually promoted to Level 3, if retained. At Level 3 it is the combined result of the selectivity of the firm's retention criterion and of the existence of an exogenous separation shock. The result that the posterior belief must be sufficiently high for a worker to be employed at Level 3, and the fact that firing a low performing worker at Level 3 can be more profitable than demoting him to Level 2 (if the change in the posterior belief after a bad performance realization is sufficiently large), together imply that workers assigned to Levels 2 and 3 might be fired and not demoted.

By comparing, from Table 3, the fraction of workers receiving in each year a high rating with the fraction of workers employed at each level, from Table 1, it follows that employees who are retained at the firm at any level, but not promoted, are those whose performance ratings is on average lower, i.e., promoted workers have highest assessed ability. This evidence is consistent with the equilibrium result that employees are progressively assigned to Levels 2 and 3, as the assessment of their talent, as revealed by their performance on the job, improves. The fact that the probability of being assigned to any level eventually decreases over time, as well as the fraction of employees receiving a high rating at each level, is also consistent with the prediction that workers whose assessed ability decreases are those more likely to leave the firm.

[17] Attention has been restricted to the sub-sample of employees with no rating information missing.

4 Empirical Analysis

4.1 Solution Method

Although the model does not admit a closed-form solution, it can be solved numerically for the firm's unknown value function $\Pi(\cdot)$ and the job-specific values $\Pi_k(\cdot)$, $k = 1, 2, 3$. As argued in Section 2, the value function $\Pi(\cdot)$ is a fixed point of a contraction mapping. Since the belief ϕ about the worker's ability being high is the only state variable in the firm's dynamic programming problem, the state space reduces to the unit interval. Therefore, $\Pi_k(\cdot)$ can be computed recursively by value function iteration. For computational reasons, the state space has been discretized in a uniform grid of 600 equidistant points on the interval [0,1].

When the distribution of revenue realizations is not symmetric across types, i.e., the probability of a high rating for a high ability worker is different from the probability of a low rating for a low ability worker (i.e., $\alpha_k \neq 1 - \beta_k$ for some $k = 1, 2, 3$, where α_k is the probability of a high rating at job k for a worker of high ability and β_k for a worker of low ability), the process of posterior beliefs visits different states along each equilibrium sample path, for given prior belief ϕ_1. This implies that, for every belief value on the grid, the updated posterior computed by Bayes' rule can be a point outside the grid. Note that this problem would also arise in the symmetric case, as long as the probability of a high rating was different across tasks.[18] To ensure that the updated posterior from each possible belief value on the grid is itself a grid point, a nearest neighborhood procedure has been adopted, to select the value on the grid closest to the exact Bayes' update.

The firm's optimal employment (i.e., retention and task allocation) policy is then computed by determining, for each belief value on the grid, the task which generates the highest expected discounted profit, by direct comparison of the alternative-specific values, as computed from $\Pi(\cdot)$.

4.2 Estimation Method

Given that in the model the firm and the worker are assumed to be endowed with the common prior $(1 - \phi_1, \phi_1)$ over the ability space $\{\underline{\theta}, \overline{\theta}\}$ at the beginning of period 1, the distribution of prior beliefs is not determined by the model. In estimation we assume that the probability ϕ_1 that the worker is of high ability is drawn from a beta distribution over the set of belief values for which the assignment of Level 1 is profitable for the firm.[19] Denote the vector of structural parameters to be estimated by

$$\psi = (a_\beta, b_\beta, \delta, (\alpha_k, \beta_k, \overline{y}_k, \underline{y}_k, \xi_k, E_k(\overline{\theta}), E_k(\underline{\theta}))_{k=1}^3),$$

[18] Only if $\alpha_k = 1 - \beta_k$ and $\alpha_k = \alpha_{k'}$, for $k, k' = 1, 2, 3$, the non-linearity of the Bayes map could be accommodated by selecting a different belief grid for each ϕ_1.

[19] The choice of the beta specification is motivated by its flexibility and the fact that it has a compact support, so that, in particular, ϕ_1 can be restricted to belong to the interval of belief values $[\phi_1^*, \phi_2^*)$. In fact, for the relevant set of parameter values selected during estimation, the resulting firm's optimal employment policy is the one predicted by the model, i.e., the interval belief strategy prescribing that the worker be assigned to job 1 as long as ϕ lies in $[\phi_1^*, \phi_2^*)$.

where a_β and b_β are the parameters of the beta distribution from which the initial prior ϕ_1 is drawn, δ indicates the firm and the worker's discount factor, α_k (for a high ability worker) and β_k (for a low ability worker) are, for each Level $k = 1, 2, 3$, the locational parameters of the Bernoulli distribution governing output realizations, which can be high, $\overline{y}_k$, or low $\underline{y}_k$, and ξ_k is the exogenous probability that the worker leaves the firm at the end of a period when assigned to Level k. Performance outcomes are assumed to be measured with error. The classification error rate, E_k, depends on the job level the worker is assigned to in a period and on the worker's true ability.[20]

Observe that the firm's reservation profit, Π, and each worker's reservation utility, U, act in the model as scale parameters of the expected one period return at each level, $y_k(\phi)$. As such, they cannot be separately identified from $\overline{y}_k$ and $\underline{y}_k$. For given α_k and β_k, in fact, a proportional change in $\overline{\Pi}$ and in $\overline{y}_k$ and $\underline{y}_k$ leaves the relative worth of the jobs in static terms unchanged. In particular, even if the one-period revenue Π from terminating the worker and the expected one period revenue $y_k(\phi)$ from Level 1, 2 or 3 increase, the firm is indifferent between any two of the employment alternatives for the same belief values. Similarly, for given ϕ, the same proportional increase in U and decrease in $\overline{y}_k$ and $\underline{y}_k$, for all k, leaves $y_k(\phi)$ unchanged. Therefore, Π and U are normalized to zero.[21]

The model is estimated by smooth simulated maximum likelihood. At any time t denote the vector of observed outcomes for individual i by $O_{it} = (L^o_{it}, R^o_{it})$, the job level the individual is assigned to in period t, L^o_{it}, and the performance realization recorded for the period, R^o_{it}. Let $\theta_1 \equiv \underline{\theta}$ denote the low level of ability and $\theta_2 \equiv \overline{\theta}$ the high level. Let $s_1 \equiv e_1 \geq 16$ indicate the number of years of education of an employee at entry. The likelihood function for a sample of N individuals, observed from period $t = 1$ to period $t = 10$, is given by the product over all individuals of the 10 period outcome histories of observed levels and performance ratings, conditional on their education at entry,

$$\mathcal{L}(\psi \mid s_1) = \prod_{i=1}^{N} \int_{\phi_1} \sum_{k=1}^{2} \Pr(\theta_k \mid \phi_1, s_1) \Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi_1, s_1) dF(\phi_1 \mid s_1).$$

[20] The interpretation of the classification error as type dependent follows the modelling hypothesis that, on average, high ability employees generate more high ratings than low ability employees. A flexible error structure allows therefore the model to fully capture differences in the probability of success across types as can be estimated from the histories of observed ratings.

[21] The estimation of a version of the model which encompasses both promotion and wage (base salary) dynamics is currently being implemented. In this formulation the worker's human capital is assumed to be perfectly transferable across identical firms and employment outcomes to be symmetrically observable to all market participants. Theoretical results for the general case, in which ability can be general or firm specific, are derived in the companion paper. Following the characterization of the equilibria of interest, the specification of the observed wage, when the belief about individual i's ability being high is ϕ_{it}, is $\ln w^o_{ikt} = \ln w_k(\phi_{it}) + \varepsilon_{ikt}$, if individual i is assigned to Level $k = 1, 2, 3$ in period t. In this expression $w_k(\phi_{it}) \equiv a + by_k(\phi_{it})$, where a is a scale correction factor, since the firm's outside option has been normalized to zero, and b reflects the relative price of the good produced by the firm in terms of money (constant 1988 US dollars). Finally, ε_{ikt} is the draw of the measurement error on wages, assumed to be normally distributed with mean zero and variance σ^2_k at Level k. This formulation then allows for an indirect test of the hypothesis that wages at the managerial level are set competitively. Details can be provided upon request.

Since the firm and the worker' initial prior distribution over the worker's ability is not observed, the probability of each individual history has to be integrated over all possible priors. In estimation the beta distribution, which parameterizes the set of potential prior distributions, is discretized in J points over the interval $[\phi_1^*, \phi_2^*)$, so that the likelihood function is approximated as

$$\mathcal{L}(\psi \mid s_1) \simeq \prod_{i=1}^{N} \sum_{j=1}^{J} \Pr(\phi_1 = \phi_1^j \mid s_1) \sum_{k=1}^{2} \Pr(\theta_k \mid \phi_1^j, s_1) \Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi_1^j, s_1) \tag{2}$$

where $\Pr(\theta_1|\phi_1^j, s_1) \equiv 1 - \phi_1^j$ and $\Pr(\theta_2|\phi_1^j, s_1) \equiv \phi_1^j$, $j = 1, ..., J$. In expression (2), $\Pr(\phi_1 = \phi_1^j|s_1)$ represents the probability that the firm and worker i's prior belief about the worker's ability being high is ϕ_1^j at the beginning of period 1. Given that an individual can be either of high or of low ability with probability $\Pr(\theta_k|\phi_1^j, s_1)$ at entry, the likelihood function is obtained as the product over all individuals of the probabilities of the type-dependent outcome histories $\Pr(O_{i1}, ..., O_{i10}|\theta_k, \phi_1^j, s_1)$. The mixture over types is obtained by integrating over the prior distribution $(\phi_1^j, 1 - \phi_1^j)$. For each individual, the probability of his observed employment history at the firm, conditional on his education, is finally computed by weighting the prior-dependent history with the probability of a particular prior being the initial belief the firm and the individual are endowed with.

For any individual the probability of a period-t outcome pair can be factored in the product of the probability of the assigned level and of the performance signal observed in the period, conditional on this level. The conditional probability of an individual i's outcome history can therefore be expressed as

$$\begin{aligned}\Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi_1^j, s_1) \;=\; & \Pr(L_{i1}^o \mid \theta_k, \phi_1^j, s_1) \Pr(R_{i1}^o \mid \theta_k, L_{i1}^o) \cdots \\ & \cdot \Pr(L_{i10}^o \mid \theta_k, \phi_1^j, R_{i1}, ..., R_{i9}, s_1) \cdot \Pr(R_{i10}^o \mid \theta_k, L_{i10}^o)\end{aligned} \tag{3}$$

where $L_{it}^o \in \{L_0, L_1, L_2, L_3\}$ indicates the level assignment, with L_0 representing no employment, and $R_{ij} \in \{\emptyset, 0, 1\}$, $j = 1, ..., 10$ the actual performance outcome realized in period t (note that the rating of a worker who has left the firm is missing by construction). The probability of each observed level is computed conditional on the worker's unobserved ability (which determines the probability distribution of the true performance signal at each task), the initial prior ϕ_1 (which determines the probability of the worker's initial job assignment at the firm) and the sequence of past realized ratings (which, together with the initial prior, determine the value of the updated posterior). The probability of the observed rating, instead, only depends upon the worker's actual ability and the level assigned, from our assumption that the distribution of revenue realizations at Level $k = 1, 2, 3$ is bernoulli with parameter α_k, if the worker is of high ability, and β_k, if the worker is of low ability.

In expression (3) it is implicit that, given the bernoulli process governing output realizations at each level, at any time t the initial prior and the sequence of *true* performance realizations are sufficient statistics for the updated posterior. Specifically, $\Pr(L_{it}^o|\theta_k, \phi_1^j, R_{i1}, ..., R_{it-1}, s_1) = \Pr(L_{it}^o|\theta_k, \phi_t^j, s_1)$, where ϕ_t^j represents the updated posterior at the beginning of period t from the prior ϕ_1^j and the sequence of actual performance outcomes from period 1 through $t-1$, $(R_{i1}, ..., R_{it-1})$.

For each individual, the probability of the per-period level $L^o_{it} = L_r$ is calculated as $\Pr(\Pi_r(\phi) = \max\{\Pi_0(\phi), \Pi_1(\phi), \Pi_2(\phi), \Pi_3(\phi)\})$, viewed, for the purpose of estimation, as a function of the parameters of the model conditional on the data, and it is computed by a kernel smoothed frequency simulator. Specifically, the probability of the observed level in each period, for given initial prior ϕ^j_1, is simulated over S possible realizations of the performance rating in the period and smoothed through a logistic kernel with bandwidth parameter τ.[22] The corresponding kernel is computed as

$$\Pr(L^o_{it} = L_r \mid \theta_k, \phi^{js}_t, s_1) = \exp\left[\frac{\Pi_r(\phi^{js}_t(\theta_k)) - \max_l\{\Pi_l(\phi^{js}_t(\theta_k))\}}{\tau}\right] \cdot \left\{\sum_{m=0}^{3} \exp\left[\frac{\Pi_m(\phi^{js}_t(\theta_k)) - \max_l\{\Pi_l(\phi^{js}_t(\theta_k))\}}{\tau}\right]\right\}^{-1}$$

at the j-th draw of the initial prior and the s-th simulation draw of the performance realization, with $L_r \in \{L_0, L_1, L_2, L_3\}$. In the expression $\phi^{js}_t(\theta_k)$ denotes the updated posterior from the prior ϕ^j_1 and the sequence of performance ratings $(R_{i1s}, ..., R_{it-1s})$ from period 1 to period $t-1$, simulated conditional on the worker's true ability, i.e., $\phi^{js}_t(\theta_k) = \varphi(\phi^j_1 \mid R_{i1s}(\theta_k), ..., R_{it-1s}(\theta_k))$.[23] The probability of the observed level is then computed as the average of the above kernel over the S simulations of the performance rating,

$$\Pr(L^o_{it} = L_r \mid \theta_k, \phi^j_t, s_1) \simeq \sum_{s=1}^{S} \frac{\Pr(L^o_{it} = L_r \mid \theta_k, \phi^{js}_t, s_1)}{S}.$$

As mentioned, to avoid zero-probability events contributing to the likelihood function, and given the inherent noisiness of the performance appraisal process, it is assumed that performance ratings are measured with error. Formally, the conditional probability of observing a rating $R^o_{it} \in \{0, 1\}$ in period t at level $L^o_{it} = L_k \in \{L_1, L_2, L_3\}$, if the true performance is R_{it} and the worker's ability $\theta_k \in \{\underline{\theta}, \overline{\theta}\}$, is given by

$$\Pr(R^o_{it} = 1 \mid R_{it} = 1, \theta_k, L^o_{it}) = E_k(\theta_k) + (1 - E_k(\theta_k)) \Pr(R_{it} = 1 \mid \theta_k, L^o_{it})$$
$$\Pr(R^o_{it} = 1 \mid R_{it} = 0, \theta_k, L^o_{it}) = (1 - E_k(\theta_k)) \Pr(R_{it} = 1 \mid \theta_k, L^o_{it})$$

where $\Pr(R_{it} = 1|\overline{\theta}, L^o_{it} = L_k) = \alpha_k$ and $\Pr(R_{it} = 1|\underline{\theta}, L^o_{it} = L_k) = \beta_k$, $k = 1, 2, 3$.[24] In this way the model of misclassification is characterized by four rates, out of all the possible combinations of

[22] In actual estimation, the bandwidth has been set to 10, based on sensitivity analysis. The procedure is an application of the measurement error technique introduced by McFadden [1989]. See also Keane and Wolpin [1997] and Eeckstein and Wolpin [1999].

[23] Notice that the number of performance ratings simulated in each period is constant across individuals and prior draws.

[24] See Keane and Wolpin [2001] and Keane and Sauer [2003]. When simulating outcomes for given parameter values, the sequence of reported choices with errors is constructed by drawing a sequence $\{U_{it}\}^T_{t=1}$ of $T = 10$ deviates from a uniform number generator for each individual i and comparing these draws with the classification error rates. The comparison determines whether choices are correctly reported, by the following rule: given $R_{it} = 1$, if $U_{it} < \Pr(R^0_{it} = 1|R_{it} = 1, \theta_k, L^o_{it})$, then $R^0_{it} = 1$, and $R^0_{it} = 0$, otherwise. Similarly, given $R_{it} = 0$, if $U_{it} < \Pr(R^0_{it} = 0|R_{it} = 0, \theta_k, L^o_{it})$, then $R^0_{it} = 0$, and $R^0_{it} = 1$, otherwise.

observed and true choices, since

$$\Pr(R_{it}^{o}=1\mid\theta_k,L_{it}^{o}) = \Pr(R_{it}^{o}=1\mid R_{it}=1,\theta_k,L_{it}^{o})\Pr(R_{it}=1\mid\theta_k,L_{it}^{o}) + \Pr(R_{it}^{o}=1\mid R_{it}=0,\theta_k,L_{it}^{o})\Pr(R_{it}=0\mid\theta_k,L_{it}^{o}).$$

In this specification, the classification error is unbiased: the (conditional) probability of observing a high output realization is the same as the (conditional) probability that a good output truly occurs, i.e., $\Pr(R_{it}^{o}=1|\theta_k,L_{it}^{o}) = \Pr(R_{it}=1|\theta_k,L_{it}^{o})$. Unbiasedness implies that the classification rates are linear in the true choice probabilities. As the probability of the true choice converges to one, the probability of the observed choice converges to one as well, i.e., unbiasedness is preserved in the limit since the probability of a correct classification increases linearly from $E_k(\theta_k)$ to one as the true choice probability approaches one. In other words, as $\Pr(R_{it}=1|\theta_k,L_{it}^{o})\to 1$, $\Pr(R_{it}^{o}=1|R_{it}=1,\theta_k,L_{it}^{o})\to 1$. In addition, when the probability of the true choice goes to zero, $E_k(\theta_k)$ approximates the conditional probability of observing the true choice, since $\Pr(R_{it}=1|\theta_k,L_{it}^{o})\to 0$ implies $\Pr(R_{it}^{o}=1|R_{it}=1,\theta_k,L_{it}^{o})\to E_k(\theta_k)$. In this sense $E_k(\theta_k)$ can be interpreted as a base classification error rate. In estimation, $E_k(\theta_k)$ is treated as a free parameter and it is the only parameter on which unbiasedness depends.

Given the assumed error structure for the performance signal, the associated probability of the observed rating is computed as

$$\Pr(R_{it}^{o}=R_r\mid\theta_k,L_{it}^{o}) \simeq \sum_{s=1}^{S}\frac{\Pr(R_{it}^{o}=R_r\mid R_{its}=R_{rs},\theta_k,L_{it}^{o})\Pr(R_{its}=R_{rs}\mid\theta_k,L_{it}^{o})}{S}$$

where R_{rs} denotes the realization of the performance signal at the s-th simulation, with $R_r,R_{rs}\in\{\emptyset,0,1\}$. The sequence $(R_{it1},...,R_{itS})$ of period t simulation draws is then used to compute the vector of period-$t+1$ updated posteriors $(\phi_{t+1}^{j1},...,\phi_{t+1}^{jS})$.

Notice that the entire set of the model parameters enters the likelihood through the choice probabilities, which are computed from solving the firm's dynamic programming problem. The maximization of the likelihood function involves an iterative process between the numerical solution of the firm's dynamic programming problem, for given parameter values, and the computation of the likelihood function.[25]

[25] Observe that, given our simulation technique, as long as the number of simulations, as compared to the number of individuals, grows arbitrarily large, the simulated maximum likelihood estimates are consistent and asymptotically normal. If this ratio is bounded away from infinity, the estimates are still consistent, but the limiting distribution is normal with mean not equal to zero, i.e., there is a bias.

5 Estimation Results

5.1 Parameter Estimates

The qualitative implication of the model that experimenting on a worker's unobserved ability is an important determinant of job to job transitions inside a firm is confirmed by estimation results.[26] Table 4 reports the value of the vector $\psi = (a_\beta, b_\beta, \delta, (\alpha_k, \beta_k, \underline{y}_k, \overline{y}_k, \xi_k, E_k(\underline{\theta}), E_k(\overline{\theta}))_{k=1}^3)$ of structural parameters, estimated from the sample of 502 managers entering the firm at Level 1 between 1970 and 1979, with at least 16 years of education and no level assignment or performance rating missing. Relevant descriptive statistics for the sample are reported in Appendix B.[27]

Table 4. Parameter Estimates

a_β	1.000	$\underline{y}_1$	$-2,446.885$
b_β	1.000	$\underline{y}_2$	$-5,986.493$
α_1	0.869	$\underline{y}_3$	$-880,226.430$
α_2	0.778	ξ_1	-
α_3	0.999	ξ_2	-
β_1	0.069	ξ_3	0.564
β_2	0.000	$E_1(\overline{\theta})$	0.010
β_3	0.700	$E_2(\overline{\theta})$	0.002
δ	0.950	$E_3(\overline{\theta})$	0.111
$\overline{y}_1$	50.940	$E_1(\underline{\theta})$	0.000
$\overline{y}_2$	$3,599.936$	$E_2(\underline{\theta})$	0.987
$\overline{y}_3$	$40,846.745$	$E_3(\underline{\theta})$	0.001

From these estimated values, as predicted by the model the firm's optimal employment policy is an interval belief strategy, which prescribes that the worker be assigned to Level 1 if $\phi \in [0.052, 0.503)$, to Level 2 if $\phi \in [0.503, 0.993)$ and to Level 3 if $\phi \in [0.993, 1]$, but that he be not employed if $\phi \in [0, 0.052)$. A number of theoretical restrictions under which this policy, characterized in Section 2, is the firm's optimal employment policy are also satisfied. Notice first that the distribution of output signals at the three levels is asymmetric across types, i.e., $\alpha_k \neq \beta_k$ for $k = 1, 2, 3$. Moreover, at each job $k = 1, 2, 3$, the distribution of output realizations when the worker is of high ability first-order stochastically dominates the corresponding distribution when he is of low ability, i.e., $\alpha_k > \beta_k$.

[26] Estimation of the model from the sample of 1,870 managers entering the firm at Level 1 between 1970 and 1979, with at least 16 years of education at entry and no level information missing (but censored performance rating histories), is currently pursued. Details can be provided upon request. In the present version of the estimation, separation rates at Levels 1 (ξ_1) and 2 (ξ_2) have been set to zero.

[27] All parameter estimates, except for the classification errors on performance ratings, are significant at standard levels.

This implies, as posited by the model, that observing a high rating improves the firm's assessment that the worker is of high ability. Given these values for α_k and β_k, the estimated size of the output realizations at the three jobs, $\underline{y}_k$ and $\overline{y}_k$, satisfies assumptions (A1)-(A2), i.e., $y_3(\overline{\theta}) > y_2(\overline{\theta}) > y_1(\overline{\theta})$ and $y_1(\underline{\theta}) > y_2(\underline{\theta}) > y_3(\underline{\theta})$, with $y_1(\underline{\theta}) = -2,274.612$, $y_2(\underline{\theta}) = -5,986.075$ and $y_3(\underline{\theta}) = -235,425.955$.

In particular, from the fact that $\Pi = 0 > y_1(\overline{\theta}) = -276.456$, while $y_2(\overline{\theta}) = 1472.475$ and $y_3(\overline{\theta}) = 40,842.563$, it follows that the expected continuation profit from assigning a worker to Level 1 is sufficiently large to compensate the one-period profit loss from employing him. The fact that, at the estimated parameter values, the expected continuation value at Level 2 exceeds the one at Level 1 at each belief also implies that the (gross) informational value at Level 2 is larger than the one at Level 1. The maximal difference is of the order of 1,360, when $y_2(\phi) = 1,472.4$. Finally, a value of $\delta = 0.95$ is consistent with the yearly observations used in estimation, given that it implies an annual interest rate on a risk free asset of 4.75 percent.

5.2 Within-Sample Fit

We will now present evidence of the model's within-sample fit by looking at the distribution of managerial employees across Levels 1, 2 and 3, over the first nine years after entry. The observed and predicted fraction of those managerial workers, entering the firm at Level 1, who are assigned to Levels 1, 2 and 3 or leave the firm, in each of the nine years after entry, are reported in Table 5.

Table 5. Proportion of Employees at Levels 1 and 2, Observed (BGH) and Predicted (DP) (16 or More Years of Education at Entry - 502 Employees)

Years Since Entry	Level 1 (BGH)	Level 1 (DP)	Level 2 (BGH)	Level 2 (DP)	Level 3 (BGH)	Level 3 (DP)	Exit (BGH)	Exit (DP)
0	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.347	0.330	0.327	0.262	0.000	0.000	0.327	0.408
2	0.100	0.058	0.277	0.087	0.042	0.168	0.582	0.686
3	0.034	0.071	0.141	0.021	0.084	0.115	0.741	0.793
4	0.014	0.010	0.074	0.020	0.070	0.064	0.843	0.906
5	0.008	0.009	0.026	0.003	0.048	0.039	0.918	0.949
6	0.002	0.004	0.012	0.003	0.034	0.019	0.952	0.974
7	0.000	0.001	0.006	0.001	0.018	0.010	0.976	0.988
8	0.000	0.001	0.002	0.000	0.008	0.005	0.990	0.994
9	0.000	0.000	0.000	0.000	0.002	0.002	0.998	0.997

As it can be seen, the model succeeds in capturing the dynamic profile of the probability of continuous assignment to Level 1, which is steeply decreasing for the sample of observed employees over the ten year period. The pattern of assignment to Level 2 implied by the model also shares the same qualitative features of the profile observed in the data: it sharply increases the second year after entry and then decreases throughout. The greatest discrepancy between the observed and predicted fraction of managers employed at Level 2 is in the second and third year since entry, when the fraction of employees in the data assigned to Level 2 is significantly greater than the proportion predicted by the model, with a difference, respectively, of 0.19 and 0.12.[28]

As for the observed and predicted fraction of employees who are assigned to Level 3, the hump-shaped pattern observed in the data, increasing in the first three years since entry and then decreasing, is successfully captured by the model. Nonetheless, the proportion of employees assigned to Level 3 in the second and third period since entry, simulated from the model, is substantially larger than the fraction observed in the data. This is a consequence of the fact that the model predicts a smaller proportion of employees at Level 2 in those same periods than the one actually observed, while the observed and predicted exit rate, as well as the observed and predicted proportion of employees at Level 1, are fairly similar in those years. In fact, the fraction of employees leaving the firm predicted by the model is close to the fraction observed in the data, with the largest difference of 0.104 in the second period after entry.

6 The Value of Information

In general there may be no obvious units to measure the amount of information available to a decision maker. The question is nonetheless meaningful in the context of a broader decision problem, which involves choosing an information structure. When a worker is assigned to a job position, the revenue generated in a period is not only the source of the firm profit, but it provides the firm with additional information about the worker's ability, given that the likelihood of observing either a high or a low rating (proxy for high or low revenue) depends on the worker's underlying ability. Specifically, the choice of a job affects the distribution of the performance signals generated in a period and therefore the distribution of the firm's posterior beliefs. In a sense, then, choosing to which job to assign the worker can be viewed as choosing which information to generate about his ability, i.e., which experiment to perform in order to learn about his unobserved human capital.

Notice that, if the firm did not observe the revenue produced by the worker on the job in a period, it would not be able to condition its future assignment decisions on it. In this case, the expected discounted profit from assigning the worker to job k would be

$$\Pi_k(\phi) = y_k(\phi) - U + \delta(1-\xi_k)\Pi_k(\phi) + \delta\xi_k\overline{\Pi}$$

[28]One of the dimensions along which the model fit could be improved is by allowing for the presence of unobserved heterogeneity in the probability of a success at Level 2, for each worker type. This modification would in fact introduce a more flexible parametrization of the determinants of profit at Level 2, which could allow the model to match more closely the pattern of allocation to Level 2 and then to Level 3 at high tenures.

so that $\Pi_k(\phi) = y_k(\phi) - U + \delta\xi_k\overline{\Pi}/[1-\delta(1-\xi_k)]$.[29] However, since the firm can condition its decision of which job to assign the worker in period $t+1$ on the performance signal observed in period t, a natural measure of the (gross) value of information is the extra expected profit, from period $t+1$ on, from choosing task k over the task which maximizes the expected period profit, task s. This value can then be quantified as the difference between the firm's maximal expected continuation profit, function of its current period choice of job k, $E_k[\Pi(\tilde{\phi})|\phi]$, minus $y_s(\phi) - U + \delta\xi_s\overline{\Pi}/[1-\delta(1-\xi_s)]$.[30] Analogously, the firm's willingness to pay for the maximal amount of information, from assigning the worker to each job k, can be measured as the difference, at each belief, between (i) the expected continuation value from the most informative experiment at job k, which would immediately reveal the worker's ability after one period, and (ii) the expected continuation value from assigning the worker to job k.

As discussed the firm incurs an opportunity cost in generating information about a worker's ability. The option value of this information is the expected one-period profit loss the firm incurs by choosing to assign the worker to job k rather than to the most profitable job for that period. Then, the above discussion implies that the net value of information to the firm can be measured as

$$\begin{aligned} I_k(\phi) &\equiv \delta\left\{(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} - \frac{y_s(\phi) - U + \delta\xi_s\overline{\Pi}}{1-\delta(1-\xi_s)}\right\} - [y_s(\phi) - y_k(\phi)] \\ &= y_k(\phi) - U + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} - \frac{(1-\delta\xi_s)[y_s(\phi) - U] + \delta^2\xi_s\overline{\Pi}}{1-\delta(1-\xi_s)} \end{aligned}$$

that is, approximately the extra payoff from the dynamic game over obtaining perpetually the static game profit. Observe that, given that the firm's value function is convex in the posterior belief, the firm always values new information as long as there is uncertainty about the worker's ability, in the sense that it always prefers a riskier distribution of posterior beliefs to a less risky one, as discussed in Section 2.

The objective of the present Section is to assess the impact on the net value of information to the firm and on learning (measured as the probability of retaining a high ability worker) of altering specific parameters of the environment from their estimated values. The effect of changes in the structural parameters of the model on these quantities are in principle not obvious. Modifications of some parameters, namely α_k and β_k, affect directly the informational content of job k. Nevertheless, all the structural parameters δ, α_k, β_k, $\underline{y}_k$, $\overline{y}_k$ and ξ_k, $k = 1, 2, 3$, have an impact on the firm's own *valuation* of information, since they affect the degree of convexity of $\Pi(\cdot)$, and, in this way, the value of information.

[29] If payoffs were normalized by $1-\delta$, so as to be expressed as per period averages, and the separation rate at each job was zero, this value would be the same as the firm's profit from the static game, i.e., on average the firm would receive his period profit.

[30] See this approach in Chade and Schlee [2002] for a discussion.

6.1 Changes in the Value of Information and Incomplete Learning

One of the purposes of the estimation exercise is to determine the value of endogenous information acquisition to the firm, and to quantitatively evaluate the changes in this value under alternative scenarios, simulated from the benchmark case, in which parameters are fixed at their estimated values. Estimates of the parameters of the model also allow to uncover the equilibrium relationship between the value of information and the amount of learning which takes place through employment, measured as the probability of retaining a high ability worker at either Level 1, Level 2 or Level 3 in each period. An important dimension along which counterfactual experiments are of interest is therefore in assessing the impact of changes in the value of information on the (inefficient) turnover of high ability workers.

Understanding the effect of learning on firm-level allocation decisions has significant implications for the labor market experience of workers and for a firm's incentive to employ them. It also enables us to make predictions about the effectiveness of policies that aim at improving monitoring of performance, i.e., the quality of the information generated through employment, at each level of the firm's hierarchy. The counterfactual evaluations we will perform aim specifically at investigating: (i) the impact on the value of information acquisition to the firm, and (ii) the resulting comparative dynamic effect on workers' career prospects, of:

(1) changes in the firm's degree of time impatience, which parameterizes the firm's incentive to experiment on ability, to $\delta = 0.50$ and $\delta = 0.99$;

(2) an increase in the precision of prior information, i.e., a reduction in the dispersion of the distribution of possible prior beliefs about the worker's unobserved ability, to $a_\beta = a_\beta = 50$;

(3) an increase in the accuracy of the firm's monitoring technology, i.e., the probability of a high rating for a worker of either ability at Levels 1 and 2;

(4) a reduction to zero in the size of the output realizations and in the probability of success for each type of worker at Level 2, i.e., the case in which only Level 1, the entry job, and Level 3, the statically most profitable job, are available.

6.2 Experiment 1: Different Degrees of Time Patience

In order to illustrate the value of information acquisition implied by the parameter estimates, we compare the model's prediction on the distribution of employees across the three managerial levels, together with the fraction of employees leaving the firm, with the predictions from a model in which $\delta = 0.50$, i.e., intermediate degree of time impatience, and a model in which $\delta = 0.99$, i.e., close to maximal time patience. Tables 6 and 7 report the predicted fraction of employees at each level and leaving the firm in each year since entry, for the benchmark case and the simulated scenarios.

As expected, when the discount factor is $\delta = 0.50$, the firm's willingness to employ workers decreases, since the value of current information for the profitability of future assignment decisions is smaller. Indeed, the range of beliefs for which employment is profitable decreases, i.e., the lowest belief for which the firm is willing to employ a worker at Level 1 is approximately $\phi = 0.20$, as compared to $\phi = 0.05$ in the benchmark case. The firm's degree of time impatience has a significant effect on the pattern of exit as well: for the same rate of exogenous separations, the fraction of managers who will leave the firm after the first period almost doubles, from 0.481 in the benchmark case to 0.876.

Table 6. Predicted Fraction of Employees at Levels 1 and 2

Years Since Entry	**Level 1** ($\delta = 0.95$)	**Level 1** ($\delta = 0.50$)	**Level 1** ($\delta = 0.99$)	**Level 2** ($\delta = 0.95$)	**Level 2** ($\delta = 0.50$)	**Level 2** ($\delta = 0.99$)
0	1.000	1.000	1.000	0.000	0.000	0.000
1	0.330	0.000	0.525	0.262	0.241	0.314
2	0.058	0.000	0.174	0.087	0.000	0.291
3	0.071	0.000	0.093	0.021	0.000	0.309
4	0.010	0.000	0.046	0.020	0.000	0.322
5	0.009	0.000	0.019	0.003	0.000	0.319
6	0.004	0.000	0.010	0.003	0.000	0.324
7	0.001	0.000	0.005	0.001	0.000	0.324
8	0.001	0.000	0.002	0.000	0.000	0.324
9	0.000	0.000	0.001	0.000	0.000	0.324

Table 7. Predicted Fraction of Employees at Level 3 and Leaving the Firm

Years Since Entry	**Level 3** ($\delta = 0.95$)	**Level 3** ($\delta = 0.50$)	**Level 3** ($\delta = 0.99$)	**Exit** ($\delta = 0.95$)	**Exit** ($\delta = 0.50$)	**Exit** ($\delta = 0.99$)
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.408	0.759	0.162
2	0.168	0.145	0.000	0.686	0.759	0.534
3	0.115	0.095	0.000	0.793	0.905	0.598
4	0.064	0.041	0.000	0.906	0.959	0.631
5	0.039	0.018	0.000	0.949	0.982	0.662
6	0.019	0.008	0.000	0.974	0.992	0.666
7	0.010	0.004	0.000	0.988	0.996	0.672
8	0.005	0.001	0.000	0.994	0.999	0.674
9	0.002	0.001	0.000	0.997	0.999	0.674

When $\delta = 0.99$, the cut-off belief which makes the firm willing to assign the worker to Level 3 is almost one. This implies that most of the workers, whose ability is sufficiently high for being retained at the firm, are assigned to Level 2 for a longer period of time. In particular, in the first ten periods of employment none of them is assigned to Level 3. This has also a clear impact on the fraction of employees leaving the firm. The fact that, when a worker is of high skill, output realizations are noisier signals of ability at Level 2 than at Level 3 (i.e., $\alpha_2 < \alpha_3$), together with the fact that Level 2 is profitable for higher belief values than in the benchmark case, imply that employees at Level 2 are more likely to be retained rather than fired, for the same sequence of observed ratings. This is reflected in the smaller fraction of employees leaving the firm after the third period since entry.

When $\delta = 0.50$, as compared to the case in which $\delta = 0.95$, the value of information is higher than in the benchmark case for any $\phi \geq 0.26$ and the increase can be as large as of the order of 444 percent. This result is due to the fact that, even if in principle the firm values information more when it is less time impatient, given that it attaches a greater weight to his future expected profit, the fact that exogenous separations at Level 3 occur with high probability depresses significantly the firm's expected discounted profit from assigning an employee to any job.

As expected, when $\delta = 0.99$ the firm has nevertheless a stronger incentive to employ a worker to learn about his ability. In fact, employment starts being profitable for the firm when $\phi = 0.002$, where the firm's expected discounted profit increases by as much as 400 percent. Otherwise, the change in the value of information, compared to the benchmark case, ranges from approximately 0 percent, when $\phi = 0.075$, to -60 percent, when $\phi = 0.85$.[31]

6.3 Experiment 2: Increased Precision of Prior Information

Recall that a beta distribution with parameters $a_\beta = b_\beta = 1$ parameterizes the set of prior distributions for the firm and a worker over the worker's true ability. The variance of a beta distribution with parameters a_β and b_β is equal to $a_\beta b_\beta/((a_\beta + b_\beta)^2(a_\beta + b_\beta + 1))$, so that, when $a_\beta = b_\beta = 50$ as compared to when $a_\beta = b_\beta = 1$, it decreases from 0.083 to 0.002.

An increase in a_β and b_β is then equivalent to a reduction in the dispersion of prior beliefs about a newly hired worker, still consistent with the worker being assigned to Level 1 at entry. Given the estimated values of the parameters, however, a reduction of $5,000$ percent in this dispersion, has no significant impact on the dynamics of job assignment inside the firm, a part for a decrease in the fraction of workers leaving the firm in the second period after entry.

[31] However, the difference in the values of information is non monotonic at high belief values.

Table 8. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm

Years Since Entry	**Level 1** ($a_\beta = 1$) ($b_\beta = 1$)	**Level 1** ($a_\beta = 50$) ($b_\beta = 50$)	**Level 2** ($a_\beta = 1$) ($b_\beta = 1$)	**Level 2** ($a_\beta = 50$) ($b_\beta = 50$)	**Level 3** ($a_\beta = 1$) ($b_\beta = 1$)	**Level 3** ($a_\beta = 50$) ($b_\beta = 50$)	**Exit** ($a_\beta = 1$) ($b_\beta = 1$)	**Exit** ($a_\beta = 50$) ($b_\beta = 50$)
0	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.330	0.625	0.262	0.291	0.000	0.000	0.408	0.085
2	0.058	0.080	0.087	0.093	0.168	0.186	0.686	0.641
3	0.071	0.103	0.021	0.032	0.115	0.119	0.793	0.746
4	0.010	0.000	0.020	0.029	0.064	0.740	0.906	0.898
5	0.009	0.013	0.003	0.000	0.039	0.048	0.949	0.939
6	0.004	0.006	0.003	0.004	0.019	0.021	0.974	0.969
7	0.001	0.001	0.001	0.002	0.010	0.012	0.988	0.985
8	0.001	0.001	0.000	0.000	0.005	0.006	0.994	0.992
9	0.000	0.000	0.000	0.000	0.002	0.003	0.997	0.997

6.4 Experiment 3: Increased Accuracy of Performance Monitoring

Suppose now that the probability of a high performance rating is one for a high ability worker and zero for a low ability worker, at either Level 1 or Level 2.[32] Since in the model the firm's production and monitoring technology coincide, a change in the probability of success for each type amounts to a change in the one-period expected revenue at either Levels 1 or 2, given that the same output realizations occur with different probabilities, as well as in the expected continuation profit from either level, given that the variance in posterior beliefs, when the output signal is perfectly informative, is maximal.

Compare the distribution of workers across levels when Level 1 is perfectly informative about the worker's true ability and in the benchmark case. As expected, the fact that observing a worker's performance for one period at Level 1 perfectly reveals his human capital makes the use of Level 1 profitable only the first year after entry. In case the worker is of low ability, in fact, the firm is better off by firing him than employing him at Level 1 afterwards, while, if the worker is of high ability, assigning him to Level 3 is for the firm the best alternative. As a consequence, then, the fraction of employees terminated is higher than in the benchmark case. Similarly, given that there is no informational value for the firm from assigning the worker to Level 2, after one period at Level 1 retained employees are only assigned to Level 3. This follows from the fact that $y_3(\overline{\theta}) = 40,842.562 > y_2(\overline{\theta}) = 1,472.475$, i.e., a high ability worker is more profitable for the firm when assigned to Level 3

[32] In the simulation of the two experiments, the probabilities of success were set to $\alpha_1 = 9 \cdot 10^{-7}$ and $\beta_1 = 10^{-7}$, in case Level 1 is assumed to be perfectly informative about the worker's ability, and to $\alpha_2 = 9 \cdot 10^{-7}$ and $\beta_2 = 10^{-7}$, in case Level 2 is supposed to be perfectly informative.

than to Level 2. In this case, the change in the value of information can be as large as 563 percent, and it decreases from approximately 1, 112 percent to 0 percent.

Table 9. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm

Years (Since Entry)	**Level 1**	**Level 1** ($\alpha_1 = 1$) ($\beta_1 = 0$)	**Level 2**	**Level 2** ($\alpha_1 = 1$) ($\beta_1 = 0$)	**Level 3**	**Level 3** ($\alpha_1 = 1$) ($\beta_1 = 0$)	**Exit**	**Exit** ($\alpha_1 = 1$) ($\beta_1 = 0$)
0	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.330	0.000	0.262	0.000	0.000	0.462	0.408	0.538
2	0.058	0.000	0.087	0.000	0.168	0.203	0.686	0.797
3	0.071	0.000	0.021	0.000	0.115	0.089	0.793	0.911
4	0.010	0.000	0.020	0.000	0.064	0.039	0.906	0.961
5	0.009	0.000	0.003	0.000	0.039	0.017	0.949	0.983
6	0.004	0.000	0.003	0.000	0.019	0.008	0.974	0.992
7	0.001	0.000	0.001	0.000	0.010	0.003	0.988	0.997
8	0.001	0.000	0.000	0.000	0.005	0.001	0.994	0.999
9	0.000	0.000	0.000	0.000	0.002	0.001	0.997	0.999

Table 10. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm

Years Since Entry	**Level 1**	**Level 1** ($\alpha_2 = 1$) ($\beta_2 = 0$)	**Level 2**	**Level 2** ($\alpha_2 = 1$) ($\beta_2 = 0$)	**Level 3**	**Level 3** ($\alpha_2 = 1$) ($\beta_2 = 0$)	**Exit**	**Exit** ($\alpha_2 = 1$) ($\beta_2 = 0$)
0	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.330	0.000	0.262	0.164	0.000	0.000	0.408	0.836
2	0.058	0.000	0.087	0.103	0.168	0.000	0.686	0.897
3	0.071	0.000	0.021	0.103	0.115	0.000	0.793	0.897
4	0.010	0.000	0.020	0.103	0.064	0.000	0.906	0.897
5	0.009	0.000	0.003	0.103	0.039	0.000	0.949	0.897
6	0.004	0.000	0.003	0.103	0.019	0.000	0.974	0.897
7	0.001	0.000	0.001	0.103	0.010	0.000	0.988	0.897
8	0.001	0.000	0.000	0.103	0.005	0.000	0.994	0.897
9	0.000	0.000	0.000	0.103	0.002	0.000	0.997	0.897

As it can be seen from Table 10, the pattern of assignments to Level 1 when, instead, Level 2 perfectly reveals a worker's true skill, is analogous to the case in which Level 1 is perfectly informative. Because of the gain from assigning a high ability worker to Level 2 or from dismissing a low ability worker, no employee is retained at Level 1 after the first period.

As a difference from the previous case, though, the fraction of employees at Level 2 does not decrease, a part from the third year after entry. This is due to the fact that, given that $\beta_2 = 0.000$ and α_2 is relatively large, observing a low rating at high beliefs (at which Level 2 is still the best assignment) implies a small belief revision. The corresponding increase in the value of information to the firm ranges from 480.6 percent, when $\phi = 0.005$, to approximately 0 percent, when the firm knows with certainty the worker's ability.

6.5 Experiment 4: A Two-Job Hierarchy

The last experiment performed is to assume that the firm's hierarchy only consists of Level 1, the entry level, and Level 3. Recall that Level 3 is the most profitable job position if the worker is truly of high ability, but it is also the one at which the firm incurs the greatest one-period profit loss if the worker's actual ability is low. The experiment is performed by setting $\underline{y}_2 = \overline{y}_2 = \alpha_2 = \beta_2 = 0$.

Table 11. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm

Years Since Entry	Level 1 (Three Levels)	Level 1 (Two Levels)	Level 3 (Three Levels)	Level 3 (Two Levels)	Exit (Three Levels)	Exit (Two Levels)
0	1.000	1.000	0.000	0.000	0.000	0.000
1	0.330	0.344	0.000	0.424	0.408	0.232
2	0.058	0.069	0.168	0.269	0.686	0.662
3	0.071	0.002	0.115	0.130	0.793	0.868
4	0.010	0.001	0.064	0.058	0.906	0.941
5	0.009	0.000	0.039	0.025	0.949	0.975
6	0.004	0.000	0.019	0.011	0.974	0.989
7	0.001	0.000	0.010	0.005	0.988	0.995
8	0.001	0.000	0.005	0.002	0.994	0.998
9	0.000	0.000	0.002	0.001	0.997	0.999

In this case, workers whose assessed ability is sufficiently high, so that they are retained at the firm, are assigned to Level 3 rather than to Level 1. Since at Level 3 performance outcomes are imperfect signals of ability, workers who receive low ratings are more likely to be terminated than in the benchmark case, given that on average retained employees are assigned to Level 3 for lower belief values, i.e., when the impact of a low output signal on posterior beliefs is still significant. This is then the reason why in the second year since entry employees exit more often than in the benchmark case, even if the proportion of employees assigned to Level 1 does not change.

The fact that, for intermediate belief values, the firm can only assign the worker to Level 3, with the risk of a greater output destruction than at Level 2 if the worker is of low ability (i.e.,

$y_3(\overline{\theta}) - y_3(\overline{\theta}) > y_2(\overline{\theta}) - y_2(\overline{\theta}))$ makes the firm less willing to employ a worker than in the benchmark case (employment starts at $\phi = 0.053$, compared to $\phi = 0.053$ in the benchmark case), with a corresponding reduction in the value of information of the order of 100 percent, when employment starts to be profitable, to almost 0 percent, when $\phi = 1$.

7 Related Literature

There are several related strands of literature. A number of papers, following Jovanovic [1979a, 1979b], have applied the one-armed Bandit framework to the study of turnover across firms.[33] Seminal paper on multi-tasking is Holmström and Milgrom [1991]. They investigate a multi-task principal-agent model in the presence of moral hazard and interpret job design as an instrument to control incentives, rather than a mechanism to generate information about an agent's unobserved ability. The implications of their model for the theory of job design is that, if measurement errors (as captured by the noise to signal ratio of performance per unit of time at a task) are correlated across tasks, grouping tasks with different performance characteristics in the same job is optimal. In this case grouping tasks allows the use of comparative performance evaluation, which in turn helps reducing the risk premium incurred by the principal in providing incentives. In our framework, on the contrary, different jobs may consist of tasks with dependent measurement errors, given that a worker's unobserved ability is correlated across tasks and, being tasks dynamically complementary in information production, workers can be assigned to different tasks only as their tenure at the firm increases.

Among the contributions which analyze job assignment inside firms, Prendergast [1993] rationalizes promotions as an equilibrium device to reward the nonverifiable acquisition of firm-specific human capital. Waldman [1984b] focuses instead on the distortions in the equilibrium assignment process which arise when promotions serve as a public signal to the market about a worker's unobserved ability. Fairburn and Malcomson [2001] offer an interpretation for the relationship between performance, incentives and promotions based on the conflict of interest between managers and firm. Specifically, they study the relative incentive power of promotions and monetary transfers, when supervisors of employees are subject to influence activities on the part of employees, and show that the use of promotions reduces the incentive for managers to be affected by them.[34]

The paper closest in spirit to ours is Gibbons and Waldman [1999b]. They develop a model of learning, job assignment and human capital acquisition which accounts for a broad pattern of evidence on wage and promotion dynamics inside firms. They assume that there exists an output interaction between learning and human capital acquisition, which both affect a worker's expected

[33] See also the discussion contained in the companion paper.

[34] There is finally a number of papers which study the interaction of strategic aspects of the oligopoly problem with a decision maker's incentive to experiment and analyze the degree of efficiency of market experimentation. An example is Bergemann and Välimäki [1996].

product in a period. In their model, as in ours, an equilibrium hierarchy of job positions results from the assumption that higher ability is more valuable at higher level jobs. Since human capital is accumulated by experience, all workers eventually reach the highest job position in the firm's hierarchy as they age. Because of learning on the part of the firm and the accumulation of skills on the part of the worker, demotions are rare. The main differences between our framework and theirs are that in our case (i) the job performed by a worker affects the rate of learning about ability, and (ii) a worker of low ability is nonprofitable for the firm. Experimentation on ability affects then dynamic screening both through retention and job assignment. In particular, in our framework workers move up the job ladder purely as a consequence of the firm's improved estimate of their ability. However, because of the informational value of lower level jobs only when uncertainty about ability is highest, demotions can be rare.

On the empirical side, due to the confidentiality of the data required, only a few studies analyze intra-firm job transitions or wage dynamics. Baker, Gibbs and Holmström [1994a] provide a detailed case study analysis of the data from which our estimation sample has been selected, finding evidence for the hypothesis that a firm's internal hierarchy acts as an information acquisition filter, to screen employees according to their unobserved abilities. Baker, Gibbs and Holmström [1994b] test whether existing explanations for wage dynamics, specifically on-the-job training, learning and stationary incentive models, are consistent with the wage policy they infer from the data. They find that none of these models can alone be reconciled with the patterns emerging from the data. Chiappori, Salanié and Valentin [1999] consider a model of wage formation, in the presence of learning and downward wage rigidity, and find evidence of a 'late beginner' property in the dynamics of wages, i.e., after controlling for the wage at t, the wage at $t+1$ is negatively correlated with the wage at $t-1$. Focussing on the analysis of short term turnover, Nagypal [2003] adopts a structural estimation approach to test the relative explanatory power of learning about ability versus learning on the job in determining the intertemporal profile of the hazard rate of employment termination, using a French matched employer-employee dataset. Her estimation results support a learning interpretation for inter-firm job transitions.[35]

Finally, analyzing the same dataset we are using, De Varo and Waldman [2004] test the empirical validity of the hypothesis that promotions signal imperfectly individual ability to the market, so that a worker's current employer and alternative firms share asymmetric information about the worker's actual productivity. They find support both for the asymmetric and for the symmetric learning hypotheses.

[35] Kwon [2004], on the other hand, tries to assess the relative importance of sorting and incentive provision in shaping the dynamic profile of the probability of dismissal. His estimation results provide evidence for the incentive model.

8 Conclusion

This paper has developed a learning model of retention and job assignment and provided a structural estimation of it, using ten years of observations on level assignments and performance ratings for the cohorts of managers employed at a single U.S. service firm between 1970 and 1979. Estimation results confirm that a firm's internal hierarchy can act as an information acquisition filter, with performance at lower level jobs being used by the firm to learn about workers' true productivity, for the benefit of future assignment decisions. The sequential screening mechanism, which characterizes the firm's employment policy in the equilibria of interest, has been shown to be also a property of the promotion dynamics estimated from the data. In particular, the estimated retention and task assignment policy is the one predicted by the model. Overall, the model succeeds in fitting the probability of retention and promotion at the job positions at which most managers are employed over the sample period.

A number of stylized facts have been documented in the literature on the internal economics of the firm about the dynamics of wages and promotions (see Gibbons and Waldman [1999a, 1999b] for a comprehensive reference). The interaction of outside labor market competition with a firm's incentive to experiment on workers' ability is an important determinant of job dynamics within and across firms, but also of individual wage variability at each level within a firm hierarchy.[36] An issue of interest within a learning framework is also the extent to which the gradual assignment of employees to higher level jobs, at which ability is more valuable, is the result of firm's learning about workers' ability or can be attributed to workers acquiring new skills on the job. The exploration of these issues constitutes the specific object of present and future research.

Appendix A

Proof of Proposition 1: The fact that $\Pi(\cdot)$ is well-defined and continuous can be shown by a standard Contraction Mapping argument. Under (A1) and (A2), it can also be shown that it is increasing. As for the convexity of $\Pi(\cdot)$, the proof is adapted from Banks and Sundaram [1992a]. Recall

$$\begin{aligned}\Pi(\phi) \;=\; & \max\{\overline{\Pi}, p_1(\phi)[\overline{y}_1 + \delta(1-\xi_1)\Pi(\phi_{1h}(\phi))] + (1-p_1(\phi))[\underline{y}_1 + \delta(1-\xi_1)\Pi(\phi_{1l}(\phi))] + \delta\xi_1\overline{\Pi}, \\ & p_2(\phi)[\overline{y}_2 + \delta(1-\xi_2)\Pi(\phi_{2h}(\phi))] + (1-p_2(\phi))[\underline{y}_2 + \delta(1-\xi_2)\Pi(\phi_{2l}(\phi))] + \delta\xi_2\overline{\Pi}, \\ & p_3(\phi)[\overline{y}_3 + \delta(1-\xi_3)\Pi(\phi_{3h}(\phi))] + (1-p_3(\phi))[\underline{y}_3 + \delta(1-\xi_3)\Pi(\phi_{3l}(\phi))] + \delta\xi_3\overline{\Pi}\}.\end{aligned}$$

[36] See the comment in Subsection 4.2 on the current estimation of the wage process.

Define the mappings T_k, $k=1,2,3$, and T on $\mathcal{C}[0,1]$, the space of continuous functions on the unit interval, as follows. For $k=1,2,3$, let

$$\begin{aligned}
T_k f(\phi) &= p_k(\phi)[\overline{y}_k + \delta(1-\xi_k)f(\phi_{kh}(\phi))] + (1-p_k(\phi))[\underline{y}_k + \delta(1-\xi_k)f(\phi_{kl}(\phi))] + \delta\xi_k\overline{\Pi} \\
&= p_k(\phi)\overline{y}_k + (1-p_k(\phi))\underline{y}_k + \delta G_k f(\phi), \\
G_k f(\phi) &= (1-\xi_k)[p_k(\phi)f(\phi_{kh}(\phi)) + (1-p_k(\phi))f(\phi_{kl}(\phi))] + \xi_k\overline{\Pi}
\end{aligned}$$

and $Tf(\phi) = \max\{\overline{\Pi}, T_1 f(\phi), T_2 f(\phi), T_3 f(\phi)\}$. We will proceed in two steps. We will first show that, if f is convex, then Tf is also convex. Let $\phi', \phi'' \in [0,1]$, $\lambda \in (0,1)$ and $\phi^* \equiv (1-\lambda)\phi' + \lambda\phi''$. Define $e(y_{kh}) \equiv \frac{\lambda p_k(\phi'')}{p_k(\phi^*)} \in (0,1)$ and $e(y_{kl}) \equiv \frac{\lambda(1-p_k(\phi''))}{1-p_k(\phi^*)} \in (0,1)$. Equivalently, $1-e(y_{kh}) \equiv \frac{(1-\lambda)p_k(\phi')}{p_k(\phi^*)} \in (0,1)$ and $1-e(y_{kl}) \equiv \frac{(1-\lambda)(1-p_k(\phi'))}{1-p_k(\phi^*)} \in (0,1)$. Note that

$$\begin{aligned}
(1-e(y_{kh}))\phi_{kh}(\phi') + e(y_{kh})\phi_{kh}(\phi'') &= \frac{(1-\lambda)\alpha_k\phi' + \lambda\alpha_k\phi''}{p_k(\phi^*)} = \phi_{kh}(\phi^*) \\
(1-e(y_{kl}))\phi_{kl}(\phi') + e(y_{kl})\phi_{kl}(\phi'') &= \frac{(1-\lambda)(1-\alpha_k)\phi' + \lambda(1-\alpha_k)\phi''}{1-p_k(\phi^*)} = \phi_{kl}(\phi^*)
\end{aligned}$$

by definition of $\phi_{kh}(\phi)$ and $\phi_{kl}(\phi)$, $k=1,2,3$. Suppose f is convex. For $k=1,2,3$, from Jensen's inequality for convex functions, it follows

$$\begin{aligned}
G_k f(\phi^*) &= (1-\xi_k)[p_k(\phi^*)f(\phi_{kh}(\phi^*)) + (1-p_k(\phi^*))f(\phi_{kl}(\phi^*))] + \xi_k\overline{\Pi} \\
&\le (1-\xi_k)p_k(\phi^*)[(1-e(y_{kh}))f(\phi_{kh}(\phi')) + e(y_{kh})f(\phi_{kh}(\phi''))] \\
&\quad +(1-\xi_k)(1-p_k(\phi^*))[(1-e(y_{kl}))f(\phi_{kl}(\phi')) + e(y_{kl})f(\phi_{kl}(\phi''))] + \xi_k\overline{\Pi}.
\end{aligned}$$

Note that $p_k(\phi^*)(1-e(y_{kh})) = (1-\lambda)p_k(\phi')$, $p_k(\phi^*)e(y_{kh}) = \lambda p_k(\phi'')$, $(1-p_k(\phi^*))(1-e(y_{kl})) = (1-\lambda)(1-p_k(\phi'))$ and $(1-p_k(\phi^*))e(y_{kl}) = \lambda(1-p_k(\phi''))$. Rearranging terms,

$$\begin{aligned}
G_k f(\phi^*) &\le (1-\xi_k)[(1-\lambda)p_k(\phi')f(\phi_{kh}(\phi')) + \lambda p_k(\phi'')f(\phi_{kh}(\phi''))] \\
&\quad +(1-\xi_k)[(1-\lambda)(1-p_k(\phi'))f(\phi_{kl}(\phi')) + \lambda(1-p_k(\phi''))f(\phi_{kl}(\phi''))] + \xi_k\overline{\Pi} \\
&= (1-\lambda)[(1-\xi_k)p_k(\phi')f(\phi_{kh}(\phi')) + (1-\xi_k)(1-p_k(\phi'))f(\phi_{kl}(\phi')) + \xi_k\overline{\Pi}] \\
&\quad +\lambda[(1-\xi_k)p_k(\phi'')f(\phi_{kh}(\phi'')) + (1-\xi_k)(1-p_k(\phi''))f(\phi_{kl}(\phi'')) + \xi_k\overline{\Pi}] \\
&= (1-\lambda)G_k f(\phi') + \lambda G_k f(\phi'').
\end{aligned}$$

Observe that $p_k(\phi^*) = (1-\lambda)p_k(\phi') + \lambda p_k(\phi'')$ and $1-p_k(\phi^*) = (1-\lambda)(1-p_k(\phi')) + \lambda(1-p_k(\phi''))$. From this,

$$\begin{aligned}
T_k f(\phi^*) &= p_k(\phi^*)\overline{y}_k + (1-p_k(\phi^*))\underline{y}_k + \delta G_k f(\phi^*) \\
&\le [(1-\lambda)p_k(\phi') + \lambda p_k(\phi'')]\overline{y}_k + [(1-\lambda)(1-p_k(\phi')) + \lambda(1-p_k(\phi''))]\underline{y}_k \\
&\quad +\delta(1-\lambda)G_k f(\phi') + \delta\lambda G_k f(\phi'') \\
&= (1-\lambda)T_k f(\phi') + \lambda T_k f(\phi'').
\end{aligned}$$

As the maximum of convex functions, Tf is convex whenever f is convex.

As for the second step, we will prove that the (unique) fixed point of the mapping T is also convex. Let $\mathcal{CX}$ be the set of all convex functions f such that $f \leq Tf$. Note that $\mathcal{CX}$ is bounded above and non-empty. Let $f^* \equiv \sup\{f(\phi)|f \in \mathcal{CX}\}$. As the pointwise supremum of convex functions, f^* is convex. Observe that T is a monotone increasing operator. Then, $f^*(\phi) = \sup\{f(\phi)|f \in \mathcal{CX}\} \leq \sup\{Tf(\phi)|f \in \mathcal{CX}\}$, by definition of $\mathcal{CX}$. Also, by monotonicity of T, $\sup\{Tf(\phi)|f \in \mathcal{CX}\} \leq Tf^*(\phi)$. These combined observations imply $f^* \leq Tf^*$ or $f^* \in \mathcal{CX}$. Recall that, for all $f \in \mathcal{CX}$, $f \leq Tf$. Thus, by monotonicity of T, $Tf \leq T(Tf)$, which implies $Tf \in \mathcal{CX}$ if $f \in \mathcal{CX}$. In particular, $Tf^* \in \mathcal{CX}$. Therefore, by the definition of f^*, it must be $f^* \geq Tf^*$. This, together with $f^* \leq Tf^*$, yields $Tf^* = f^*$ or, equivalently, f^* is a fixed point of the mapping T. But since T is a contraction, it has a unique fixed point. This completes the proof of the claim.

Proof of Proposition 2: Let ϕ_1^* be the cut-off belief value which makes the firm indifferent between not hiring the worker and employing him at task 1, i.e., $\overline{\Pi} = y_1(\phi_1^*) - U + \delta E_1\Pi(\phi_1^*)$. Notice that $\phi'' > \phi'$ implies that the distribution of the updated posterior conditional on ϕ'', after revenue realizes at a job, first-order stochastically dominates the one conditional on ϕ'. Then, $E_1\Pi(\cdot)$ is increasing in ϕ if $\Pi(\cdot)$ is increasing. Since $y_1(\cdot)$ is strictly increasing, it follows that $\Pi_1(\cdot)$ is also strictly increasing (by a similar argument, it can be shown that $\Pi_2(\cdot)$ and $\Pi_3(\cdot)$ are strictly increasing in ϕ as well). Thus, ϕ_1^* is uniquely determined and, with $y_1(\overline{\theta}) > \Pi > y_1(\underline{\theta})$, $\phi_1^* \in (0, 1)$. Suppose now that the following condition holds

$$\Pi_1(\phi_1^*) = y_1(\phi_1^*) - U + \delta E_1\Pi(\phi_1^*) > \Pi_2(\phi_1^*) = y_2(\phi_1^*) - U + \delta E_2\Pi(\phi_1^*). \tag{4}$$

Then, together with $y_2(\overline{\theta}) > \overline{\Pi} > y_2(\underline{\theta})$, condition (4) yields

$$\Pi_2(\overline{\theta}) \equiv y_2(\overline{\theta}) - U + \frac{y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}}{1 - \delta(1 - \xi_3)} > \overline{\Pi} > \Pi_2(\phi_1^*) = y_2(\phi_1^*) - U + \delta E_2\Pi(\phi_1^*)$$

which implies that there exists a unique value $\phi_{0,2}^* \in (0, 1)$, with $\phi_{0,2}^* > \phi_1^*$, such that $\overline{\Pi} = \Pi_2(\phi_{0,2}^*)$. Moreover, since $y_2(\overline{\theta}) > y_1(\overline{\theta})$, from (4) and

$$\Pi_2(\overline{\theta}) > \Pi_1(\overline{\theta}) = y_1(\overline{\theta}) - U + \frac{\delta[y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}]}{1 - \delta(1 - \xi_3)}$$

it follows that there exists a value $\phi_2^* \in (0, 1)$, with $\phi_2^* > \phi_1^*$, satisfying

$$y_1(\phi_2^*) - U + \delta E_1\Pi(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2\Pi(\phi_2^*)$$

so that tasks 1 and 2 are equally profitable. Since $\phi_2^* > \phi_1^*$, by definition of ϕ_1^* it follows $\Pi(\phi_2^*) > \overline{\Pi}$ and, then, $\phi_2^* > \phi_{0,2}^*$. Suppose now that the following condition holds as well

$$\Pi_2(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2\Pi(\phi_2^*) > \Pi_3(\phi_2^*) = y_3(\phi_2^*) - U + \delta E_3\Pi(\phi_2^*). \tag{5}$$

With $y_3(\overline{\theta}) > y_2(\overline{\theta})$ and $\phi_2^* < 1$,

$$\Pi_3(\overline{\theta}) \equiv \frac{y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}}{1 - \delta(1 - \xi_3)} > \Pi_2(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2\Pi(\phi_2^*)$$

which, together with (5), implies that there exists a value $\phi_3^* \in (0,1)$, with $\phi_3^* > \phi_2^*$, satisfying

$$y_2(\phi_3^*) - U + \delta E_2\Pi(\phi_3^*) = y_3(\phi_3^*) - U + \delta E_3\Pi(\phi_3^*) > \overline{\Pi}.$$

Observe that, if the difference $\Pi_2(\phi) - \Pi_1(\phi)$ is strictly increasing, when $\phi \in [\phi_1^*, \phi_3^*]$, and the difference $\Pi_3(\phi) - \Pi_2(\phi)$ is strictly increasing, when $\phi \in [\phi_2^*, 1]$, then the cut-off values ϕ_2^* and ϕ_3^* are uniquely determined. What we will show next is first that, under the conditions stated in the Proposition, (4) and (5) hold and, then, that $\Pi_2(\phi)-\Pi_1(\phi)$ and $\Pi_3(\phi)-\Pi_2(\phi)$ are strictly increasing in ϕ over the specified belief ranges.

Let $k = 1, 2$. Notice that, if $\alpha_k\beta_{k+1} > \alpha_{k+1}\beta_k$, $\phi_{hk}(\phi) > \phi_{hk+1}(\phi)$ for $\phi \in (0,1)$. Similarly, $\alpha_{k+1} > \beta_{k+1}$ and $\alpha_k > \beta_k$ imply, respectively, $\phi_{hk+1}(\phi) > \phi_{lk+1}(\phi)$ and $\phi_{hk}(\phi) > \phi_{lk}(\phi)$, if $\phi \in (0,1)$. Moreover, if $(1-\alpha_{k+1})(1-\beta_k) > (1-\alpha_k)(1-\beta_{k+1})$, it follows $\phi_{lk+1}(\phi) > \phi_{lk}(\phi)$ for $\phi \in (0,1)$. A sufficient condition for $\alpha_k\beta_{k+1} > \alpha_{k+1}\beta_k$ and $(1-\alpha_{k+1})(1-\beta_k) > (1-\alpha_k)(1-\beta_{k+1})$ to hold is $\alpha_k \geq \alpha_{k+1}$ and $\beta_{k+1} \geq \beta_k$. Consider now the distributions of the next period value of ϕ, ϕ', conditional on its current period value and the worker being assigned to tasks k or $k+1$. Denote the two corresponding cumulative distribution functions, respectively, by $F(\phi'; k)$ and $G(\phi'; k+1)$. Observe that the mean of the two distributions is ϕ. Now, the fact that $\phi_{hk}(\phi) > \phi_{hk+1}(\phi) > \phi_{lk+1}(\phi) > \phi_{lk}(\phi)$, and $F(\phi'; k)$ and $G(\phi'; k+1)$ are two-outcome distributions, implies that $F(\phi'; k)$ constitutes a mean-preserving spread of $G(\phi'; k+1)$. Equivalently, $G(\phi'; k+1)$ second-order stochastically dominates $F(\phi'; k)$. By definition, for any two distributions $F(x)$ and $G(x)$ with the same mean, G second-order stochastically dominates F if $\int \psi(x)dF(x) \geq \int \psi(x)dG(x)$ for every increasing convex function $\psi : \mathbb{R}_+ \to \mathbb{R}$. It then follows $E_k\Pi(\phi) \geq E_{k+1}\Pi(\phi)$, by convexity of $\Pi(\cdot)$, if the exogenous separation rates ξ_k and ξ_{k+1} are sufficiently small. This argument, for $k = 1, 2$, ensures that $E_1\Pi(\phi) \geq E_2\Pi(\phi) \geq E_3\Pi(\phi)$. Observe now that $E_1\Pi \geq \overline{\Pi}$ implies $\phi_1^* \leq \phi_{0,1}$. The condition $\phi_{0,1} < \phi_{1,2}$ in turn implies $y_1(\phi_{0,1}) > y_2(\phi_{0,1})$. Note that $y_2(\overline{\theta}) > y_1(\overline{\theta})$ and $y_1(\underline{\theta}) > y_2(\underline{\theta})$ imply that the difference $y_1(\phi) - y_2(\phi)$ is strictly decreasing. With $\phi_1^* < \phi_{0,1}$, from $y_1(\phi_{0,1}) > y_2(\phi_{0,1})$ it follows $y_1(\phi_1^*) > y_2(\phi_1^*)$. Then, condition (4) holds true. Recall that $\overline{\phi}$ is the belief value for which the firm is indifferent, in the static case, between tasks 1 and 2, if task 1 is perfectly informative about ability, while task 2 is completely uninformative. Then, $\overline{\phi}$ can be computed as

$$\overline{\phi} \equiv \frac{y_1(\underline{\theta}) - y_2(\underline{\theta}) + \frac{\delta(1-\xi_2)\Pi}{1-\delta(1-\xi_2)}}{\frac{y_2(\overline{\theta})-y_2(\underline{\theta})}{1-\delta(1-\xi_2)} - y_1(\overline{\theta}) + y_1(\underline{\theta}) - \frac{\delta(1-\xi_1)(y_3(\overline{\theta})-U-\Pi)}{1-\delta(1-\xi_3)}}.$$

Note that $\overline{\phi} \in (0,1)$, if ξ_k, $k = 1, 2, 3$, is sufficiently small. Also, $\phi_2^* < \overline{\phi}$. For $y_3(\overline{\theta}) > y_2(\overline{\theta})$ and $y_2(\underline{\theta}) > y_3(\underline{\theta})$, the difference $y_2(\phi) - y_3(\phi)$ is strictly decreasing. Moreover, $\phi_2^* < \overline{\phi}$ implies that, if $y_2(\overline{\phi}) - y_3(\overline{\phi}) > 0$, then $y_2(\phi_2^*) - y_3(\phi_2^*) > 0$. Observe finally that $y_2(\overline{\phi}) - y_3(\overline{\phi}) > 0$ is equivalent to $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$. The fact that $\phi_1^* < \phi_{0,1}$, $\phi_2^* > \phi_{1,2}$ and $\phi_3^* > \phi_{2,3}$ is consequence that at ϕ_1^*, ϕ_2^* and ϕ_3^* the firm's value function is kinked, so that, respectively, $E_1\Pi(\cdot) > \overline{\Pi}$, $E_1\Pi(\cdot) > E_2\Pi(\cdot)$ and $E_2\Pi(\cdot) > E_3\Pi(\cdot)$.

We will now show that the cut-off belief values ϕ_2^* and ϕ_3^* are uniquely determined. Define, analogously to the proof of Proposition 1, $Tf(\phi) = \max\{\overline{\Pi}, T_1f(\phi), T_2f(\phi), T_3f(\phi)\}$, where, for

$k = 1, 2, 3,$

$$T_k f(\phi) = p_k(\phi)[\overline{y}_k + \delta(1-\xi_k)f(\phi_{kh}(\phi))] + (1-p_k(\phi))[\underline{y}_k + \delta(1-\xi_k)f(\phi_{kl}(\phi))] + \delta\xi_k\overline{\Pi}.$$

Suppose that $T_2f - T_1f$ and $T_3f - T_2f$ are increasing in ϕ. From $E_1f \geq E_2f \geq E_3f$, and $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, it follows that, when $T_1f < T_0f$, also $T_2f < T_0f$ and $T_3f < T_0f$. The fact that $y_1(\overline{\theta}) > \Pi > y_1(\underline{\theta})$ implies that there exists a $\hat{\phi}_1 \in (0,1)$ such that $T_1f = \overline{\Pi}$. Since $E_1f \geq E_2f \geq E_3f$, and $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, it follows that, at $\hat{\phi}_1$, $T_1f \geq T_2f$ implies $T_1f \geq T_3f$. Therefore, $Tf(\hat{\phi}_1) = T_1f(\hat{\phi}_1)$. From the fact that $T_2f - T_1f$ is increasing, and that $y_2(\overline{\theta}) > y_1(\overline{\theta})$, it also follows $Tf = T_1f$ for $\phi \in [\hat{\phi}_1, \hat{\phi}_2)$, for some $\hat{\phi}_2 < 1$. Then, for $\phi \geq \hat{\phi}_2$, $Tf = T_2f$ or $Tf = T_3f$. Now, $y_2(\overline{\phi}) > y_3(\overline{\phi})$, with $\overline{\phi} > \hat{\phi}_2$, and $E_2f \geq E_3f$ yield that, at $\hat{\phi}_2$, $Tf = T_2f$. Since $T_3f - T_2f$ is increasing and $y_3(\overline{\theta}) > y_2(\overline{\theta})$, it follows that there also exists $\hat{\phi}_3$ such that $Tf = T_2f$ for $\phi \in [\hat{\phi}_2, \hat{\phi}_3)$ and $Tf = T_3f$ for $\phi \geq \hat{\phi}_3$. Then, Tf can be rewritten as

$$Tf = \begin{cases} T_3f, & \text{if } \phi \in [\hat{\phi}_3, 1] \\ T_2f, & \text{if } \phi \in [\hat{\phi}_2, \hat{\phi}_3) \\ T_1f, & \text{if } \phi \in [\hat{\phi}_1, \hat{\phi}_2) \\ T_0f, & \text{if } \phi \in [0, \hat{\phi}_1). \end{cases}$$

Consider now the differences $T_3(Tf) - T_2(Tf)$ and $T_2(Tf) - T_1(Tf)$. Notice that they can be rewritten, respectively, as[37]

$$T_3(Tf) - T_2(Tf) = [T_3(Tf) - T_3(T_2f)] + [T_2(T_3f) - T_2(Tf)] + [T_3(T_2f) - T_2(T_3f)] \tag{6}$$
$$T_2(Tf) - T_1(Tf) = [T_2(Tf) - T_2(T_1f)] + [T_1(T_2f) - T_1(Tf)] + [T_2(T_1f) - T_1(T_2f)]. \tag{7}$$

Suppose that, for any real-valued function f on $[0,1]$, $T_3f - T_2f$ increasing over $[\hat{\phi}_2, 1]$ implies that $T_3(Tf) - T_2(Tf)$ is strictly increasing for the same values of ϕ and, similarly, that $T_2f - T_1f$ increasing over $[\hat{\phi}_1, \hat{\phi}_3]$ implies that $T_2(Tf) - T_1(Tf)$ is strictly increasing for the same values of ϕ. Then, since $\Pi(\cdot)$ is the unique fixed point of T, $\Pi_3(\phi) - \Pi_2(\phi)$ and $\Pi_2(\phi) - \Pi_1(\phi)$ must be strictly increasing over those belief ranges. To prove that $T_3f - T_2f$ increasing implies that $T_3(Tf) - T_2(Tf)$ is strictly increasing and, similarly, that $T_2f - T_1f$ increasing implies that $T_2(Tf) - T_1(Tf)$ is strictly increasing, it is enough to show that each term in the right-hand side of (6) and (7) is increasing, and at least one strictly increasing, over the desired belief range. Notice that (6) can then be rewritten as

$$\begin{aligned} 1_{\{\phi\in[\hat{\phi}_2,1]\}}T_3(Tf) \;-\; T_2(Tf) &= 1_{\{\phi\in[\hat{\phi}_3,1]\}}\{T_3(T_3f) - T_3(T_2f)\} + 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{T_2(T_3f) - T_2(T_2f)\} \\ &\quad +1_{\{\phi\in[\hat{\phi}_2,1]\}}(1-\delta)^2(y_3(\phi) - y_2(\phi)) \\ &= 1_{\{\phi\in[\hat{\phi}_3,1]\}}\{\delta E_3[T_3f - T_2f]\} + 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{\delta E_2[T_3f - T_2f]\} \\ &\quad +1_{\{\phi\in[\hat{\phi}_2,1]\}}(1-\delta)^2(y_3(\phi) - y_2(\phi)). \end{aligned} \tag{8}$$

The first equality follows from the definition of Tf and the exchangeability of the output signal, which implies $E_3E_2f = E_2E_3f$. With $T_3f - T_2f$ increasing, and the fact that $\phi'' > \phi'$ implies that

[37] The decomposition is analogous to the one used in the proof of Lemma 1 in Kakigi [1983].

the distribution of the updated posterior at any job, conditional on ϕ'', first-order stochastically dominates the one conditional on ϕ', it follows that the first two terms in (8) are increasing. Since the difference $y_3(\phi) - y_2(\phi)$ is strictly increasing, it follows $T_3(Tf) - T_2(T_2 f)$ is strictly increasing as well for $\phi \in [\phi_2^*, 1]$. Similarly, when $\phi \in [\hat{\phi}_1, \hat{\phi}_3]$, (7) can be rewritten as

$$\begin{aligned} 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}} T_2(Tf) \; - \; T_1(Tf) &= 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{T_2(T_2 f) - T_2(T_1 f)\} + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_2)\}}\{T_1(T_2 f) - T_1(T_1 f)\} \\ &\quad + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}}(1-\delta)^2(y_2(\phi) - y_1(\phi)) \\ &= 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{\delta E_2[T_2 f - T_1 f] + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_2)\}}\delta E_1[T_2 f - T_1 f] \\ &\quad + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}}(1-\delta)^2(y_2(\phi) - y_1(\phi)). \end{aligned}$$

As before, the first equality follows from the definition of Tf and the fact that $E_2 E_1 f = E_1 E_2 f$. With $T_2 f - T_1 f$ increasing, it follows that $E_k[T_2 f - T_1 f]$ is also increasing, for $k = 1, 2$. Since $y_2(\phi) - y_1(\phi)$ is strictly increasing, the difference $T_2(Tf) - T_1(T_2 f)$ is strictly increasing for $\phi \in [\phi_1^*, \phi_3^*]$. This completes the proof of the claim. □

Proof of Proposition 3: The argument can be adapted from the equivalent proof in the companion paper (Pastorino [2005]). □

Appendix B

B.1 Sample Without Level or Performance Rating Missing

Table B1. Fraction of High Ratings (Original Sample (OS), Estimation Sample (ES))

Rating	OS (Total)	OS	ES (Total)	ES
1	13,994	0.306	763	0.398
2	22,756	0.498	954	0.497
3	8,455	0.185	197	0.103
4	438	0.010	2	0.001
5	30	0.001	3	0.002
Total	45,673	1.000	1,919	1.000

Table B2. Distribution of Employees Across Levels
(12 Years of Education at Entry - 43 Employees)

Years (Since Entry)	Level 1	Level 2	Level 3	Exit	Total
0	1.000	0.000	0.000	0.000	1.000
1	0.698	0.116	0.000	0.186	1.000
2	0.465	0.093	0.070	0.372	1.000
3	0.233	0.140	0.093	0.535	1.000
4	0.047	0.140	0.116	0.698	1.000
5	0.000	0.070	0.047	0.884	1.000
6	0.000	0.023	0.000	0.977	1.000

Table B3. Hazard Rates of Exit and Promotion by Level
(12 Years of Education at Entry - 43 Employees)

Years (at Level)	Level 1 to Exit	Level 1 to Level 2	Level 2 to Exit	Level 2 to Level 3	Level 3 to Exit
1	0.186	0.116	0.200	0.400	0.000
2	0.233	0.067	0.000	0.500	0.000
3	0.350	0.150	0.000	0.000	0.667
4	0.500	0.300	0.000	0.000	1.000
5	1.000	0.000	1.000	0.000	-
6	-	-	-	-	-

Table B4. Distribution of Employees Across Levels
(13-15 Years of Education at Entry - 70 Employees)

Years (Since Entry)	Level 1	Level 2	Level 3	Exit	Total
0	1.000	0.000	0.000	0.000	1.000
1	0.700	0.100	0.000	0.200	1.000
2	0.300	0.143	0.014	0.543	1.000
3	0.171	0.100	0.043	0.686	1.000
4	0.114	0.057	0.071	0.757	1.000
5	0.043	0.057	0.014	0.886	1.000
6	0.014	0.043	0.014	0.929	1.000
7	0.014	0.029	0.029	0.929	1.000
8	0.000	0.000	0.014	0.986	1.000

Table B5. Hazard Rates of Exit and Promotion by Level (13-15 Years of Education at Entry - 70 Employees)

Years (at Level)	Level 1 to Exit	Level 1 to Level 2	Level 2 to Exit	Level 2 to Level 3	Level 3 to Exit
1	0.200	0.100	0.571	0.000	0.000
2	0.408	0.143	0.000	0.667	0.000
3	0.333	0.095	0.000	0.000	0.000
4	0.250	0.083	0.000	0.000	0.000
5	0.500	0.125	1.000	0.000	0.000
6	0.667	0.000	-	-	1.000
7	0.000	0.000	-	-	-
8	1.000	0.000	-	-	-

Table B6. Distribution of Employees Across Levels (All Education Groups - 698 Employees)

Years (Since Entry)	Level 1	Level 2	Level 3	Exit	Total
0	1.000	0.000	0.000	0.000	1.000
1	0.431	0.264	0.000	0.305	1.000
2	0.168	0.236	0.036	0.560	1.000
3	0.070	0.136	0.070	0.723	1.000
4	0.033	0.076	0.066	0.825	1.000
5	0.013	0.030	0.039	0.918	1.000
6	0.004	0.016	0.026	0.954	1.000
7	0.001	0.009	0.016	0.974	1.000
8	0.000	0.001	0.007	0.991	1.000
9	0.000	0.000	0.001	0.999	1.000

Table B7. Hazard Rates of Exit and Promotion by Level (All Education Groups - 698 Employees)

Years (at Level)	Level 1 to Exit	Level 1 to Level 2	Level 2 to Exit	Level 2 to Level 3	Level 3 to Exit
1	0.305	0.264	0.402	0.125	0.360
2	0.346	0.259	0.437	0.333	0.200
3	0.359	0.222	0.500	0.150	0.417
4	0.367	0.163	0.286	0.286	0.571
5	0.522	0.087	1.000	0.000	0.333
6	0.556	0.111	-	-	0.500
7	0.667	0.000	-	-	0.000
8	1.000	0.000	-	-	-
9	-	-	-	-	-

B.2 Descriptive Statistics for the Estimation Sample

When restricting attention to individual histories for which no level or performance rating is missing, the largest reduction in the number of observations is due to missing performance ratings. Of the original 21,905 employee-years (2,714 individuals) entering the firm at Level 1 between 1970 and 1979, 20,212 employee-years (2,557 individuals, of which 1,552 have 16 or more years of education at entry) have no level information missing.

Only 1,921 (699 individuals, of which 502 have 16 or more years of education at entry) have no level or performance rating missing. Of these 699, only individual who was assigned to Level 3 in period 2. Of the remaining 698, 43 employees have 12 years of education at entry, 70 individuals have 13 to 15 years and 502 have 16 or more years (for 83 individuals education information is missing at the time of entry).

As for the distribution of employees across levels (Table B8), restricting attention to the individuals with at least 16 years of education at entry, the patterns are very similar across the groups of 1,552 individuals, for which no level information is missing, and the group of 502 individuals, for which no level or performance rating is missing. Still, exit in this latter group is much less pronounced, given that individuals with missing ratings tend to have longer tenures, as reflected in the hazard rate of separation reported in Table B9.

Table B8. Distribution of Employees Across Levels (16 or More Years of Education at Entry - 502 Employees)

Years (Since Entry)	Level 1	Level 2	Level 3	Exit	Total
0	1.000	0.000	0.000	0.000	1.000
1	0.347	0.327	0.000	0.327	1.000
2	0.100	0.277	0.042	0.582	1.000
3	0.034	0.141	0.084	0.741	1.000
4	0.014	0.074	0.070	0.843	1.000
5	0.008	0.026	0.048	0.918	1.000
6	0.002	0.012	0.034	0.952	1.000
7	0.000	0.006	0.018	0.976	1.000
8	0.000	0.002	0.008	0.990	1.000
9	0.000	0.000	0.002	0.998	1.000

Table B9. Hazard Rates of Exit and Promotion by Level (16 or More Years of Education at Entry - 502 Employees)

Years at Level	Level 1 to Exit	Level 1 to Level 2	Level 2 to Exit	Level 2 to Level 3	Level 3 to Exit
1	0.327	0.327	0.390	0.128	0.429
2	0.368	0.345	0.468	0.329	0.273
3	0.280	0.380	0.563	0.188	0.375
4	0.353	0.235	0.250	0.500	0.600
5	0.286	0.143	1.000	0.000	0.500
6	0.500	0.250	-	-	0.000
7	1.000	0.000	-	-	0.000
8	-	-	-	-	-
9	-	-	-	-	-

References

[1] Baker, G., M. Gibbs and B. Holmström (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data", *Quarterly Journal of Economics* 109(4), 881-919.

[2] Baker, G., M. Gibbs and B. Holmström (1994b): "The Wage Policy of a Firm", *Quarterly Journal of Economics* 109(4), 921-955.

[3] Baker, G., M. Gibbs and B. Holmström (1995): "Internal Labor Markets: Too Many Theories, Too Few Facts", *American Economic Review* 85(2), 255-259.

[4] Banks, J.S., and R.K. Sundaram (1992a): "Denumerable-Armed Bandits", *Econometrica* 60(5), 1071-1096.

[5] Bergemann, D., and J. Välimäki (1996): "Learning and Strategic Pricing", *Econometrica* 64(5), 1125-1149.

[6] Chade, H., and E.E. Schlee (2002): "Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information", *Journal of Economic Theory* 107(2), 421-452.

[7] Chiappori, A.P. (2003): "Testing Contract Theory: a Survey of Some Recent Work", in Advances in Economics and Econometrics - Theory and Applications, Eighth World Congress, M. Dewatripont, L. Hansen and P. Turnovsky, ed., Econometric Society Monographs, Cambridge University Press, Cambridge, 2003, 115-149.

[8] Chiappori, A.P., B. Salanié and J. Valentin (1999): "Early Starters Versus Late Beginners", *Journal of Political Economy* 107(4), 731-760.

[9] De Varo, J., and M. Waldman (2004): "The Signaling Role of Promotions: Further Theory and Empirical Evidence", mimeo, Cornell University.

[10] Eeckstein, Z., and K. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities", *Econometrica* 67(6), 1295-1339.

[11] Fairburn, J.A., and J.M. Malcomson (2001): "Performance Promotion and the Peter Principle", *Review of Economic Studies* 68(1), 45-66.

[12] Gibbons, R., and M. Waldman (1999a): "Careers in Organizations. Theory and Evidence", in *Handbook of Labor Economics*, Volume III, O. Ashenfelter and D. Card Eds., North-Holland, Amsterdam, the Netherlands, 2373-2437.

[13] Gibbons, R., and M. Waldman (1999b): "A Theory of Wage and Promotion Dynamics Inside Firms", *Quarterly Journal of Economics* 114(4), 1321-1358.

[14] Gibbs, M. (1995): “Incentive Compensation in a Corporate Hierarchy”, *Journal of Accounting and Economics* 19(2-3), 247-277.

[15] Holmström, B., and P. Milgrom (1991): “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”, *Journal of Law, Economics and Organization* 7, 24-52.

[16] Holmström, B., and J. Tirole (1989): “The Theory of the Firm”, in *Handbook of Industrial Organization*, Volume I, Chapter 2, R. Schmalensee and R.D. Willig Eds., North-Holland, Amsterdam, the Netherlands, 61-133.

[17] Jovanovic, B. (1979a): “Job Matching and the Theory of Turnover”, *Journal of Political Economy* 87(5), 972-990.

[18] Jovanovic, B. (1979b): “Firm-specific Capital and Turnover”, *Journal of Political Economy* 87(6), 1246-1260.

[19] Kakigi, R. (1983): “A Note on Discounted Future Two-Armed Bandits”, *Annals of Statistics* 11(2), 707-711.

[20] Keane, M.P., and R.M. Sauer (2003): “A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables”, mimeo, Yale University.

[21] Keane, M.P., and K. Wolpin (2001): “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment”, *International Economic Review* 45(2), 601-641.

[22] Keane, M.P., and K. Wolpin (1997): “The Career Decisions of Young Men”, *Journal of Political Economy* 105(3), 473-522.

[23] Kwon, I. (2004): “Threat of Dismissal: Incentives or Sorting?”, University of Michigan, mimeo.

[24] Nagypal, E. (2003): “Learning-by-Doing Versus Learning About Match Quality: Can We Tell Them Apart?”, Northwestern University, mimeo.

[25] Pastorino, E. (2005): “Career Dynamics in Firms Under Uncertainty and Learning”, University of Iowa, mimeo.

[26] Prendergast, C. (1999): “The Provision of Incentives in Firms”, *Journal of Economic Literature* 37(1), 7-63.

[27] Prendergast, C. (1993): “The Role of Promotion in Inducing Specific Human Capital Acquisition”, *Quarterly Journal of Economics* 108(2), 523-534.

[28] Waldman, M. (1984b): “Job Assingment, Signalling, and Efficiency”, *Rand Journal of Economics* 15(2), 255-267.

The Market for Intellectual Property: Evidence from the Transfer of Patents

Carlos J. Serrano*

University of Toronto

July, 2005
Preliminary

Abstract

This paper studies the economics of intellectual property transfer using new data on the transfers of patents. We first present evidence about how the transfer and the expiration of patents relate to age, patent citations received, whether the patent has been previously traded or not, and the timing of the last trade. Next, we develop a model of the transfer and renewal of patents including costly technology transfer and gains from trade. We structurally estimate the parameters of the model and use the estimates to calculate the value of a patent, and in particular, to quantify what are the gains from trade in the market for patents. The estimation uses information on patent citations received, which relate to the whole spectrum of patents returns . Thus, the identification of our estimates does not uniquely rely on patents with value located at left tail of the distribution of patent returns.

*I am especially grateful to Thomas Holmes and Sam Kortum for their advice and patience. I would also like to thank Chris Laincz, Nicolas Figueroa, Suqin Ge, Valeriu Omer, Marcel Boyer, Zvi Eckstein, Mike Golosov, Antonio Merlo, Katherine Lande, Jim Schmitz and participants at several workshops and universities for their comments. I also acknowledge financial support from the Bank of Spain Graduate Fellowship and the Connaught Start-Up Award from the University of Toronto. All errors are mine.

1 Introduction

NOWADAYS INTANGIBLES ASSETS (i.e., patents, trademarks, copyrights, etc) have become an increasing large share of develop economies. For instance, in the early 1980's they represented 38% of the portfolios of U.S. firms. In the mid 1990's this share rose to 70% (WIPO [27])). However, we know little about measuring them, whether there are significant spillovers in their use, how and why they get transferred between firms, and what are the efficiency gains from the transfer of patents. In this paper, we focus in the dissemination of patents across firm's boundaries. In particular, we estimate what are the gains from trade in the market for patents. We find that the market for patents represents 14.21% of the estimated average value of a patent.

Our work is distinct to previous literature in that it makes use of data on the transfer of patents. The transfer of patents is a significant event of the life cycle of a patent. We find that nearly 20% of all U.S. patents issued to small innovators (i.e., firms that were issued no more than 5 patents in a given year) are traded at least once over their life cycle. Thus the event of a transfer is a common aspect of the data that concerns the life cycle of patents.

We have derived five key facts using this data. First, the number of patents traded as a percentage of active patents (i.e., the transfer rate) varies over the life cycle. It monotonically decreases with the exceptions of the renewal dates. In particular, immediately after renewal, transfer rates discontinuously increase. Second, the number of expired patents at a renewal date as a percentage of all active patents (i.e., expiration rate) monotonicall increases as a function of age. Third, patents with higher number of total citations received by a given renewal date are less likely to be allowed to expire[1]. Fourth, patents with higher number of total citations received are more likely to be traded. Fifth, patents that have been previously traded, and in particular the recently traded, are more likely to be retraded and less likely to expire. These patterns are robust to different patent cohorts and patent classes (i.e., industries) as shown in Serrano [22].

There is an extensive empirical literature investigating patent data (BLS [6], Griliches [7], [8], Pakes and Griliches [20], and Hall, Griliches and Hausman [11], Jaffe, Henderson and Tratjenberg [12] and Tratjenberg, Henderson and Jaffe [25], Hall, Jaffe and Tratjenberg [9], Hall, Jaffe and Tratjenberg [10] and others). Our work is different in that it uses data on transfers of patents. The U.S. patent office registers transfers of patents in the same

[1]Each patent when granted lists references to previous patents, that is citations made. Instead, citations received by a patent is the number of times that this patent has been referenced by other patents. Previous empirical studies on patents have found that citations received by a patent is a measure of the economics value of a patent.

way that counties register the transfer of houses. As we show here, the market for trade in patents is large. In our study, we make use of all the records of titles transferred and link this information to the basic patent data (e.g., patent's grant date, renewals, citations received, etc.) that others have used.

The dataset we have constructed is a panel with the histories of trades and renewal decisions for patents granted since the early 1980's. In addition, it contains characteristics such as citations received, industry to which the patented technology belongs, size of the firm that was the owner at the grant date, and other relevant information[2].

There is also extensive work in the theoretical literature on patents. The starting point for my theory is Schankerman and Pakes [21] and Pakes [19]. They examine the problem of a patent owner deciding in each period whether or not to pay the renewal fee and thereby extend the life of a patent. The contribution of my theory is to introduce into the model, in each each period, an alternative potential owner who may have a greater valuation for the patent than the owner at the beginning of a given period. To transfer a patent to a new owner involves a resource cost, a transaction cost. In summary, whereas Schankerman and Pakes' model has one margin, should the patent owner pay the fee for renewing the patent, my model has a second margin, should the cost of technology transfer be paid to reallocate the patent to an alternative owner.

The intuition for the model is simple. Initially, patents are granted to a fraction of firms. The rest of the firms are potential buyers. At every period, patents are traded because some firms are more productive than others in the use of a given patent. However, any gain from trade in transferring a patent to the potential buyer must be weighed against the resource cost of technology transfer.

In the model, there are two mechanisms. First, the cost of technology transfer creates a selection effect so that better patents are more likely to be traded. This selection effect accounts for the discontinuous increase in the transfer rate after renewal, and also for the reason that traded patents are more likely to be retraded and less likely to expire. Second, there is a horizon effect that explains why transfer rates decrease as patents get closer to their expiration date. A shorter horizon implies less time to amortize the cost of technology transfer.

The parameters of the model are estimated using the simulated general method of moments. The main departure of our estimation strategy with respect to the previous literature is the use of data on the transfer of patents, and of patent citations received. Both type of data complement the information about renewals already used in the previous

[2]The size of a firm is defined as the number of patents that were granted to the firm who was the owner of the patent at the time that the patent was granted.

literature. The advantage of using patent citations is that it permit us to estimate the parameters of the distribution of patent returns using information about the whole spectrum of patents. Thus, the identification of our estimates does not uniquely rely on patents with value located at left tail of the distribution of patent returns such as the estimates of Schankerman and Pakes [21], Pakes [19] and Lanjouw [15].

In detail, we estimate the parameters of the model in order to fit the proportion of active patents that are traded conditional on having been previously traded or not previously traded, the proportion of active patents that are allowed to expire conditional on having been previously traded or not previously traded, the mean of the number of citations received at ages 2 to 17 for previously traded and not previously traded patents, and the mean number of citations received of active patents at age 1. We find that the average value of a patent is $US (2003) 57,910, however patent returns are very skewed. For instance, we find that the bottom 50% of all granted patents only account for 7.6% of the cumulative value of all patents, and the 90% accounts for 44.9% of the total value. We also find a large difference between the present value of a patent at age 1 depending on whether that patent will be eventually traded or not. For example, the average value of a traded and non-traded patent is, respectively $130,155 and $42,426. Finally, we find that the fixed per patent cost of technology transfer is estimated to be substantial, $26,298.

The previous literature on markets for intellectual property can be summarized in three groups. One strand aims to demonstrate evidence of the existence of this market. The method generally used has been the analysis of industry case studies, such as the works collected in Arora, Fosfuri and Gambardella [2]. In addition, a sequence of papers by Lamoreaux and Sokoloff [17], [18] provide an account of organized markets for technology in the late 19^{th} and early 20^{th} century, prior to the growth of in house R&D laboratories by large firms. The second strand of the literature has suggested the existence of potential gains from specialization and diffusion of technology (Arrow [3], Arora, Fosfuri and Gambardella [2]). The third strand has analyzed the limitations of the market, such as the appropiation problems in the transfer of knowledge (Arrow [4], Teece [24] and Williamson [26]) and the cost of transfer of technology (Teece [23]).

Finally, our work opens new avenues of research. First, to study the sources of innovation and to characterize who are the buyers and sellers of technology. In particular, to trace the flow of technology transfer and to analyze whether small firms specialize in the creation of innovations that eventually are sold to their larger counterparts. Second, to examine to what extent a higher level of patent protection has facilitated specialization and, consequently trade in patents. Lastly, to evaluate the use of taxation on intellectual property transfer to promote innovation. These questions have not been previously empirically addressed due to a lack of data on how patents are traded.

This paper is organized as follows. Section 2 explains the data and presents the stylized facts. Section 3 develops the model. Section 4 solves the model and links the stylized facts with the results of the model. Section 5 present the estimation strategy and discuss the identification of the estimates. Section 6 shows what are the estimation results, in particular it quantifies what are the gains from trade in the market for patents. Section 7 concludes the paper. Finally, all proofs and a data summary are included in the Appendix.

2 Data

A patent for an invention is the grant of a property right to the inventor in order to exclude others from making, using, or selling the invention. The life cycle of a patent begins at the grant date.[3] By the end of year 4, 8 and 12 upon the grant date renewal fees are due. If they are not paid, then the patent expires.[4] Such renewal events have been studied for patents granted in European countries in an extensive and important literature (Schankerman and Pakes [21], Pakes [19], Lanjouw [15] and others).

Another event that can happen in the life cycle of a patent is what the U.S. patent office calls "reassignments", and what we will call a "transfer" or "trade". In principle, the event can happen many times during the life of a single patent. The U.S. patent office maintains a registry of these events. We have obtained these records for all transfers that occurred from 1981 to 2002, of which there were 1,041,083. The records have information about patent numbers, making it possible to merge the patent level data on renewals and citations that has been used in the previous literature. The details of the procedures we used to deal with the transfer data are explained in Serrano [22].

A particular issue we treat in detail in Serrano [22] is that some of the transfers recorded with the patent office are administrative events, like a name change, as opposed to a true economic transfer between two distinct parties. Fortunately, for each transaction there is a data field that records the "brief", which is the nature of the trade. We separate out traded patents where the reason is a name change, a security interest, a correction, etc. The remaining accounts for 508,756 patents, 49% of all traded patents.

A second issue is that in cases where there is a merger between two large companies, patents are traded in large blocks. When Burroughs Corporation merged with Sperry

[3]The term of new patents applied for prior to 1995 was 17 years from their grant date. This term was subsequently modified to 20 years from the date in which the patent application was filed.

[4]The USPTO states that the renewal fee by the end of years 4, 8 and 12 since the grant date of the patent are respectively, US $910, $2,090 and $3,220 as of October 7, 2003. The USPTO began charging renewal fees in 1984 on patents applied for after December 12, 1980.

Corporation to create Unisys Corporation in September 1986, this event appears in my data as transactions totalling 2261 patents (the largest single transaction includes 1702 patents). Our theoretical analysis will focus on decision making at the *patent* level. There are costs and benefits of transferring a particular patent. Obviously, in a wholesale trade such as Burroughs merging with Sperry, the decision making is not at the level of a single patent. To parallel our focus in the theory, in our empirical analysis we focus on *small innovators*. In doing so, the economic forces that we highlight will be more salient than in transactions involving the likes of Burroughs or Sperry.[5] In addition, small innovators are interesting in their own right, given the importance they play in the innovation process (Arrow [5], Acs and Audretsch [1]). Indeed, we operationalize this focus on small innovators by restricting attention to patents granted to firms with no more than 5 patents granted to them that year.

The dataset we have compiled is a panel of patents detailing their histories of trade and renewal decisions. The panel contains patents that were applied for after December 12, 1980 and issued since January 1, 1983 to U.S. or foreign businesses.[6] In addition, it has characteristics such as citations received, industry to which the patented technology belongs, size of the firm that was the owner at the grant date, and other relevant information.

The panel includes 453,683 patents granted to small innovators. This sample contains about a third of all granted patents to U.S. or foreign businesses.

The next section presents the key facts of the transfer of patents.

2.1 Facts

This section presents the basic facts that describe the underlying quality of traded patents, how the transfer rate varies over the life cycle of a patent, and the effects of a transfer on the renewal and trading decision.[7] The key facts are the following:

1. The transfer rate[8] monotonically decreases following the first year after a patent has

[5] A companion paper, "Measuring the Transfer of Patents" shows that patents granted to large corporations are more likely to be traded for other reasons than the technology that they represent. For instance, they can be recorded as a result of large acquisitions pursued to increase the buyer's market share in a particular product, etc.

[6] Patents applied for after December 12, 1980 are subject to renewal fees. To create a comphensive sample we consider January 1, 1983 as the starting grant date of the patents contained in the panel. Finally, issued to U.S. or foreign business means that at the date the patent was granted, the owner was a U.S. or foreign business.

[7] For explanation, the age of a patent is defined as follows. Its age when it is traded is the number of years between the trade date and the grant date. In particular, if a patent was traded during its second year of life (e.g., 17 or 22 months since being issued), I consider that the patent was traded at age 2.

[8] Transfer rate at a given patent age is defined as the proportion of patents that are traded conditional on having survived up to that period.

been granted, with the exceptions of the renewal dates. Immediately after renewal, the transfer rate discontinuously increases. Moreover, the transfer rate increases during the application period of a patent.

2. The expiration rate[9] monotonically increases as a function of the renewal dates.

3. Patents with higher number of total citations received by a given renewal date are less likely to be allowed to expire.

4. Patents with higher number of total citations received by a given age are more likely to be traded.

5. Among patents of the same age, those that have been previously traded, in particular those recently traded, are more likely to be retraded and less likely to expire.

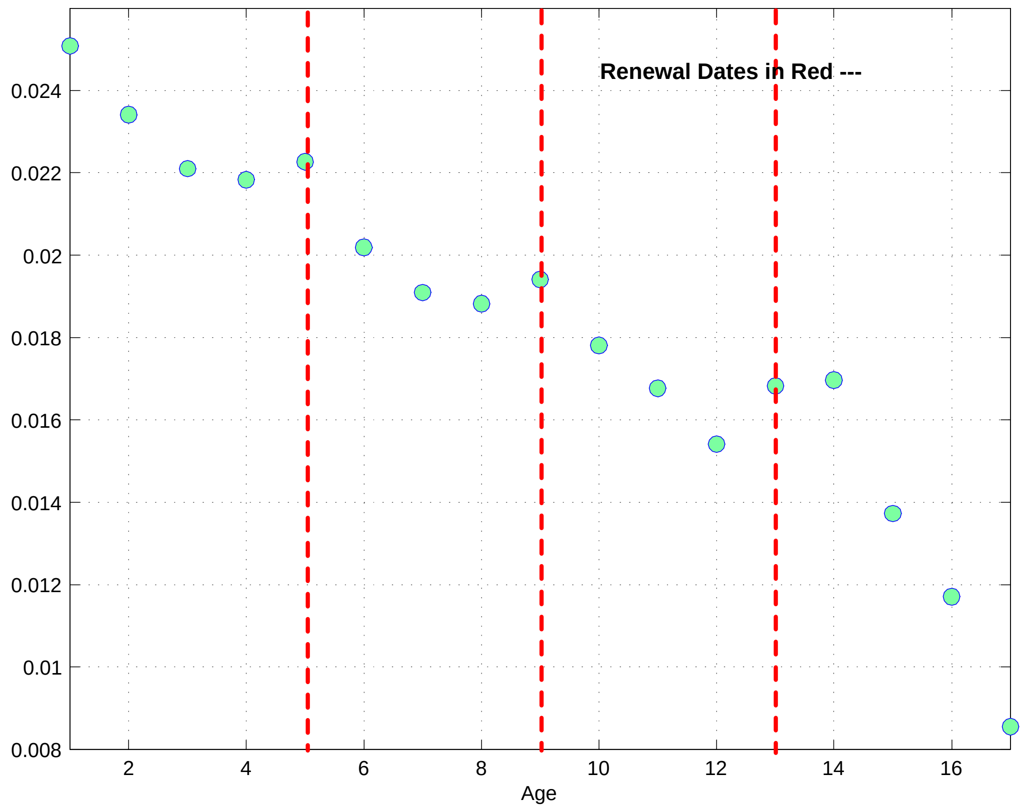
Fact 1: The transfer rate monotonically decreases from the grant date until the end of a patent's life except at the renewal dates. Immediately after renewal, this rate increases abruptly. For instance, the proportion of traded patents drops from 2.51 to 2.18% respectively, from age 1 to age 4. Then in age 5 it increases to 2.23 and in age 6 drops again to 2.02%. This evidence is consistent for all three renewal dates in which renewal fees are due according to the U.S. patent system, and controlling by patent classes.

Fact 2: The expiration rate monotonically increases as a function of the renewal dates. In Table 1 we show the expiration rate at each renewal date. For instance, among patents of age 5, 18.08% are allowed to expire by that date. Four years after, at age 9, 28.39% of the remaining patents are allowed to expire. The last renewal fee, due by the beginning of age 13, implies that 32.02% of all active patents become not active. The finite horizon of a patent life together with the "depreciation" of their per period revenue due to imatation of the arrival of superior technologies might the forces that are behind the increasing pattern of the expiration rate.

Fact 3: Patents with higher number of total citations received by a given age are less likely to be allowed to expire. In Table 1A, we present the predicted probabilities of calculated with the estimates of a logit model. We have regressed the renewal decision on renewal date dummies, patent class dummies, and total citations received. We find that among patents of the same age, as the number of total citations increases, the expiration

[9]Expiration rate at a given patent age is defined as the proportion of patents that are expired conditional on having survived up to that period.

Figure 1: The Proportion of Traded Patents Conditional on Renewal



rate decreases. This feature confirms the robustness of using dynamic citations as a measure of the economic value of patents. Furthermore, according to the regression estimates, adding an extra citation received to a patent decreases the log of the probability of being expired by 0.043 units[10]. The following table presents the predicted expiration rate conditional on the number of total citations received by each renewal date.

Fact 4: Patents with a higher number of total citations citations received are more likely to be traded. Logit analysis shows that an extra citation received by a patent increases the log of the probability of being traded by 0.011 units. For instance, in Table 1B we show the predicted probability to be traded of active patents conditional on the total citations received. We find that patents with 1 total citations received at age 8 have an estimated probability of being traded of 0.0177 at age 8, this probability jumps to 0.038 if total citations received are 60, and it spikes to 0.063 if total citations received are 100.

Fact 5: This fact focuses on the effects of the trading decision on the retrading and the future renewal decision of the patent. In Table 2, we show a combination of results

[10]Details about the logit analysis can be found in a companion paper: "Measuring the Transfer of Patents," manuscript, at http://www.econ.umn.edu/~carles/research.htm

Table 1: Transfer Rate and Expiration Rate Conditional on Citations Received

A. Expiration Rate as a Function of Total Citations Received

		Total Citations Received				
Age	Unconditional	0	1	10	20	30
5	.1808	.1990	.1920	.1369	.0920	.0607
9	.2839	.3342	.3243	.2427	.1699	.1155
13	.3202	.3981	.3874	.2969	.2124	.1468

B. Transfer Rate as a Function of Total Citations Received

		Total Citations Received				
Age	Unconditional	0	1	10	20	30
1	.0251	.0250	.0254	.0285	.0324	.0369
8	.0182	.0175	.0177	.0199	.0226	.0258
17	.0085	.0073	.0074	.0084	.0095	.0110

that explain how the trading decision and its timing show the difference between previously traded and non-traded patents.

In particular, columns labeled as "Not Previously Traded" and "Previously Traded - Any year" of Table 2 show that among patents of the same age, traded patents are more likely to be retraded and less likely to expire. We see that previously traded patents are twice more likely to be retraded. With respect to the renewal decision, traded patents are about 5 percentage points less likely to expire at each of the renewal dates.

In addition, columns labeled as "1 year" and "4 years" show that these differences are even more striking when we consider the timing at which a previous trade took place. For instance, patents traded a year before a renewal date are about half as likely to expire at that renewal date than patents traded four years ago. Finally, patents traded one year ago are twice as likely to be traded than patents traded four years ago.

The next section develops a model that interprets the key facts.

3 A Model of Patent Trades

The main objectives of the model are to shed light on the underlying quality of patents that are traded, and to study how the cost of technology transfer determines the probability of being traded conditional on renewal as a function of the age of the patent (i.e., the transfer rate as a function of patent age).

To do so, we consider an economy with time indexed by $a = 1, ..., L$ where $L < \infty$

Table 2: Percentage of Active Patents Traded and Expired

	Age	Not Previously Traded	Previously Traded (Years since last trade)		
			Any Year	1 year	4 years
Expiring Decision	5	18.6	12.6	7.1	14.4
	9	29.2	23.4	11.8	23.5
	13	33.0	28.3	15.6	28.5
Trading Decision	4	2.01	4.47	5.05	-
	8	1.56	3.78	4.85	3.58
	12	1.27	2.62	3.27	2.47

represents the finite life of a patent.[11] The economy is populated by a large number of firms $\mathcal{F} < \infty$. In period 1, the economy is also endowed with a large number of patents $K < \mathcal{F}$, which are randomly paired to firms. Some firms are more productive and can obtain more revenues than other firms from a given patent. Thus at the beginning of every period a fraction of all the firms holds exactly one patent, and the rest of the firms can be alternative owners by making succesful acquisition offers.

Firms obtain their revenue from the patents they own and each patent is owned by at most one firm.[12] For simplicity, we also assume that each firm holds at most one patent.

Firms maximize profits. Profits are defined as the expected discounted value of a sequence of per period patent returns x_a minus a renewal fee c_a. The discount factor is β. The payment of the fee extends patent protection. If the fee is not paid, then returns are zero thereafter. Consequently, the value of a firm is exactly the value of its patent. Therefore, the growth of a firm is the growth of patent returns.

The timing is as follows. At the beginning of every period, firms know the per period return of their patent in case it is allowed to expire ($zero$), kept (x), or sold (y) to a potential buyer. Firms choose whether to sell, keep or let the patent expire. Next, the per period return of the patent, let's us $z \in \{0, x, y\}$ be that return, is collected by the new or old owner if the patent has been respectively sold or kept. Finally, at the end of the period, the returns for next period become known, the owner of the patent meets a new potential buyer, and an acquisition offer is received.

In the model, there are two sources that account for how per period returns evolve over the life cycle of a patent. Patent returns can grow within the firm and between firms.

[11]Notice that periods and the age of a patent are interchangeable.

[12]Licensing of patents is in the background. Licensing affects the per period revenue of a patent but not its ownership.

First, the internal process of returns takes place within the firm. This process represents either the arrival of new potential applications that increase the value of the innovation, or the depreciation of returns due to the discovery by other firms of similar technologies. Second, the external growth of returns occurs if the patent is acquired by another firm. This growth accounts for the possibility that some firms are more efficient and can obtain more revenues from the same patent.

Growth of Returns Within the Firm The internal growth of returns is modeled as a stochastic process that captures all sources that affect the growth of patent returns other than the ones associated with the efficiency gains attained from the transfer of patents. For instance, possible interpretations can be (i) lower growth of returns due to imitation, arrival of superior technologies to produce a similar good; or (ii) higher growth of returns due to the arrival of new applications that enhance the returns of the innovation, or learning about the product market of the innovation.

These characteristics motivate a process of returns within the firm that might depend on the age of the patent, a, and the per period patent return, z_a. Notice that z_a is determined right after the decision of selling, keeping or renewing the patent. Thus, z_a can be either zero, x_a, or y_a if the renewal fee was not paid (i.e., patent is allowed to expire), the patent was kept, or the patent was traded. The next period returns are

$$x_{a+1} = g_a^i z_a \qquad a \in \{1, ..., L-1\}$$

where $g_a^i \in [0, B^i]$ is a random variable that represents the growth of internal returns. The random variable is distributed with

$$F_{g_a^i}(u^i; z_a, a) = \Pr[g_a^i \leq u^i; z_a, a]$$

where $F_{g_a^i}(.)$ is the distribution function of g_a^i.

For simplicity, we focus on a process in which the growth of returns within the firm is independent of their level. This case helps to disentangle and highlight two important effects present in the data: the selection and horizon effects. In section 4.4, implications of a process of growth of returns that allows for dependence on the level of returns are also studied.

The case for which the growth of returns is independent of their level has a counterpart in empirical industrial organization. Many studies have persistently found evidence for which the growth of firms is independent of their size, known as Gibrat's law. This general case has not been previously considered in the patent literature.[13] However, we argue that in the theory developed in this paper, the growth of a firm is equivalent to the growth of patent returns. Since each firm holds a patent, and patent protection is their unique source of revenue, success, survival and exit for innovative firms and patents have effectively the same essence.

The following assumption specifies Gibrat's law within the model.

Assumption G: The process of returns of a patent follow Gibrat's law if the internal growth of the returns is independent of their level

$$F_{g_a^i}(u^i; z_a, a) = \Pr[g_a^i \leq u^i; a]$$

Growth of Patent Returns Between Firms In this environment some firms are more efficient and can obtain more revenue than other firms from a given patent. Thus, growth of patent returns between firms is a result of efficiency gains attained through the trading of patents.

In the model, at the end of every period, each holder of a patent meets with a potential buyer. The potential buyer draws an efficiency factor g^e. This efficiency gain represents the efficiency of the potential buyer relating to that of the current owner of the patent. So, at the beginning of every period, the patent return of the potential buyer is defined as

$$y_{a+1} = g^e x_{a+1} \qquad a \in \{1, ..., L-1\}$$

where the random variable g^e is drawn independently and identically from a distribution F_{g^e} with support $[0, B^e]$. Thus, the probability that an efficiency factor is lower than any given number u^e is

$$F_{g^e}(u^e) = \Pr[g^e \leq u^e]$$

The distribution function F_{g^e} is common to all patents at all ages.

However, the diffusion of innovations across the boundaries of a firm is not cost free. An investment must be pursued to make efficient use of the acquired knowledge. This is assumed a fixed cost of technology transfer, τ, independent of the age of the patent and the potential gains from trade. The existence of significant costs of technology transfer is a well known fact documented in the literature of intellectual property transfer. Teece [23] is an early reference.

[13] However, linear depreciation, which is a particular case of Gibrat's law, has been previously used in the literature on estimating the value of a patent (see Lanjouw, Pakes and Putnam [16])

Finally, during the meeting between the potential buyer and the owner of the patent, an offer is made. This offer to buy a patent can be summarized by the age of a patent, a, and the patent return of the potential buyer, y_a. Further, any efficiency gain from transferring the patent must be weighted against the cost of technology transfer. For simplicity, it is assumed that the seller gets all the surplus.

The next section studies the problem that a firm solves.

3.1 The Maximization Problem of a Firm

Consider the problem of a firm that holds a patent prior to its a^{th} renewal. At the beginning of any period the holder of a patent knows its current return if respectively, the patent is kept x, it is sold y, or allowed to expire, zero. Consequently, the decision is made and the patent return is $z \in \{x, y, 0\}$.

Define $V(a, z)$ as the discounted expected value of a patent of return z at age a when the firm is commited to pay the renewal fee.

$$V(a, z) = z - c_a + \beta E[\widetilde{V}(a+1, x', y')|a, z]$$

This value is equal to the current return of the patent, z_a, minus the renewal fees, c_a, plus its discounted option value. The next period internal returns are $x' = g_a^i z$, and the external returns are $y' = g^e x'$. In addition, the operator $E[.]$ denotes an expectation conditional on the return of the current owner and the age of the patent. The option value of a patent is defined as

$$E[\widetilde{V}(a+1, x', y') \mid a, z] = \int \int \widetilde{V}(a+1, u^i z, u^e u^i z) \mathrm{d}F_{g_a^i}(u^i; a, z) \mathrm{d}F_{g^e}(u^e)$$

The value of keeping a patent is $\widetilde{V}^K(a, x, y)$. This only depends on the age of the patent and the per period return of the current owner.

$$\widetilde{V}^K(a, x, y) = V(a, x)$$

Instead, the value of selling is $\widetilde{V}^S(a, x, y)$. This is equal to the value of the patent for a firm with current returns y minus the cost of technology transfer τ. For simplicity, it is assumed that the owner of a patent makes a take it or leave it offer to the buyer.[14] Then,

[14] Alternative bargaining methods do not affect the *qualitative* results of the model.

the value of selling, which coincides with the price of the patent, is

$$\widetilde{V}^S(a, x, y) = V(a, y) - \tau$$

Finally, the holder of a patent decides whether to sell, keep or let its patent expire by solving the following

$$\widetilde{V}(a, x, y) = \max\{\widetilde{V}^S(a, x, y), \widetilde{V}^K(a, x, y), 0\} \qquad a = 1, ..., L$$

where the letters S, K denote the value of sell and keep and zero is the return if a patent is let to expire.

Prior to the description of the equilibrium of the model, several assumptions are needed to characterize basic properties of the value function. A sufficient condition is that returns are bounded, or that they depreciate with age.

Basic continuity conditions for the process of growth of returns within the firm are assumed (i.e., generality conditions to guarantee the continuity and existence of the value function): First, there exists an ε such that $E[{x_a}^{1+\varepsilon} \mid a = 1] < \infty$. Second, $F_{g_a^i}(u^i\ ; z_a, a)$ is continuous in z at every u^i except, possibly, at values of u^i at which $F_{g_a^i}(u^i\ ; z_a, a)$ has a discontinuity in u^i.

In the model it is also considered that the expected growth of returns within the firm decreases with the age of the patent. This is equivalent to assuming that $F_{g_a^i}(u^i; z_a, a)$ is weakly increasing in a. We focus on processes of growth of patent returns such that patents with high returns today are more likely to have high returns tomorrow. This translates to assuming that the $\Pr(zu_a^i \leq \overline{u}|z)$ is weakly decreasing in z, that is first order stochastic dominance in the evolution of per period returns within the firm.

The above conditions will be carried out through the paper (further assumptions are used later in the paper, and they are stated in those cases).

The following Lemma states that the value function of a patent is continuous, weakly increasing in the returns of the patent and weakly decreasing in patent age.

Lemma 1 *The value function $\widetilde{V}(a, x, y)$ is continuous and weakly increasing in the current return of the holder of the patent, x, and the return of the potential buyer, y. The option value $E\widetilde{V}(a + 1, x', y'|z, a)$ is weakly decreasing in a.*

Proof. See Appendix. ■

The next section develops the main implications of the model.

4 The Selection and Horizon Effect

This section is devoted to the selection and horizon effects and their role in explaining the stylized facts observed in the data. To do so, the characteristics of the policy functions of the problem of the firm are analyzed. First, we focus on the selection effect. In order to do this, the comparative statics for the case of a fixed patent age are studied. Subsequently, we study the foundations of the horizon effect. This conveys the analysis of how the two cutoff rules vary over the life cycle of a patent. Finally, we examine the case in which the growth of returns within the firm depend on their level.

The policy functions are two cutoff rules $\{\widehat{x}_a(\tau)\}_{a=1}^L$ and $\{\widehat{g}_a(x,\tau)\}_{a=1}^L$.[15] They depend on the current return of a patent, x, the return of a potential buyer, y, and the parameters of the model such as the renewal fees c_a, and the cost of technology transfer τ. The policy space can be summarized by these policy rules. First, $\widehat{x}_a(\tau)$ is defined as the patent return that makes the holder of a patent indifferent between keeping or letting a patent of age a expire. Second, the cutoff $\widehat{g}_a(x,\tau)$ represents the potential external growth in returns that makes a firm indifferent on whether to trade a patent or not.

The two rules describe three regions, E, K, and S in the policy space of patent of age a. These regions correspond respectively to: let the patent expire, keep it or sell it. The regions E and K are separated by a straight line $\widehat{x}_a(\tau)$, which is independent of the external growth of returns g^e. For low current patent returns, that is $x < \widehat{x}_a(\tau)$, the firm chooses between not renewing and selling the patent. So, the cutoff $\widehat{g}_a(x,\tau)$ separates the areas E and S. Finally, for sufficiently high returns, that is $x > \widehat{x}_a(\tau)$, letting the patent expire is not an optimal choice, thus $\widehat{g}_a(x,\tau)$ delineates the K and S regions. The following figure shows a particular example of the policy space, under Gibrat's law.

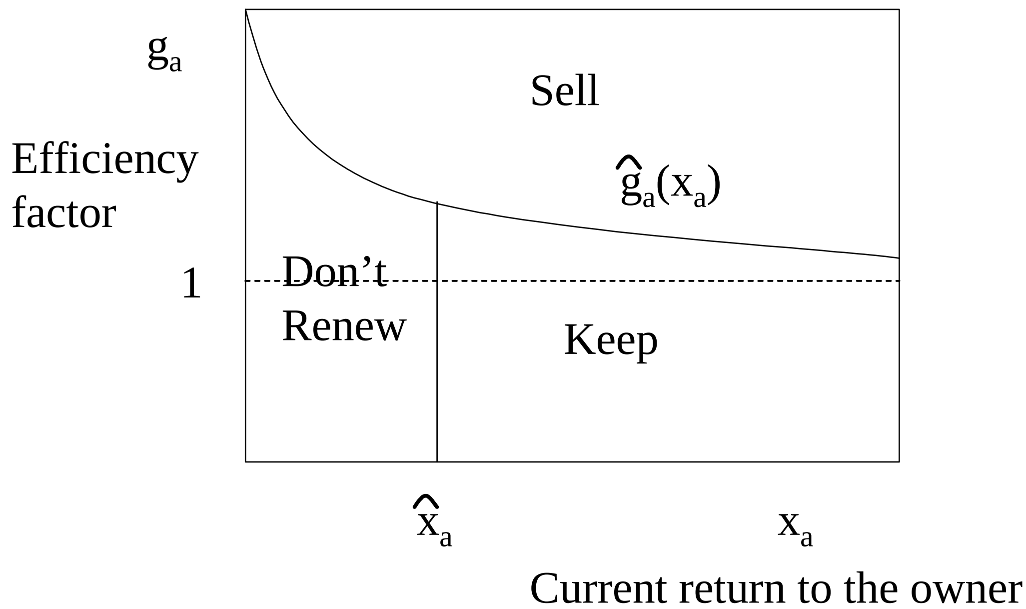
The next section studies the selection effect. This coincides with the examination of the problem of selling, keeping or letting expire a patent of age a. In other words, it characterizes the underlying properties of the patents that are traded.

4.1 The Selection Effect: Policy Functions for Fixed Age

This section explores the individual heterogeneity in patent characteristics that makes some patents more likely to be traded. In the model, patents are defined by a pair (a,x), where a is the patent age and x is the per period return. The analysis of this section is focused on how ex ante patent returns determine the patent's likelihood of being traded, and similarly its likelihood of being allowed to expire.

[15] Both cutoff rules also depend on the states (x,y,a), the renewal fees, c_a, and the cost of technology transfer, τ. However, c_a and y have been omitted.

Figure 2: The Policy Space



4.1.1 Which Patents Are Traded?

Evidence presented in Section 2 shows us a few things. First, patents with a higher number of total citations received are more likely to be traded. Second, traded patents are more likely to be retraded and less likely to expire. These patterns correspond, respectively to stylized facts 4 and 3 mentioned earlier.

The model can account for these patterns. The mechanism is simple. The cost of technology transfer creates a selection effect so that patents with higher ex ante per period returns are more likely to be traded. In addition, since new owners are more efficient in using the patent, its return increases even further after being traded. Thus, patents of higher quality are more likely to be traded, and traded patents are more likely to be retraded and less likely to expire.

The selection effect arises in the following way. In the model, any patent owner will sell a patent if it receives an offer from another firm with relative efficiency higher than the efficiency that offsets the cost of technology transfer, that is $g^e > \widehat{g}_a(x, \tau)$. So, the probability that a patent is traded concides with the likelihood of receiving such offers, which is $\Pr[g^e \geq \widehat{g}_a(x, \tau)]$. This probability, which is a function of age and patent return, is defined as the transfer rate of a patent of age a.

The characteristics of the transfer rate are ultimately determined by those of the cutoff $\widehat{g}_a(x, \tau)$. In other words, if $\widehat{g}_a(x, \tau)$ is decreasing in x, then the probability of being traded increases in x. Therefore, better patents are more likely to be traded if $\widehat{g}_a(x,\tau)$ is decreasing with x.

In order to show that $\widehat{g}_a(x,\tau)$ is decreasing in x, we must consider two separate cases: (i) $x < \widehat{x}_a(\tau)$, and (ii) $x > \widehat{x}_a(\tau)$. This separation is necessary because the function $\widehat{g}_a(x,\tau)$ is defined differently in these two parts of the policy space. In the first case, when patent returns are sufficiently low, that is $x < \widehat{x}_a(\tau)$, the proof is obvious. However, I briefly explain it because it provides a clear intuition of the selection effect.

When per period patent returns are low, the choices of the firm are just whether to sell or let the patent expire. A seller is indifferent to selling or allowing the patent to expire if $\widetilde{V}^S(.) = 0$. We can show that there exist $\widehat{g}_a(x,\tau)$ such that the firm is indifferent between these two choices, i.e., $V(a, \widehat{g}_a(x,\tau)x) - \tau = 0$. The left hand side of the equation is the value of selling a patent, and the right hand side is the value of an expired patent. If x increases, then it must be the case that $\widehat{g}_a(x,\tau)$ decreases with x to keep the equality holding. So, the higher x is, the lower $\widehat{g}_a(x,\tau)$ is to cover the cost of technology transfer. The result we want to show follows because the value function of a patent is weakly increasing in the level of current returns as shown in Lemma 1.

Instead, if per period returns are sufficiently large, that is $x > \widehat{x}_a(\tau)$, then the firm acts at the margin between selling and keeping the patent. The proof is a bit more elaborate than the argument of low returns, but the result still holds. It can be shown that if the internal growth of returns is independent of the level, that is Gibrat's law, then the cutoff $\widehat{g}_a(x,\tau)$ is indeed monotonically decreasing in x. This result means that owners of patents with larger returns demand less external growth of returns in acquisition offers for their patents to be traded. The following proposition summarizes the result.

Proposition 2 *If assumption G holds and $\tau > 0$, then (i) the function $\widehat{g}_a(x,\tau)$ is weakly decreasing for all x, (ii) $\widehat{g}_a(x,\tau) > 1$, (iii) the probability of being traded is increasing with x.*

Proof. See Appendix. ■

The intuition of the proof is not difficult. For instance, consider the case of a myopic firm, which is when the discount factor β is zero. Later in this section the case of positive discount factor will be considered. For a myopic firm, the value of a patent at a given age is just the current returns. Consequently, the value of keeping and selling is the current return x and the return of a potential buyer minus the cost of technology transfer $y - \tau$. A firm is indifferent between both choices if there exists an efficiency gain $\widehat{g}_a(x,\tau)$ such that $y = x\widehat{g}_a(x,\tau)$ holds. Now, fix y. The larger x is, the lower $\widehat{g}_a(x,\tau)$ must be to cover the cost of technology transfer and for the equality to hold. Therefore, patents that are traded are ex ante of higher quality that the average patent.

For the general case in which the discount factor is positive, the argument for a proof is as follows. We must show that for a given efficiency gain between a potential buyer and seller, let's say $\overline{g}_a^e$, and patent age a, the difference between the option value of selling and keeping is weakly increasing in the current return of the patent, z_a. In other words, let us increase the current return of the patent z_a a bit, now we can show that the difference between selling and keeping also increases. Then, there exists a lower efficiency gain g_a^e lower than $\overline{g}_a^e$ such that the difference in expectations (see below) goes back to the level before increasing z_a. Therefore, in order to show that $\widehat{g}_a(x, \tau)$ is decreasing in x, it is sufficient to show that the following expression increases as a function of z_a.

$$E\left[\widetilde{V}(a+1, x', y')|a, g_a^e z_a\right] - E\left[\widetilde{V}(a+1, x', y')|a, z_a\right]$$

However, the proof is a bit subtle. This is due to the fact that, because both option values are weakly increasing in z_a, then the change in the difference with respect to z_a can be ambiguous. In fact, the result depends on the characteristics of the process of internal growth of returns, which is $F_{g_a^i}(u^i; z_a, a)$. Proposition 1 shows that proportional growth of returns, that is Gibrat's law, guarantees that the above difference in expectations is weakly increasing in the current return of the patent. Section 5 explores other processes of internal growth of returns that might depend on the level of returns.

This section has shown that patents with higher ex ante returns are more likely to be traded, and its implications are traded patents are more likely to be retraded and less likely to expire. This result had relied on two key assumption of the model: (i) a fixed cost of technology transfer, and (ii) that gains from trade are relative to the current return of the patent (i.e., the Gibrat's assumption, that is a proportional growth rate, is a particular case of this).

First, if the cost of technology transfer, instead of a fixed cost, is a proportion of x (i.e., $\tau = \alpha x$, where $\alpha \in (0, 1]$), then the result of proposition 1 still holds from any $\beta > 0$. However, if the cost of technology is either zero or proportional to the value of the buyer prior to taxes, $V(a, y)$, then the probability of being traded is independent of the return of the patent (i.e., it is flat with respect to x). So, the implication of these two hypothetical cases would not coincide with the facts observed in the data.

Second, if the process of external growth of returns that allows for trades in patents was in levels rather than in growth rates (i.e., the random draw could be a direct return y, rather than a g^e), then the probability of being traded would be decreasing in the return of the patent. This would contradict the facts observed in the data. Thus, the model shows that arrival offers must be linked to the current return level of a patent in order to explain the data.

Therefore, the model presents minimal assumptions that account for the stylized facts in the data. In particular, it predicts the necessity to incorporate costs of technology transfer. Finally, it shows the need of modeling potential gains from trade conditional on the current return of the patent to explain the stylized facts (i.e., a relative efficiency gain).

4.1.2 What Patents Are Allowed to Expire?

Patents expire because their returns are sufficiently low and not enough to cover the renewal fees. This is an obvious implication of the construction of patent renewal models (see Schankerman and Pakes (1986) and Pakes (1986)).

In the model, the result holds because the option value of a patent in increasing in its return, which was shown in Lemma 1. Any patent of age a , if current returns are $x < \widehat{x}_a(\tau)$ and a potential buyer offer is characterized by an efficiency gain $g^e < \widehat{g}_a(x, \tau)$, then the patent will be allowed to expire.

The next section focuses on the effects of age in the trading and renewal decision of a patent.

4.2 The Horizon Effect: How do Policy Functions Change with Age?

This section studies how the transfer rate of patents varies over their life cycle. In particular, it analyzes the comparatives statistics, with respect to the age of a patent, of the policy functions for a fixed patent return. The interplay between the cutoff $\widehat{g}_a(x, \tau)$ and the age of a patent determines the probability of being traded conditional on survival. This probability might be of interest because it explains the expected efficiency gain from a patent transfer. In addition, the section also looks at how the likelihood to expire changes over the patent's life cycle. This is interpreted by how the cutoff $\widehat{x}_a(\tau)$ affects the probability of being allow to expire as a function of age.

4.2.1 The Transfer Rate of a Patent Over Its Life Cycle

The stylized fact 1 shows that the transfer rate of patents monotonically decreases since their issue date with the exception of the renewal dates. Immediately after renewal, the transfer rate discontinuously increases.

In the model, there is a horizon effect that explains that transfer rates decrease as the patent gets closer to its expiration date. A shorter horizon implies less time to amortize the fixed cost of technology transfer.

The transfer rate of a patent is the probability of being traded conditional on survival. Patents are traded if efficiency gains from purchasing offers are such that $g^e > \widehat{g}_a(x, \tau)$. So the probability of being traded is just $\Pr[g^e > \widehat{g}_a(x, \tau)]$, which, in particular, depends on a. In order to show that the transfer rate is decreasing with age, it suffices to prove that the cutoff $\widehat{g}_a(x, \tau)$ is increasing in a for a fixed x.

The argument to show the result can be divided into two parts. First, it is easy to show that the result holds if returns are low, $x < \widehat{x}_a(\tau)$. When returns are low, $x < \widehat{x}_a(\tau)$, the firm's optimal choices are between keeping the patent or allowing it to expire. A firm is indifferent between these two choices if $V(a, x\widehat{g}_a(x, \tau)) = 0$. By Lemma 1, we know that the value of a patent is decreasing in age and increasing in returns. So, if a increases, then it must be the case that $\widehat{g}_a(x, \tau)$ also increases to keep the equality holding. Thus for a fixed low patent return x, offers with higher efficiency gains are required for a trade to take place as a patent gets older.

Second, proposition 2 generalizes the result for all returns and for any period other than the renewal dates. It is assumed that the growth of internal returns is independent of the level.

Proposition 3 *If assumption G holds and $\tau > 0$, then for all x and τ, (i) $\widehat{g}_a(x, \tau)$ is increasing in a, and (ii) the probability of being traded conditional on survival is weakly decreasing in age, a.*

Proof. See Appendix. ■

The argument of the proof when returns are high (i.e., $x > \widehat{x}_a(\tau)$) is to show that, for a fixed efficiency gain g^e, the difference between the value of selling and keeping a patent monotonically decreases, and also converges to zero at most at period L. If the internal growth of returns is independent of the level, then the result follows because both the value functions of keep and sell decrease in average proportionally to their current return level.

4.2.2 The Expiration Rate of a Patent Over Its Life Cycle

This section studies the probability of being expired as a function of the age of the patent. The foundations of this result were shown in Schankerman and Pakes [21] and Pakes [19].

Evidence from U.S. patents presented in fact 3 shows that the proportion of patents allowed to expire increase with the age of a patent. In the model, to explain this result it is sufficient to show that the cutoff $\widehat{x}_a(\tau)$ is increasing in the age of the patent. The following proposition shows that.

Proposition 4 *If the renewal fees schedule is weakly increasing, then the cutoff* $\widehat{x}_a(\tau)$ *is weakly increasing in age. Therefore, the probability of being allowed to expire weakly increases as a function of patent age.*

Proof. See Appendix. ■

More interesting, however, are the effects of the renewal decision on the distribution of patent returns as a function of patent age. This is particularly relevant when the number of renewal dates is low, as it is in the U.S. patent system. In this case, the number of patents that might expire in each of the renewal dates can be substantially large. Thus, immediately after a large fraction of patents expires, the average patent return increases. The next section develops and explores this result in-depth.

4.3 A Horse Race: The Selection Compared to the Horizon Effect

Renewal dates and their implications into the trading decision are interesting events on which to focus.

These events link the results of the model with the data. The model predicts that immediately after renewal, the selection and horizon effects display opposite trends towards the probability for a patent of being traded. On the one hand, right after the renewal date, the average patent return increases, so patents are more likely to be traded according to the selection effect. On the other hand, the horizon effect implies that as the age of a patent increases, patents are less likely to be traded . Therefore, there exists a horse race between the two effects. The race determines the observed proportion of patents that are being traded.

In addition, the model can separately identify the impact of the two effects. To do so, let us consider the possibility of shrinking a period in the model to just less than seconds (i.e., for understanding, a continuous-time model would be a good approximation). If at equilibrium a positive measure of patents expire, then immediately after renewal the distribution of returns dominates stochastically the one existing before the renewal decision. Thus, the selection effect implies that the probability of being traded increases. However, as the time passes by since the renewal date, the horizon effect becomes stronger so that eventually it offsets the selection effect. Therefore, the model helps to disentangle the forces under the selection and horizon effect.

Evidence from the data seen in fact 1 shows that the probability of being traded discontinously increases in all renewal dates. Then this probability monotonically decreases until the next renewal date. Therefore, according to the model this is a sign that shortly after renewal the selection effect is sufficiently stronger than the horizon effect, but also the selection effect vanishes fast.

4.4 Level Dependence in the Growth of Returns within the Firm

Previous sections of the paper have focused on the case for which the growth of patent returns within the firm was independent of the level.

However, the existing patent literature that has estimated the value of patents, such as Lanjouw and Pakes, have considered variations of a process of growth of internal returns that also depends on their level. The following assumption defines an explicit stochastic process of growth of returns within the firm with level dependence.

Assumption L: The random variable g_a^i is distributed with a function that depends on the return z_a and patent age a.

$$F_{g_a^i}(u^i; z_a, a) = \left\{ \begin{array}{lll} 0 & \text{if} & u^i < \delta \\ \Pr[g_a^i \leq u^i; z_a, a] & \text{if} & u^i \geq \delta \end{array} \right\}$$

such that the function $F_{g_a^i}(u^i; z_a, a)$ is increasing in z.

If $F_{g_a^i}(u^i; z_a, a)$ does not depend on age, then the process is defined as constant learning (LC). If $F_{g_a^i}(u^i; z_a, a)$ is increasing in a, then the process is defined as diminishing learning (LD). An extreme case of LD, let is call it LDE, is one in which there exist an $\overline{a} < L$ such that $\Pr[g_a^i = \delta; z_a, \overline{a}] = 1$.

This assumption considers a process such that in every period returns either depreciate at a rate $\delta \in (0, 1)$ or grow at a rate $g^i > \delta$.

The term learning has been previously used in the literature on estimating the value of a patent. Learning allows for the possibility of new opportunities or applications that enhance the returns of a patent to be discovered (i.e., so that their growth rate can be higher than depreciation).

The following proposition proves that, if the internal growth of returns also depends on the level as in the learning process specified in assumption L, then $\widehat{g}_a(x, \tau)$ is not necessarily increasing for all a. In particular, it states sufficient conditions so that $\widehat{g}_a(x, \tau)$ is weakly decreasing for sufficiently young patents. The main implication is that the probability of being traded is hump-shaped as a function of the age of the patent.

Lemma 5 *If the support of the random variables g_a^i and g^e is bounded above, then $\left|\frac{\Delta V(a+1,x',y'|z)}{\Delta a}\right| \leq Q$.*

Proof. See Appendix. ■

Proposition 6 *Let $\overline{a}$ be the patent age at which learning completely vanishes. If assumption LDE, the support of the random variables g_a^i and g^e is bounded above, and the $\Pr[g_a^i \leq u^i; z_a, a]$ is sufficiently concave with a, then there exists an age $a^* < \overline{a}$ such that for $a < a^*$ the function $\widehat{g}_a(x,\tau)$ is weakly decreasing in a. In addition, if $a \geq a^*$, then $\widehat{g}_a(x,\tau)$ is weakly increasing in a. Therefore, the probability of being traded conditional on survival is hump-shaped as a function of the age of a patent.*

Proof. See Appendix. ■

The rational of this proposition is that if the probability of learning decreases sufficiently fast with age and is also decreasing in the return of a patent, then the option value of keeping a patent with low return experiences larger proportional losses than patents with high returns. If patent returns are low, then learning possibilities initially explain a large part of the option value of a patent because learning is more likely for patents with low current returns. Instead, if current returns are high (for instance, upon a potential trade), then learning, although still important, is less weighted in the option value of a patent because newer profitable applications are less likely. Consequently, if learning fades sufficiently fast as age increases, then the option value of a patent with lower current returns experiences larger proportional losses on average than a patent with higher current returns. Thus, given a fixed efficiency gain, patents that were not traded when very young might be traded when they are slightly older. So, $\widehat{g}_a(x,\tau)$ is weakly decreasing in a. However, as time passes the learning effect vanishes and then the process of returns converges to a scenario in which returns depreciate, a particular case of Gibrat's law. Thus, the effect of trading soon surely dominates again. Then, the function $\widehat{g}_a(x,\tau)$ is weakly increasing in a, and the probability of being traded is weakly decreasing in a. Therefore, $\widehat{g}_a(x,\tau)$ is U-shape, which implies that the probability to be traded conditional on survival is hump-shaped with age.

Nevertheless, a learning process by itself does not necessarily imply that $\widehat{g}_a(x,\tau)$ is U-shaped in a. The previous proposition illustrates the necessity of strong assumptions to show that $\widehat{g}_a(x,\tau)$ is not always weakly decreasing in a. Diminishing learning guarantees that the slope of the option value of a patent for the seller is steeper, however it might be the case that it has less steepness than the one of the buyer. In fact, if learning does not diminish fast enough, then the horizon effect dominates and consequently $\widehat{g}_a(x,\tau)$ is increasing in a.

As a matter of fact, we can show that if the process of learning is independent of the age of the patent, then $\widehat{g}_a(x,\tau)$ is weakly increasing in a. This means that the probability to be traded conditional on survival is weakly decreasing. The following proposition shows this result.

Proposition 7 *If $F_{g_a^i}(u^i; z, a)$ is independent of a, then (i) the function $\widehat{g}_a(x,\tau)$ is weakly increasing in a. And (ii) the probability of being traded is weakly decreasing in the age of the patent.*

Proof. See Appendix. ■

5 Estimation and Identification

In this section we discussed the estimation and identification of the parameters of the stochastic specification of the model. A specific feature of our estimation strategy is the use of citations received in order to estimate the value of patents. The previous literature estimating the value of patents has uniquely relied on methods that provide information on the left tail of the distribution of the value of patents, such as the renewal proportions. The advantage of our method is that we complement the information contained in the renewal proportions by using the deciles of the distribution of citations received at given ages. Since, the distribution of citations received apply to the whole spectrum of patent values, the identification of our estimates do not rely on the tail of a given distribution as the estimations carried out by Schankerman and Pakes [21], Pakes [19], Lanjouw [15], etc.

The rest of the section is organized as follows. First we introduce a stochastic specification of the model of patent transfers and renewals developed in previous sections. Second, we discuss the estimation strategy and present preliminary estimates of the parameters of the model. Finally, we argue the identification of the estimated parameterers.

5.1 Stochastic Specification of the Model

The stochastic specification that is estimated contains 12 parameters. We decided to set the discount factor $\beta = 0.9$ as in Pakes [19]. The rest of the parameters, which are jointly estimated, are those that are contained in the initial distribution of returns, the internal growth of returns, the external growth of return, the cost of technology transfer, the probability of a random trade, the probability that a patent return drops to zero, and three parameters of a process that generates citations received.

In the model the invention decision is exogenous. So, patents are granted to a fraction of firms. Since patent renewal fees are due by the end of the 4^{th}, 8^{th} and 12^{th} year since the grant date of a patent, we define the age of a patent a as the number of years from its grant date.

The initial quality of an innovation, at the time a patent is granted, is assumed to be distributed lognormally according to:

$$\log(x_a) \sim F_{IR}(\mu, \sigma_R)$$

where the age at the grant year, $a = 1$, and $F_{IR}(.)$ is a normal distribution.

There are two sources that explain how returns evolve over time: internal and external growth of returns. The internal growth of returns occurs within the firm. This process is stochastic and illustrates both the possibility that patent per period returns increase over time as a result of the discovery of new succesful applications, and perhaps, patent returns decrease due to the arrival of competing technologies. The stochastic process is specified as a random variable g_a^i with distribution function $F_{g_a^i}(u^i; a)$.

$$F_{g_a^i}(u^i; a) = \left\{ \begin{array}{lll} 1 - \theta & \text{if} & 0 \leq u^i < \delta \\ 1 - \theta + \theta(1 - \exp(-\frac{u^i}{\sigma_a})) & \text{if} & u^i \geq \delta \end{array} \right\}$$

where the parameter σ_a is defined as $\sigma_a = \phi^{a-1}\sigma$, $a = \{1, ..., 17\}$. Since it is assumed that $\phi \leq 1$, then σ_a is decreasing as age increases.

Alternatively, the process of internal growth of returns is equivalent to one in which on the one hand there is an obsorving state of zero returns with probability $(1 - \theta)$, and if returns are not obsorbent then the internal growth of returns is the maximum between δ and a draw of a single parameter exponential distribution. So the likelihood of arrival of succesful application coincides with the the probability that the internal growth of returns are higher than the depreciation factor, $\Pr(g_a^i > \delta)$, which we define as the *learning likelihood*.

The depreciation of patent returns is characterized by parameters δ and θ. The parameter θ determines the proportion $(1 - \theta)$ of all active patents that in a given period loose all their patent returns (i.e., next period return will be zero). The parameter δ determines the depreciation rate $(1 - \delta)$ of those patents that their returns did not become zero as a result of the θ obsorbing state, and that their learning draw is not higher than δ.

The parameters ϕ and σ determine the learning likelihood as a function of the number of years since the grant date of a patent. A small ϕ involves that potential learning possibilities vanish very fast. A low σ implies that the opportunities of learning are not sizeable and that the probability that a patent is valuable is also small.

The growth of returns between firms is modeled as a random variable g^e with an exponential distribution function

$$F_{g^e}(u^e) = 1 - \exp(-\frac{u^e}{\sigma^e})$$

The next period returns of the owner of a patent depend on the internal growth of returns and whether the patent was traded or not.

$$x_{a+1} = g_a^i * z_a$$

Where $z_a \in \{x_a, y_a, 0\}$ is the per period return if the patent is respectively, kept, sold and allowed to expire. And $y_a = g^e x_a$.

Moreover, in the estimation specification we allow for transactions of patents that are random with probability ε. In the data, we conjecture that there are patent recorded for other reasons other than technology transfer (i.e., in the model terms these transactions do not involve neither a cost of technology transfer not a gain from trade). To be consistent with that potential source of noise of the data, we allow for noisy trades in our estimation strategy. We understand the conservative nature of our assumption in that we will obtain a lower bound on the gains from trade.

Finally, we introduce citations received into the model. To do so, we consider an adhoc generating process of per period citations received. We assume that the number of citations received in a year, k, is distributed as a Negative Binomial distribution with parameters (γ, p), where γ is as follows.

$$\gamma = \exp(\sum_{j=1}^{17} 1_j \rho_j + \lambda_1 \ln(1 + z_a))$$

z is the per period patent return, and $1_j = 1$ if $a = j$, and $1_j = 0$ otherwise.[16]

The specification we propose allows patents to received citations independently of their revenue process, this is captured by the dummies of age ($\rho_j' s$). We propose such a method because by examining the data we conjecture that there might exist a process underlying citations which is dependent of age but independent of patent's revenue. An element that will be benefitial to estimate part of this process is the fact that expired patents still get citations.

The observation that expired patents still get citations will allow us to calibrate param-

[16]The mean of a Negative Binomial with parameters (γ, p) is $E(k) = \frac{\gamma(1-p)}{p}$. The variance of a Negative Binomial with parameters (γ, p) is $Var(k) = \frac{\gamma(1-p)}{p^2}$.

eters $\rho_5, ..., \rho_{17}$. In particular, given p, the parameters $\rho_5, ..., \rho_{17}$ can be pinned down by the mean of the citations received by expired patents.

The complete stochastic specification to be estimated contains 15 parameters to be estimated, $w = (\mu, \sigma_R, \sigma^e, \tau, \varepsilon, \theta, \delta, \phi, \sigma, \rho_1, \rho_2, \rho_3, \rho_4, p, \lambda_1)$.

5.2 Estimation Strategy and Estimates

The parameters of the model are structurally estimated using the simulated generalized method of moments. This method involves finding parameters w so that they minimize the distance between the empirical moments, defined as those from the data, and the simulated moments generated by the model. The moments generated by the model are simulated because they cannot be solved analytically due to the structure of the model. In particular, our estimation strategy consists in fitting the proportion of active patents that are traded conditional on being previously traded and not previously traded (32 moments), the proportion of active patents allowed to expire conditional on having been previously traded and not previously traded (6 moments), the mean of the number citations received at a given age from age 2 to age 17 conditional on previously traded and conditional on not previously traded (32 moments), and the mean of citations received at age 1 of active patents. In total there are 71 moments. The following algorithm explains the details of the procedure.

5.2.1 Estimation Algorithm

The parameters of the model are structurally estimated using the simulated generalized method of moments. A simulating procedure was first applied in a patent renewal model by Pakes [19], however Pakes' approach used a maximum likelihood estimator. Lanjouw [15] used the simulated generalized method of moments as we use here.

To estimate the parameters of the model we find the simulated minimum distance estimator[17], $\widehat{w}_N$, of the true k parameter vector, w_0[18], and N is the sample size.

$$\widehat{w}_N = \arg\min_{w} v_j ||h_N - \eta_N(w)||$$

The vector h_N is defined as the sample hazard proportions or empirical moments, and the vector $\eta_N(w)$ as the ones that are simulated. In particular, the vector of empirical moments contains the following.

[17] The capital letter N denotes the sample size.

[18] Pakes and Pollard (1989) have showed conditions under which $\widehat{w}_N$ converges to w_0, and $\sqrt{n}(\widehat{w}_N - w_0)$ satisfies a central limit theorem.

The vector w_0 is defined as the unique solution to the equation

$$G(w) = \sum_{j=1}^{71} v_j \left(\frac{\eta_j(w) - \eta_j}{\eta_j} \right)^2 = 0$$

where η is the vector of the true hazard probabilities, $\eta(w)$ are the hazards predicted by the model with paramete vector w, and the v_j are the elements of the diagonal of the weighting matrix. The weighing matrix is defined as

$$v(w) = \text{diag}[\sqrt{n/N}]$$

where $n = [n_1, ..., n_j, ..., n_{71}]$, and n_j is the number of patents in which moment j is conditioning on. Note that the metric of the distance we choose is the proportional distance between the empirical and simulated moment. We use this metric because the value of the moments of the expiring decision is

Using the simulated minimum distance estimator requires to simulate the moments $\eta_N(w)$ and to minimize the distance between these moments and their empirical counterparts.

The simulated moments $\eta_N(w)$, given a vector of parameters w, are generated in the following way. First, we solve the model recursively. We calculate the value function at age a at a number of selected grid points. Next we find the cutoff rules $\widehat{x}_a$ and $\widehat{g}_a(x)$ as function of w and c_a. In order to calculate the value function at age $(a-1)$, we approximate the double integral that defines the option value of a patent with quadrature approximation methods. Note that the limits of the integral will be defined as a function of the cutoff rules $\widehat{x}_a$ and $\widehat{g}_a(x)$.

Second, we calculate the simulated moments as the average obtained from S simulated populations of N patents. Each simulated population of patent consists of taking pseudo random draws from the distribution of initial returns F_{IR}, and then we pass each initial patent return through the stochastic process of returns of the model implied by distribution of internal growth of returns $F_{g_d^i}$ and the distribution of external returns F_{g^e}, etc.

Finally, we average all simulations and calculate the simulated moments $\eta_N(w)$[19]. And to minimize the distance between the empirical and the simulated moments we use a minimization algorithm based on simulated annealing methods.[20]

[19]We run 15 simulations, and in each simulation we have 453,683 patents with their respective draws of initial returns, internal and external growth of returns.

[20]In particular, we use simulating annealing methods. In particular, the amebsa and amotsa subroutines as describes in the Numerical Recipes for Fortran at http://www.library.cornell.edu/nr/cbookfpdf.html. See Section 10, Minimization or Maximation of Functions.

Table 3: Estimates

Description (Parameter)	Estimate
Depreciation factor (δ)	0.9226
Full Depreciation (θ)	0.9577
Learning Factor from Internal Growth of Returns (ϕ)	0.4242
Fixed Component from the Mean of the Internal Growth of Returns (σ)	0.9221
Mean parameter of the Lognormal Initial Distribution (μ)	8.1740
Std. Deviation parameter of the Lognormal Initial Distribution (σ_R)	1.2938
Cost of Technology Transfer (τ)	26,298
Mean External Growth of Returns (σ^e)	0.4133
Random Trades (ε)	0.00817
Citations process parameters	-

5.2.2 Estimates

Our estimates and standard errors are presented in Table 3[21].

5.2.3 Identification

We decided to set the discount factor $\beta = 0.9$ as in Pakes [19]. The rest of the parameters of the model are jointly estimated following the algorithm explained above. The main source of identification of the value of patents is the use of the schedule of renewal fees and their Dollar value together with the functional form assumptions. So, there is no information in the data we currently use in the estimation that identifies an upper bound of the value of patents that are renewed at all renewal dates. The fact that owners allow their patents to expire indicate that the present value of their future expected patent returns is below the one of the renewal fee. Obviously, it is being assumed the fact that owners of patents are willing to pay renewal fees expecting that their future returns will be high enough to compensate these costs.

The parameters of the process of internal growth of returns, which are (δ, ϕ, σ), and the ones from the initial returns, (μ, σ_R), are jointly indentified from the proportion of active patents that expire conditional on having been previously traded or not, and the proportion of active patents that are traded conditional having been previously traded or not. In particular, ϕ is identified from the shape of the transfer rate when a patent is young. A higher ϕ implies that the speed at which learning vanishes is smaller, so the the

[21] We use a bootstrap method to obtain the standard errors of the parameters estimates of the model. In particular, we do the following. We take the parameter estimates and calculate the simulated moments generated by the model. Then, we draw a S simulations of N patents, and apply the estimating procedure as if the newly simulated moments where the empirical moments. We repeat this procedure 10 times in order to simulate the distribution of parameter estimates.

transfer rate has more curvature early in the life of a patent. The parameter δ is identified in part from the expiration rate and the curvature of the transfer rate as patents get closer to their expiration date. Moreover, the parameters μ, σ_R are identified jointly from the moments that relate to the expiration decision and also from the level of the transfer rate conditional on previously traded or not, especially early in the life of a patent

The parameters ε and σ_e are identified as follows. The parameter ε is jointly identified from both the level of the proportion of active patents that expire conditional on having been traded, and the proportion of active patents that are traded conditional on having been previously traded. A higher ε implies that a higher proportion of active and previously traded patents will expire, and a lower proportion of active and previously traded patents will be retraded. Furthermore, σ^e is identified from the level of the transfer rate, conditional on previously traded or not, over the whole life cycle of a patent.

Finally, the parameters of the process that generate citations received, i.e., $\lambda_1, p, \rho 1, .., \rho 4$ are identified using the mean of citations received of previously traded and not traded patents from age 2 to 17 and the mean number of citations received of active patents at age 1.

The cost of technology transfer is identified as follows.

Cost of Technology Transfer (τ) The cost of technology transfer is identified by the size of the jumps of the transfer rate that are observed immediately after each renewal date and the curvature of the proportion of active patents that are trade conditional on previously traded or not.

The identification strategy is as follows. Let us start considering the case in which the transaction cost is zero. In this case, the model predicts that the transfer rate will be *flat* over the life cycle of a patent except at the renewal rates. Immediately after renewal, the transfer rates increase. In particular, the size of the jumps, which are temporary (i.e. only at the renewal date), depend on the size of the renewal fee c_a. The renewal fee together with the process of external returns determine the curvature of the policy function $\widehat{g}_a(x)$ for any $x < \widehat{x}$ (notice that in this case $\widehat{g}_a(x) = 1$ for all $x > \widehat{x}$). The size of the area delimited by the intersection of $\widehat{g}$ with the upper bound of g^e and $\widehat{x}$ determines the jump of the transfer rate. Finally, as age increases, $\widehat{g}_a(x)$ will shift up, then the area between $\widehat{g}_a(x)$ and the upper bound of its support determines the number of patents traded and consequently the curvature of the transfer rate as a function of age.

Table 4: Fit of the Expiration Rate Conditional on Previously or Not Previously Traded

A. Number of Patents Allowed to be expired as a Percentage of Previously Traded Active Patents

Age of the Patent (Years)	Model	Data
5	14.02	12.57
9	20.03	23.39
13	22.07	28.35

B. Number of Patents Allowed to be expired as a Percentage of Not Previously Traded Active Patents

Age of the Patent (Years)	Model	Data
5	19.46	18.62
9	29.78	29.25
13	33.18	33.01

6 Estimation Results

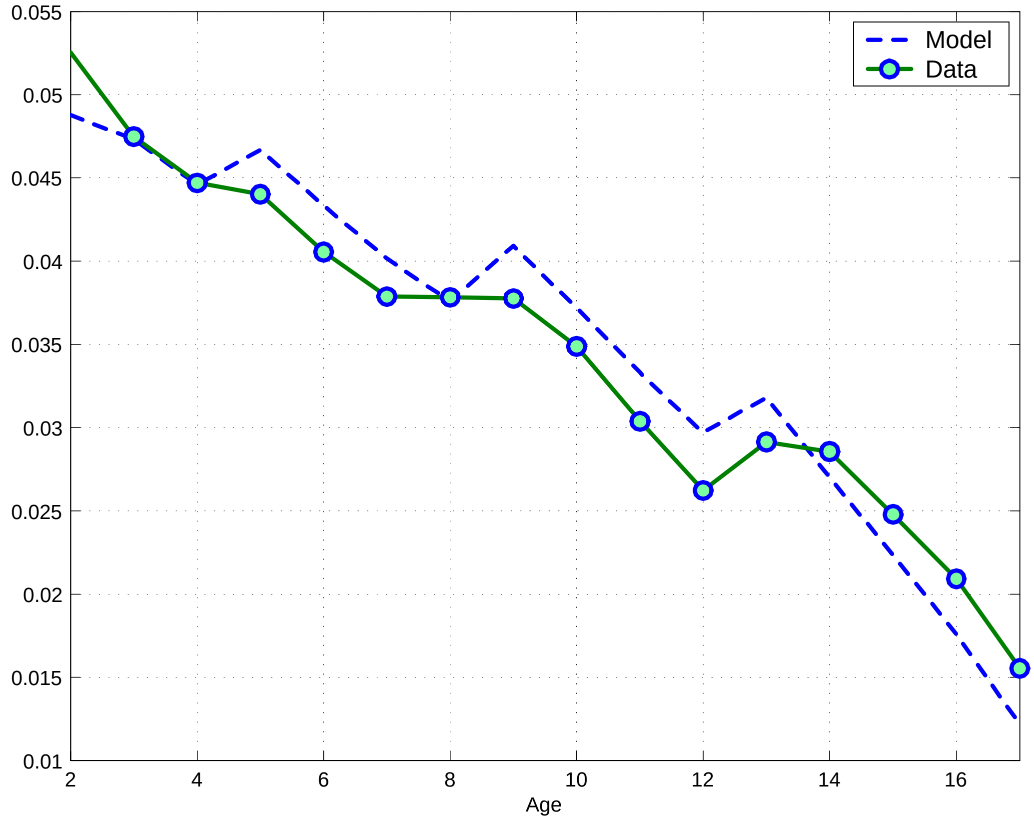
This section first discuss the fit and implications of the parameter estimates. Next, it quantifies what are the gains from trade in the market for patents, and the value of a patent.

6.1 Fit of the Model

An indicator of how the estimated model fits the data is to compare the empirical moments and the simulated moments from the model. We show that the model fits well the moments we try to match. In Table 4, we present the number of patents allowed to be expired as a percentage of traded and not traded active patents. We see that both empirical and simulated expiration rate are increasing as a function of the age of the patent. However, the model tends to predict a lower steepness in the moment that relates to the probability to be traded conditional on been active and having been previously.traded. The moment of the probability to be traded conditional on renewal and not having been previously traded is remarkably well fitted.

Moreover, Figure 3 and Figure 4 show respectively, the number of patents that are traded as a percentage of active previously traded and not previously traded patents. Figure 3 shows that the model is able to capture the decreasing shape of moments to be matched, including the discountinuous jumps immediately after renewal, and the sharp decrease when patent get closer to their expiration date. Nevertheless, the simulation generated by the model tend to overpredict the jumps of the transfer rate immediately after renewal. This feature might be as a result of the timing in which the trading decision occurs in the

Figure 3: The Proportion of Patents that are Traded Conditional on Renewal and Previously Traded

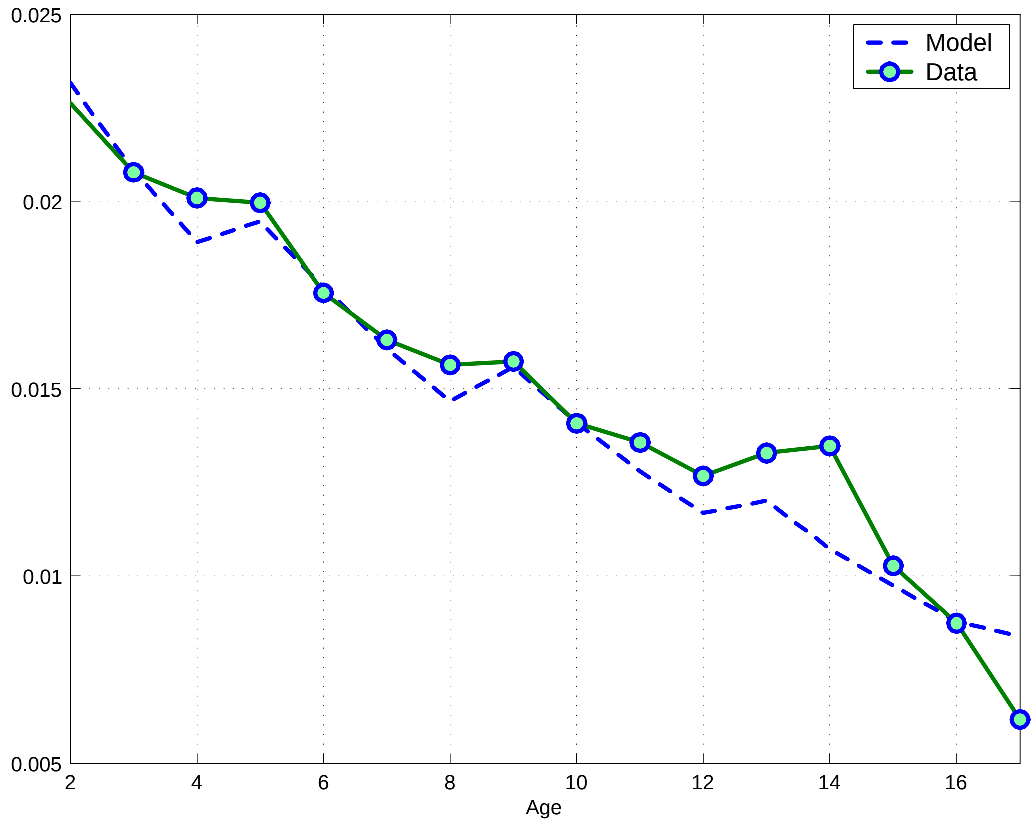


model. In the model, choices are made at the beginning of the period, while in reality these decisions take place continuously within a year. Figure 4 also shows that the model fits well the decreasing pattern of the data and the jumps of this statistic observe immediately after a renewal date.

An additional measure of how the estimated model fits the data is assessing the predictions of the model in other moments that the ones used in the estimation strategy. For instance, we can us the transfer rate, that is the the number of traded patents as a percentage of all active patents. For instance, Figure 5 show the fit of the model generating this moment and the one from the data. We see that the fit is good.

We can also look at the number of patents that are allowed to expire at a renewal date as a percentage of all active patents. In Table 5 we present the expiration rate in the data and the one generated by the model. We can also observe a good fit. Thus, the model does a remarkable job in fitting other moments that the ones we tried to match.

Figure 4: The Proportion of Patents that are Traded Conditional on Renewal and Not Previously Traded



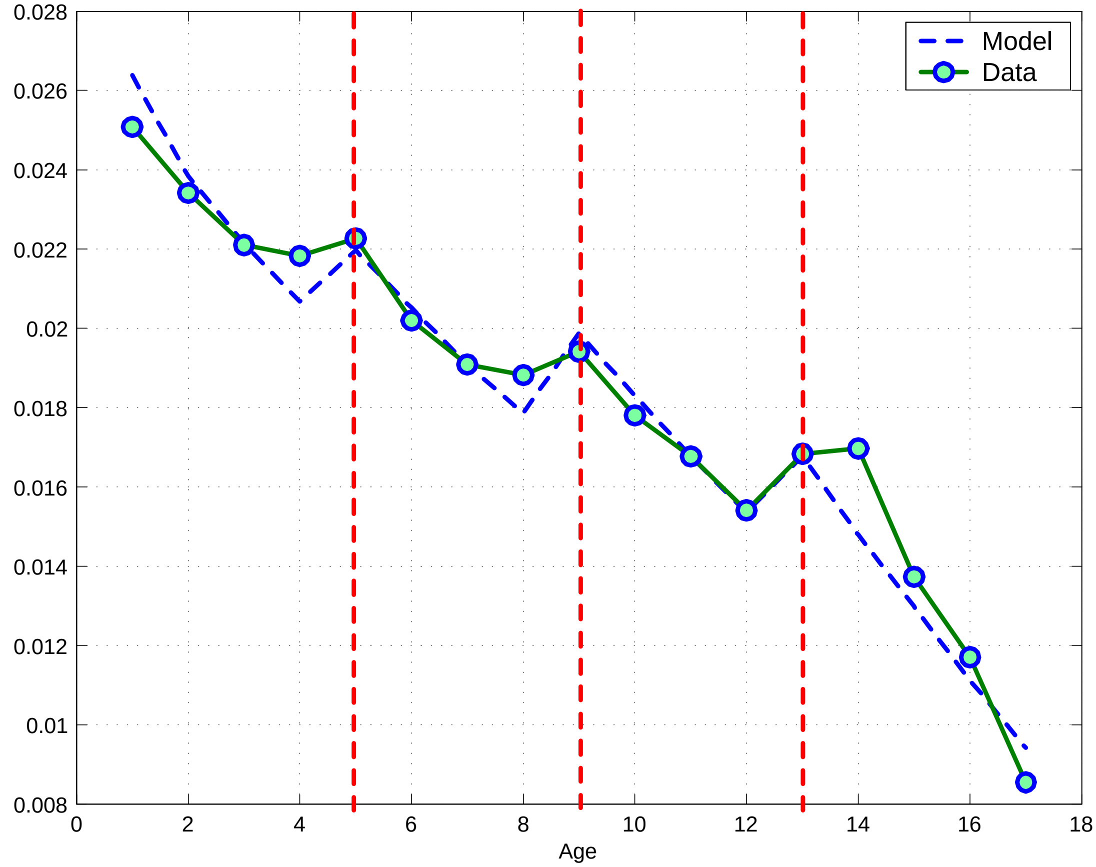
6.2 Discussion of Estimated Parameters

The parameter estimates are presented in Table 3. The estimates suggest the following.

Initial Returns The parameters μ and σ_R determine the initial distribution of returns. A high σ_R implies a large heterogeneity in quality among patents. A low μ implies that initially the quality or returns of patents are low. For instance, according to the estimated parameters, 15.06% of all patents have per period returns at age 1 below US$ (2003) 250, and the median patent has per period returns US$ (2003) 3559 at age 1.

Internal Growth of Returns The parameters δ, ϕ, σ determine the process of internal growth of returns. A high σ implies a higher probability of discovering new uses or successful applications that will increase the returns of the patent. The parameter ϕ measures the speed at which vanishes the learning possibilities. According to the estimates, the learning likelihood of internal growth of patent returns is exhausted by the end of the 4^{th} year of

Figure 5: The Proportion of Traded Patents Conditional on Renewal



age of a patent[22]. For instance, the following table uses the estimated parameters of the model to shows the predicted speed at which learning, vanishes. We define *learning* as the Pr $(g_a^i > \delta)$.

The estimate of the depreciation rate shows that when the learning opportunities are exhausted patents returns depreciate fast. Parameters δ and θ determine the full rate of depreciation. For instance, an estimate of $\delta = 0.92$ and $\theta = 0.97$ imply that the per period return of a patent depreciate in average at the rate of 11.64 % a year. Let us consider a patent with per period returns US$ (2003) 100,000 at age 6. The per period returns would in average depreciate offering approximately US$ (2003) 29,000 by age 16. This suggests that either competing technologies or imitation erode fast the profits from the protection of intellectual property.

[22]Pakes (1986) and Lanjouw (1998) report similar results for the case of German patents.

Table 5: Expiration Rate

Age	Data	Model
5	18.99	19.06
9	28.30	28.20
13	30.81	30.74

Table 6: Learning Likelihood and Patent Age

Patent Age	$\Pr(g_a^i > \delta)$
1	0.352
2	0.090
3	0.004
4	1.94E-06
5	3.64E-14
6	0

External Growth of Returns The parameter σ^e of an exponential distribution describes the process of potential gross growth of returns. A high σ^e implies both that is more likely that potential buyers are might be more efficient and that the heterogeneity among efficiency gains is also greater. The parameter is estimated to be $\sigma^e = 0.4133$. This suggests that the probability that in a given period a potential buyer has larger per period returns is 0.0890. However, since the cost of technology transfer is positive, then the decision whether a patent is transferred ultimately depends on both the current per period patent return of its owner and the cost of technology transfer.

Costs of technology transfer The estimate of the cost of technology transfer suggests is US$ (2003) 26,298. This estimate suggests that costs of adopting technologies developed by other firm require important expenditures, perhaps in personel, R&D, or restructuring firm's organization to efficiently use the acquired technology.

Random Trades The parameter ε accounts for patents recordations that are not as a result of the process of gains from trade and cost of technology transfer. The estimate of this parameter is 0.008, i.e., 0.8 percentage points of the proportion of active patents that are traded in every period cannot be explained with the specification of the transfer of technology that we have used. This accounts approximately for 30% of the trades of the patents of age 1, 34% at age 2, 37% at age 5, 44 at age 10, 48% at age 13, 59% at age 15.

Table 7: Distribution of the Value of Patents at Age 1

Percentile	Value (US$ 2003)	Cum % of Total
50	21,186	7.6
75	55,442	22.7
80	71,424	28.1
90	137,083	44.9
98.5	461,043	78.4
99.8	1,185,477	93.7
Mean value	57,910	

and 87% at age 17. Thus, since most of the patents are traded early in their life, the model can explain in average more than 60% of all patents traded.

Distribution of Citations Received The distribution of citations received, a Negative Binomial distribution, is characterized by the parameters (γ, p), where $\gamma(z)$ is a function of revenue that depends on parameters $\rho_1, ..., \rho_{17}, \lambda_1$. As explained before, given p, the parameters $\rho_5, ..., \rho_{17}$ can be pinned down by the mean of the citations received by expired patents. The parameters $\rho 1, ...\rho 4, \lambda_1$ and p are jointly estimated using the mean number of citations of previously and not previously traded patents and the mean number of citations of active patents at age 1.

Distribution of the Value of Patents In this section the parameter estimates are used to simulate the distributions of the value of patents and show how this distribution evolves as patents become older. These distributions were calculated by generating a 7.5 millions of patents with their respective initial returns, and consequentely following their returns using the estimated process of internal and external growth of returns taking into account whether patent were renewed or traded. The value of a patent is obtained as the discounted present value of its stream of patent returns at a given age.

The simulation show that the distribution of patent returns is very skewed as can be seen by estimated value of patents at their grant date, see Table 7. In particular, the value of the median patent is estimated to be US$ (2003) 21,186, but the bottom 50% of all patents only account for the 7.6% of the total value of all patents in the cohort.

In addition, the distribution of patent returns becomes more skewed as patents become older. The model predicts that low return patents are less likely to be traded and they mainly depreciate either at a rate δ or their returns become zero with probability $1 - \theta$.

Table 8: Distribution of the Value of Patents at Age 4

Percentile	Value (US$ 2003)	Cum % of Total
50	10,559	2.8
75	35,353	14.5
80	46,925	19.2
90	99,356	34.8
98.5	410,094	71.1
99.8	1,226,341	90.1
Mean value	43,350	

We find that patent with higher per period returns are more likely to be traded, so their per period returns in average depreciate less fast than are the ones that are not traded. This trading effect implies that the distribution of patent values is more skewed as patents are older. For instance, in Table 7 we show that the median patent at age 1 has a value of \$21,186 and the 99.8 percentile is \$1,185,477. However, in Table 8 we show that the median patent at age 4 is \$10,559, while the 99.8 percentile is 1,226,341. Thus, the distribution is more skewed.

Moreover, we find that the average present value at age 1 of a patent that is eventually traded is \$130,155, and the average value of one that is not traded ever is \$42,426. In terms of the average per period return at age 1 for traded and never traded patents, they are respectively 18,209 amd 6,280. Thus, traded patents in average are three times more valuable than their non-traded counterparts.

The next section focuses on the gains from trade in the market for patents.

6.3 Gains from Trade

Patents are traded because some firms are more productive than others in the use of a given patent. So, upon the transfer of a patent, gains from trade are realized. In the model, we define the gains from trade as the value of patent rights generated by the ability to transfer patents between firms.

We calculate the value of the gains from trade by running the following expirement. First, we use the above estimates and run simulations to obtain the average value of a patent at its grant date. We find that the average present value of a patent is US\$ (2003) 57,910. Next we shutdown the market for patents by setting the cost of technology transfer high enough so that no patent is traded as a result of technology transfer. Consequently, we find that the average value of a patent by its grant date and when no market exists is US$ (2003) 49,682. The gain from trade is the percentage gain in value as a result of the existence of the market, that is 14.21%[23].

This estimate is a lower bound of the gains from trade in the market for intellectual property. We assumed that the seller gets all the surplus in a patent transfer. Thus, we are underestimating all the gains trade that could occur if in an actual transaction the potential buyer accrues a benefit larger than zero. In addition, patents might be licensed too and licensing also results in gains from trade. We cannot account for these gains because there is not systematic data on licensing. Licensing transactions tend to be private agreements between firms.[24].

A statistic that is also of interest is to quantify what are the *gross* gains from trade, defined as those when the cost of technology transfer equal to zero. We find that the gross gains from trade are 26.82%. Given the assumptions previously discussed, this statistic is an upper bound of how much can be achieved by using policy addressed at reducing the cost of technology transfer to promote the transfer of technology. In particular, we show that reducing the cost of technology transfer by 50% implies that the gains from trade of the market would increase from 14.21% to 17.63%, this is 3.42 percentage points more, or a 24% increased.

7 Conclusion

This paper is the first systematic work that uses data on the transfer of patents to shed light on the workings of the market for intellectual property. It is a significant advancement because the previous literature on patents and intellectual property transfer have been hampered by a lack of systematic data on how intellectual property assets are traded. This study attains two objectives. First, it presents the stylized facts about the transfer of patents. Second, it develops and analyzes a model of patent trades that explains the stylized facts. The paper finds evidence that the market for trading patents increases efficiency, with patents rising in value as they are better matched with firms.

The data presented in this paper has been collected using recorded transfer of patents at the USPTO. The dataset compiled is a panel of patents with their histories of renewal and

[23]The gain from trade is calculated as $\frac{57{,}910-49{,}682}{57{,}910}$. Note that the gains from trade are net of taxes and the cost of technology transfer.

[24]Note that if the cost of technology transfer was equal to infinity, no licensing would occur either. Licensing a technology also involves a cost of adoption. What might be ocurring in the market is that the size of the cost of adopting a technology might determine whether technologies are either licensed or sold (assigned).

trading decisions over their life cycle. The paper moves on to present the patterns observed in the data. First, the number of patents traded as a percentage of all active patents monotonically decreases as a function of age except at the renewal dates. Immediately after renewal, this rate discontinuously increases. Second, the number of patents that are allowed to expire as a percentage of all active patents is increasing as a function of the renewal dates. Third, patents with higher number of total citations received by a given renewal dates are less likely to be allowed to expire. Fourth, among patents of the same age, patents with higher number of total citations received are more likely to be traded. Finally, among traded patents of the same age, and especially those recently traded, they are more likely to be retraded and less likely to be allowed to expire.

These facts and intrinsic characteristics of patents motivate a dynamic model of patent trades with costly technology transfer and gains from trade. In the model, there are two mechanisms. First, the cost of technology transfer creates a selection effect so that better patents are more likely to be traded. This explains the discontinuous increase of the transfer rate after the renewal decision, and the evidence that traded patents are more likely to be traded and less likely to expire. Second, there exists a horizon effect that explains that the transfer rate decreases as the patent gets closer to its expiration date. This is because the shorter horizon implies less time to amortize the cost of technology transfer. This accounts for the observed decreasing transfer rate over the life cycle of patents.

The parameters of the model are estimated using the simulated general method of moments. We find estimates that show that the market for patents accounts for 14.21% of the value of a patent. This number is a lower bound of the gains from trade in the market for intellectual property because we have assumed that the seller gets all the surplus, and because licensing opportunities have not been considered in this accounting.

This work opens new avenues of research. Perhaps most interesting would be to study the sources of innovation and to characterize who are the buyers and sellers of technology. In particular, to trace the flow of technology transfer, and to analyze whether small firms specialize in innovating and then selling their inventions to larger firms, which might have a comparative advantage in their management. Second, to evaluate to what extent the move toward higher protection of patent rights that occurred in the 1980's has facilitated specialization, and consequently trade in patents. Lastly, this work can also be extended to examine alternatives to promote innovation such as lower taxation on intellectual property transfer.

These questions have not been previously addressed empirically due to a lack of data on how intellectual property assets are traded.

References

[1] Acs, Z. and Audretsch, D. (1988), "Innovation in Large and Small Firms: An Empirical Analysis." *The American Economic Review*, Vol. 78, No. 4, 678-690.

[2] Arora, Fosfuri and Gambardella (2002), "Market for Technology: The Economics of Innovation and Corporate Strategy." The MIT Press. Cambridge, Massacheusetts.

[3] Arrow, K. (1962). "Economic Welfare and the Allocation of Resources for Invention." In the Rate and Direction of Inventive Activity. Special conference series NBER. Princeton, NJ: Princeton Univ. Press.

[4] Arrow, K. (1962A). "Comment on Case Studies. In The Rate and Direction of Inventive Activity. Special conference series NBER. Princeton, NJ: Princeton Univ. Press.

[5] Arrow, K. (1983). "Innovation in Large and Small Firms." In Entrepreneurship, edited by Joshua Ronen. Lexington, Mass.: Lexington Books.

[6] BLS (1979). *The Impact of Research and Development on Productivity Growth*. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics, Bulletin 2331.

[7] Griliches, Z. (1979). "Issues in Assesing the Contribution of R&D to Productivity Growth." *Bell Journal of Economics*, 10, 92-116.

[8] Griliches, Z. (1992). "The Search of R&D Spillovers." *Scandinavian Journal of Economics*, 94, 29-47.

[9] Hall, B., Jaffe, A. and Tratjenberg M. (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

[10] Hall, B., Jaffe, A. and Tratjenberg M. (2004). "Market Value and Patents Citations." *Rand Journal of Economics*, forthcoming 2004.

[11] Hall, B., Griliches, Z., Hausman, J. (1986). "Patents and R and D: Is There a Lag?" *International Economic Review*, 27, 265-283.

[12] Jaffe, A., Henderson R. and Tratjenberg M. (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*. Vol. CVIII, issue 3, p. 577.

[13] Holmes, T. and Schmitz, J. (1990). "A Theory of Entrepreneurship and Its Application to the Study of Business Transfers." *Journal of Political Economy*, vol. 98, no. 2.

[14] Holmes, T. and Schmitz, J. (1995). "On the Turnover of Business Firms and Business Managers." *Journal of Political Economy*, vol 103, no. 5.

[15] Lanjouw, J. O. (1998). "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value". *Review of Economic Studies*, vol 65.

[16] Lanjouw, J. O., Pakes, A. and Putnam, J. (1998), "How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data." *Journal of Industrial Economics.*

[17] Lamoreaux, N. and Sokoloff, K. (1997). "Inventors, Firms, and the Market for Technology: U.S. Manufacturing in the Late Nineteenth and Early Twentieth Centuries." NBER Working Paper H0098.

[18] Lamoreaux, N. and Sokoloff, K. (1999). "Inventive Activity and the Market for Technology in the United States, 1840-1920." NBER Working Paper 7107.

[19] Pakes, A. (1986). "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica*, vol. 54, no. 4

[20] Pakes, A. and Griliches, Z. (1984). "Patents and R&D at the Firm Level: A First Look." In *R&D, Patents and Productivity*, ed. by Z. Griliches. Chicago: University of Chicago Press.

[21] Schankerman M. and Pakes, A. (1986). "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period." *The Economic Journal,* Vol. 96: No. 384, pp. 1052-1076

[22] Serrano, C. (2005). "The Dynamics of the Transfer and Renewal of Patents." Manuscript, University of Toronto.

[23] Teece, D. J. (1977). "Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-how." *The Economic Journal*, Vol 87: 242-261.

[24] Teece, D. J. (1986), "

[25] Tratjenberg M., Henderson R. and Jaffe A. (1997). "University versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation and New Technology*, 5, 19-50.

[26] Williamson O.E. (1991). "Comparative Economic Organization -The Analysis of Discrete Structural Alternatives." Administrative Science Quarterly 36 (4): 269-296.

[27] WIPO (2003). "Intellectual Property: A Power Tool for Economic Growth." World Intellectual Property Organization (WIPO)

Appendix: Proofs and Tables

Proof of Lemma 1:

The proof is an extension based on Pakes (1986) results.

Proof of Proposition 2:

For convenience of notation, let us rewrite the value function $V(a,x,y)$ as $V_a(x,y)$. I want to show that $\widehat{g}_a(x,\tau)$ is weakly decreasing with respect x. I have divided the proof into two parts. The first one studies the case where $\widehat{g}_a(x,\tau)$ is defined as the external growth of returns that makes a firm indifferent between selling and allow the patent to expire. This result is straightforward. Formally the function $\widehat{g}_a(x,\tau)$ is defined as

$$V(a,\widehat{g}_a(x,\tau)x) = 0$$

From Lemma 1 we know that the value function is weakly increasing in per period returns of a patent. Let us suppose that x increases, it must be the case that $\widehat{g}_a(x,\tau)$ is decreasing with x to keep the equality holding.

The second part of the proof analyzes the subtle case in which $\widehat{g}_a(x,\tau)$ is defined as the external growth of returns that makes a firm indifferent between selling and keep the patent. Consider assumption G, that states the internal growth of returns is independent of the level. Patent returns evolve over time according to

$$\begin{aligned} g_a^i &= \frac{x'}{z} \\ g^e &= \frac{y'}{x'} = \frac{y'}{g^i z} \end{aligned}$$

where in the case of assumption G, the joint density function of the random variables g_a^i and g^e, defined as $f_a(g^x,g^y)$ depends upon a but not z.

The argument of the proof is by induction on the age of the patent. Let us start by considering the last period, $a = L$.

$$\begin{aligned} [\hat{g}_L(x,\tau) - 1]\, x &= \tau \\ \hat{g}_L(x,\tau) &= \frac{\tau}{x} + 1 \end{aligned}$$

which is decreasing in x.

Now, assume true for $a' > a$, I will show is also true for a.

The decision of whether to sell or keep relies on the following expression.

$$\begin{aligned}
& \widetilde{V}_a^S(x,y) - \widetilde{V}_a^K(x,y) \\
= \;& V_a(y) - \tau - V_a(x) \\
= \;& V_a(g_a^y x) - \tau - V_a(x) \\
= \;& (g_a^y - 1)\,x - \tau \\
& + \beta \left[E\left[\widetilde{V}_{a+1}(x',y') | a, g_a^y x \right] - E\left[\widetilde{V}_{a+1}(x',y') | a, x \right] \right]
\end{aligned}$$

It is sufficient to show that the above is weakly increasing in x. The first term is increasing in x. Now look at the second

$$\begin{aligned}
& E\left[\widetilde{V}_{a+1}(x',y') | a, g_a^e x \right] - E\left[\widetilde{V}_{a+1}(x',y') | a, x \right] \\
= \;& \int_{g_{a+1}^x} \int_{g_{a+1}^y} \left[\begin{array}{c} \widetilde{V}_{a+1}(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x) - \\ \widetilde{V}_{a+1}(g_{a+1}^i x, g_{a+1}^e g_{a+1}^i x) \end{array} \right] f_a(g^i, g^e) dg_{a+1}^i dg_{a+1}^e
\end{aligned}$$

For general a, recall that as an induction hypothesis we assumed

$$\widetilde{V}_{a+1}^S(x,y) - \widetilde{V}_{a+1}^K(x,y)$$

was weakly increasing in x. It suffices (given conditions on the joint density funtion of the growth of returns) to prove that the interior of the double integral is increasing in x.

$$\widetilde{V}_{a+1}(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x) - \widetilde{V}_{a+1}(g_{a+1}^i x, g_{a+1}^e g_{a+1}^i x)$$

There are four cases to study.

1. in K region with $(g_{a+1}^x x, g_{a+1}^e g_{a+1}^i x)$, in K region with $(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x)$

$$\begin{aligned}
& \widetilde{V}_{a+1}(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x) - \widetilde{V}_{a+1}(g_{a+1}^i x, g_{a+1}^e g_{a+1}^i x) \\
= \;& V_{a+1}(g_{a+1}^i g_a^e x) - V_{a+1}(g_{a+1}^i x)
\end{aligned}$$

Define $\lambda = g_a^e$. It suffices to show that the above is weakly increasing in x. The above is an increasing transformation of the induction hypotheses because λ is independent of $g_{a+1}^i x$. Then the above expression is increasing in x.

2. in K region with $(g_{a+1}^i x, g_{a+1}^e g_{a+1}^i x)$, in S region with $(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x)$

$$\begin{aligned}
& \widetilde{V}_{a+1}(g_{a+1}^i g_a^e x, g_{a+1}^e g_{a+1}^i g_a^e x) - \widetilde{V}_{a+1}(g_{a+1}^i x, g_{a+1}^e g_{a+1}^i x) \\
= \;& V_{a+1}(g_{a+1}^e g_{a+1}^i g_a^e x) - \tau - V_{a+1}\left(g_{a+1}^i x\right)
\end{aligned}$$

I want to show that the above in increasing in x. Let $\lambda = g^e_{a+1} g^e_a$. Since λ is independent of $g^i_{a+1}x$ and by the previous induction argument, then the above is weakly increasing in x.

Similarly, it can be shown for the remaining two cases.

3. in E region with $(g^i_{a+1}x, g^e_{a+1}g^i_{a+1}x)$, in E region with $(g^i_{a+1}g^e_a x, g^e_{a+1}g^i_{a+1}g^e_a x)$

4. in S region with $(g^i_{a+1}x, g^e_{a+1}g^i_{a+1}x)$, in S region with $(g^i_{a+1}g^e_a x, g^e_{a+1}g^i_{a+1}g^e_a x)$

This completes the proof.

Proof of Proposition 3:

I want to show that $\widehat{g}_a(x,\tau)$ is weakly increasing in a. Consider assumption G, that states the internal growth of returns is independent of the level. Patent returns evolve over time according to

$$\begin{aligned} g^i_a &= \frac{x'}{z} \\ g^e_a &= \frac{y'}{x'} = \frac{y'}{g^i_a z} \end{aligned}$$

where in the case of assumption G, the joint density function of the random variables g^i_a and g^e, defined as $f_a(g^x, g^y)$ depends upon a but not z.

For convenience of notation, let us rewrite the value function $V(a,x,y)$ as $V_a(x,y)$. To prove the proposition, it suffices to show that the difference $\widetilde{V}^S_a(x,y) - \widetilde{V}^K_a(x,y)$ is decreasing in a, so that $\hat{g}_a(x,\tau)$ is increasing in a.

$$\begin{aligned} & \widetilde{V}^S_a(x,y) - \widetilde{V}^K_a(x,y) \\ = \; & V_a(y) - \tau - V_a(x) \\ = \; & V_a(g^e_a x) - \tau - V_a(x) \\ = \; & (g^e_a - 1)\, x - \tau \\ & + \beta \left[E\left[\widetilde{V}_{a+1}(x',y')|a, g^e_a x\right] - E\left[\widetilde{V}_{a+1}(x',y')|a, x\right]\right] \end{aligned}$$

The argument of the proof is by induction. First, we start in the case of $a = L$ and $L-1$, that is the last and penultimate period of life of a patent

$$\begin{aligned} & \widetilde{V}^S_L(x,y) - \widetilde{V}^K_L(x,y) \\ = \; & x g^e_L - \tau - x \\ = \; & x(g^e_L - 1) - \tau \end{aligned}$$

For $a = L - 1$, it is

$$\begin{aligned}
& \widetilde{V}_{L-1}^{S}(x,y) - \widetilde{V}_{L-1}^{K}(x,y) \\
= \quad & x(g_{L-1}^{e} - 1) - \tau + \beta[E\left[\widetilde{V}_{L}(x',y')|L-1, g_{L-1}^{e}x\right] - E\left[\widetilde{V}_{L}(x',y')|L-1, x\right]]
\end{aligned}$$

There are three cases to study. The first case is the one in which the patent was kept in period $L-1$ as well as in period L. The second case considers the possibility of sale in period $L-1$ and being kept in period L. Finally, the third one analyzes the case that the patent is sold in both periods.

(1) K - K

$$\begin{aligned}
& x(g_{L-1}^{e} - 1) - \tau + \beta[E[\widetilde{V}_{L}^{K}(x',y')|L-1, g_{L-1}^{y}x] - E\left[\widetilde{V}_{L}^{K}(x',y')|L-1, x\right] \\
= \quad & x(g_{L-1}^{e} - 1) - \tau + \beta[E[V_{L}(g_{L}^{i}g_{L-1}^{e}x)] - E[V_{L}(g_{L}^{i}x)]] \\
= \quad & x(g_{L-1}^{e} - 1) - \tau + \beta\int[v_{L}(g_{L}^{i}g_{L-1}^{e}x) - v_{L}(g_{L}^{i}x)]f_{L}(g^{i},g^{e})dg^{i}
\end{aligned}$$

Taking as given x, g_{L-1}^{e}, and g_{L}^{e}, $\widetilde{V}_{a}^{S}(x,y) - \widetilde{V}_{a}^{K}(x,y)$ is decreasing in a because the integral is larger or equal than zero.

Similarly we can also show it for the next two cases.

For a general a,we know that $\left[E\left[\widetilde{V}_{a}(x',y')|a-1, g_{a-1}^{e}x\right] - E\left[\widetilde{V}_{a}(x',y')|a-1, x\right]\right] = 0$ for $a = L$. So, we have to show that

$$\{E\left[\widetilde{V}_{a+1}(x',y')|a, g_{a}^{e}x\right] - E\left[\widetilde{V}_{a+1}(x',y')|a, x\right]\} \rightarrow 0 \text{ uniformly}$$

Since we know that

$$\begin{aligned}
E\left[\widetilde{V}_{a}(x',y')|a-1, g_{a-1}^{e}x\right] \quad & \rightarrow \quad 0 \text{ as } a \text{ approaches } L \\
E\left[\widetilde{V}_{a}(x',y')|a-1, x\right] \quad & \rightarrow \quad 0 \text{ as } a \text{ approaches } L
\end{aligned}$$

and by proposition 1 we know that given a and $g^{e} > 1$ if a trade takes place

$$E\left[\widetilde{V}_{a}(x',y')|a-1, g_{a-1}^{e}x\right] \geq E\left[\widetilde{V}_{a}(x',y')|a-1, x\right]$$

Therefore $\widehat{g}_{a}(x,\tau)$ increases in a because $f_{a}(g^{i},g^{e})$ depends upon a but not on the return of the patent, z.

Finally, I show that the probability of being traded conditional on survival is weakly decreasing as a function of age. Since it is assumed that the process of arrival of offers, $F_{g^{e}}(u^{e})$, is independent of age, and we have already shown that $g_a(x,\tau)$ is weakly increasing in a, then the probability of being traded is weakly decreasing in a.

Proof of Lemma 4:

We want to show that $\left|\frac{\Delta V(a+1,x',y'|z)}{\Delta a}\right| \leq Q$, which rearranging is $|V(a+1,x',y'|z) - V(a,x',y'|z)| \leq Q$. It is sufficient to show that $(V(a,x',y'|z) - V(a+1,x',y'|z)) \leq Q$.

Let M be the maximum value that a return can achieve. It is finite because the distribution of the growth of internal and external returns are bounded above. Then

$$
\begin{aligned}
V(a,x',y'|z) &= z + \beta E(z') + \beta^2 E(z'') + ... + \beta^{17-a} E(.) \\
&\leq M + \beta M + \beta^2 M + ... + \beta^{17-a} M \\
&= M\frac{(1-\beta^{18-a})}{1-\beta}
\end{aligned}
$$

Let m be the minimum value that a return can achieve (i.e., considering that the process of internal growth of returns is bounded below by δ). So, we can show that

$$
\begin{aligned}
V(a+1,x',y'|z) &= z + \beta E(z') + \beta^2 E(z'') + ... + \beta^{17-a-1} E(.) \\
&\leq m + \beta m + \beta^2 m + ... + \beta^{17-a-1} m \\
&= m\frac{(1-\beta^{18-a-1})}{1-\beta}
\end{aligned}
$$

Therefore, it suffices for Q to be such that $M\frac{(1-\beta^{18-a})}{1-\beta} - m\frac{(1-\beta^{18-a-1})}{1-\beta} \leq Q$.

Proof of Proposition 5:

Given any patent return z_1 and z_2 such that $z_2 > z_1$, I want to show that

$$
\frac{\Delta[V(a,x',y'|z_2) - V(a,x',y'|z_1)]}{\Delta a} = \left\{ \begin{array}{ll} > 0 & \text{if} \quad a < a^* \\ \leq 0 & \text{if} \quad a \geq a^* \end{array} \right\}
$$

It is sufficient to examine the sign of the relationship between the slopes of the option value of a patent, that is the following expression.

$$
\left(\frac{\Delta EV(a+1,x',y'|z_2)}{\Delta a} - \frac{\Delta EV(a+1,x',y'|z_1)}{\Delta a} \right)
$$

Without loss of generality, I focus on the case in which the holder of the patent chooses to keep it. Similarly, we can also show the result for the case in which the patent is sold.

$$
\begin{aligned}
&\left(\frac{\Delta EV(a+1,x'|z_2)}{\Delta a}-\frac{\Delta EV(a+1,x',z_1)}{\Delta a}\right)\\
=\;&\frac{\Delta V(a+1,\delta z_2)}{\Delta a}F_{g_i}(\delta;a,z_2)-\frac{\Delta V(a+1,\delta z_1)}{\Delta a}F_{g_i}(\delta;a,z_1)\\
&+V(a+1,\delta z_2)\frac{\Delta F_{g_i}(\delta;a,z_2)}{\Delta a}-V(a+1,\delta z_1)\frac{\Delta F_{g_i}(\delta;a,z_1)}{\Delta a}\\
&+\int_{\delta^+}^{\infty}[\frac{\Delta V(a+1,u^i z_2|z_2)}{\Delta a}f_{g_i}(u^i;a,z_2)-\frac{\Delta V(a+1,u^i z_1|z_1)}{\Delta a}f_{g_i}(u^i;a,z_1)]\mathrm{d}u^i\\
&+\int_{\delta^+}^{\infty}[V(a+1,u^i z_2|z_2)\frac{\Delta(f_{g_i}(u^i;a,z_2)}{\Delta a}-V(a+1,u^i z_1|z_1)\frac{\Delta(f_{g_i}(u^i;a,z_1)}{\Delta a}]\mathrm{d}u^i
\end{aligned}
$$

where f_{g_i} is the density function of F_{g_i}.

The first term is always negative. To show this, first notice that $\frac{\Delta V(a+1,\delta z_i)}{\Delta a}<0$ for $i\in\{1,2\}$. It was assumed that $F_{g_i}(\delta;a,z_2)>F_{g_i}(\delta;a,z_1)$. Using the particular case in which F_{g_i} is independent of z we can show that $\frac{\Delta V(a+1,\delta z_2)}{\Delta a}<\frac{\Delta V(a+1,\delta z_1)}{\Delta a}$ always holds. Therefore, the first term is negative.

We can also show that the sign of the second term is positive. By Lemma 1 we know that the value function is increasing in the patent per period returns, so $V(a+1,\delta z_2)>V(a+1,\delta z_1)$. Also we have assumed that the probability of "no learning" (i.e., $g_a^i=\delta$) is increasing in a, that is equivalent to $\frac{\Delta F_{g_i}(\delta;a,z_2)}{\Delta a}>\frac{\Delta F_{g_i}(\delta;a,z_1)}{\Delta a}$. The term is clearly positive.

The sign of the third term is ambiguos. On the one hand, we know that $\frac{\Delta V(a+1,u^i z_2)}{\Delta a}<\frac{\Delta V(a+1,u^i z_1)}{\Delta a}$, however $f_{g_i}(u^i;a,z_2)<f_{g_i}(u^i;a,z_1)$. This result is not surprising. For instance, in the case in which the the process F_{g_i} is independent of z the equivalent of this term would be always negative. Introducing the assumption for which learning is less likely for patents with higher returns makes the expected value of the potential buyer smaller than in the case of independence.

Since we assumed that $\frac{\Delta(f_{g_i}(u^i;a,z_2))}{\Delta a}>\frac{\Delta(f_{g_i}(u^i;a,z_1))}{\Delta a}$, then the sign of the fourth term is positive.

Next, we have to show that for sufficiently small a, the difference between the slopes of the option value for the potential buyer and seller, $\frac{\Delta EV(a+1,x'|z_2)}{\Delta a}-\frac{\Delta EV(a+1,x',z_1)}{\Delta a}$, increases with age. To do so, it is sufficient to show that the second and fourth terms are larger than the first and third term. The strategy is to construct a bound for the term $\frac{\Delta V(a+1,x',y'|z)}{\Delta a}$. By Lemma 2 we know that $\left|\frac{\Delta V(a+1,x',y'|z)}{\Delta a}\right|\leq Q$, where Q is a number sufficiently small and positive. Therefore, this bound permit us to make the first and third term small enough to compare to the other two terms. There exists a a^* such that $a<a^*\leq\overline{a}$ that the second and fourth term dominate the first and the third one. If $a\geq a^*$, then the last three terms become monotonically small converging

to zero when learning vanishes at age $a = \overline{a}$ (i.e., the probability of learning is nul) (i.e., this is a particular case of Gibrat's law, see Proposition 2).

Proof of Proposition 6:

The argument of the proof is as follows. It is assumed that the returns due to internal growth of returns are subject to first order stochastically dominance in z. In other words, the higher today's return is, the more likely it is that the return of tomorrow will be high. Now, consider the problem of whether to sell a patent in period L, that is the last period of life of a patent. A patent is sold if $y \geq x + \tau$. Consider now a period before the last, which is $L - 1$. A patent is now sold if the following condition holds

$$y + \beta E(y'|y) \geq x + \beta E(x'|x) + \tau$$

Rearranging this condition, we obtain that

$$y \geq x + \beta[E(x'|x) - E(y'|y)] + \tau$$

However $\beta[E(x'|x) - E(y'|y)] \leq 0$ because of first order stochastically dominance. Then, it must be the case that $\widehat{g}_{L-1} \leq \widehat{g}_L$. The proof can be extended backwards by any number of finite periods. Therefore, $\widehat{g}_a(.)$ is weakly increasing in a.

Proof of Proposition 7:

I want to show that $\widehat{x}_a(\tau)$ is increasing in a. First I show that for any age a there exists a unique $\widehat{x}_a(\tau)$ for which firms are indifferent between keeping and discontinuing a patent. So, $\widehat{x}_a(\tau)$ is defined as $V(a, \widehat{x}_a(\tau)) = 0$. It is assumed that the schedule of renewal fees c_a, when positive, is increasing with age. It is also assumed in the paper that the probability that tomorrow's return is larger than a given number u is weakly decreasing with age. Given that, Lemma 1 demonstrates that the option value function of a patent, $E\widetilde{V}(a + 1, x', y'|a, z)$ is weakly decreasing in a. Therefore, the returns that make a firm indifferent between keeping and discontinuing, $\widehat{x}_a(\tau)$, must be increasing as age increases.

Table 9: Summary Statistics of the Unbalanced Panel

Age of Patent (Years)	Active Patents	Not Yet Traded	Already Traded
A. Number of Patents (Small Innovators)			
1	453,683	442,303	11,380
2	417,372	397,124	20,248
3	382,367	355,802	26,565
4	349,013	317,856	31,157
5	258,655	229,154	29,501
6	236,443	205,712	30,731
7	214,895	184,012	30,883
8	194,289	163,847	30,442
9	123,381	101,128	22,253
10	108,948	88,102	20,846
11	94,978	75,795	19,183
12	81,378	64,122	17,256
13	46,095	35,288	10,807
14	36,894	27,968	8,926
15	29,138	21,973	7,165
16	21,512	16,109	5,403
17	14,977	11,110	3,867

Table 10: Summary Statistics of the Unbalanced Panel (All innovators)

Age of Patent (Years)	Active Patents	Not Yet Traded	Already Traded
B. The Size of Owner of a Patent at grant date: Median (and 25 percentile)			
1	51 (5)	51 (5)	17.5 (2)
2	48 (5)	50 (5)	14 (2)
3	49 (5)	52 (5)	15 (3)
4	48 (5)	51 (5)	14 (3)
5	35 (4)	38 (4)	11 (2)
6	37 (4)	41 (4)	11 (2)
7	33 (4)	36 (4)	10 (2)
8	35 (4)	39 (4)	12 (2)
9	35 (4)	39 (4)	11.5 (2)
10	34 (4)	36 (4)	11 (2)
11	32 (3)	36 (4)	10 (2)
12	29 (3)	33 (3)	10 (2)
13	29 (3)	33 (4)	10 (2)
14	24 (3)	27 (3)	9 (2)
15	28 (3)	30 (4)	14 (2)
16	25 (3)	27 (3)	14 (2)
17	28 (3)	30 (4)	20 (3)

Table 11: Summary Statistics of the Unbalanced Panel (All innovators)

Age of Patent (Years)	Active Patents	Not Yet Traded	Already Traded
C. The Number of Patents: (All innovators)			
1	1,586,346	1,559,776	26,570
2	1,443,092	1,395,121	47,971
3	1,309,045	1,245,627	63,418
4	1,179,492	1,105,492	74,000
5	914,169	843,106	71,063
6	831,048	756,591	74,457
7	750,597	675,343	75,254
8	677,442	602,549	74,893
9	451,399	395,089	56,310
10	396,787	343,052	53,735
11	344,790	294,836	49,954
12	295,563	249,988	45,575
13	170,879	141,532	29,347
14	135,886	111,166	24,720
15	108,620	87,977	20,643
16	80,122	64,202	15,920
17	57,034	45,340	11,694

Pairwise-Difference Estimation of a Dynamic Optimization Model*

Han Hong
Dept. of Economics
Duke University

Matthew Shum†
Dept. of Economics
Johns Hopkins University

This draft: December 7, 2004
Comments welcome

Abstract

We develop a new estimation methodology for a dynamic optimization model with unobserved state variables. We propose a pairwise-difference approach which exploits two common features of the dynamic optimization problem we consider: (1) the weak monotonicity of the agent's decision (policy) function in the unobserved state variable, conditional on the observed state variables; and (2) the state-contingent nature of optimal decision-making which implies that, conditional on the observed state variables, the variation in observed choices across agents must be due to randomness in the unobserved state variable across agents. We illustrate our estimation procedure by estimating a dynamic trading model for the milk production quota market in Ontario, Canada.

Keywords: Pairwise-difference Estimators, Dynamic Optimization Models
JEL: C13, C50

---

*We thank Phil Cairns at the Dairy Farmers of Ontario for providing the data and patiently answering our questions. We thank J. Adda, V. Aguirregabiria, P. Bajari, L. Benkard, R. Blundell, B. Honore, H. Ichimura, S. Khan, C. Meghir, E. Miravete, S. Ng, A. Pakes, M. Pesendorfer, G. Ridder, L. Roberts, F. Wolak, and seminar participants at Iowa, Princeton, Rochester, Stanford, UCL, USC, the Cowles Conference on Structural Estimation at Yale University, and the 2001 Econometric Society Meetings at the University of Maryland for helpful discussions. We are grateful to the NSF (SES-0079495, SES-0003352) and SSHRC for financial support.

†Corresponding author: Matthew Shum, Dept. of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218; mshum@jhu.edu (e-mail).

1 Introduction

In this paper, we propose a new estimation methodology for a dynamic optimization model with preference and/or payoff shocks which are unobserved to the econometrician (but observed by agents when they make their dynamic choices). The two-step estimator we propose relies on two common features of the dynamic optimization problem we consider. First, we exploit the monotonicity of the agent's decision (policy) function in the unobserved shocks, conditional on the observed state variables. Second, we exploit the state-contingent nature of optimal decision-making which implies that, conditional on the observed state variables, the variation in observed choices across agents must be due to randomness in the shocks across agents.

The two-step pairwise-difference estimator we propose represents a new approach to estimating continuous-discrete choice dynamic models. To our knowledge, our approach represents the first application of pairwise-differencing methods, which have primarily been used in cross-sectional contexts (cf. Honore and Powell (1994)), to structural dynamic optimization problems. It complements the existing literature on identification and estimation in discrete-choice dynamic optimization models (cf. Pakes and Simpson (1989), Hotz and Miller (1993), Taber (2000), Magnac and Thesmar (2002)).

Our approach is related to some recent work which exploits monotonicity assumptions to identify and estimate structural equations. Earlier, Olley and Pakes (1996) exploited such an assumption in order to invert out the unobserved shock (or unobserved state variable) to derive a semiparametric estimator for production functions with serially correlated unobservables. Matzkin (2003) also exploited the quantile invariance implication of monotonicity to estimate nonparametrically functions which are nonlinear in the error term. Bajari and Benkard (2001) also used this principle in their study of hedonic discrete choice models of demand for differentiated products.

One advantage of our approach over alternative methods for estimating continuous choice dynamic optimization models, such as Euler Equation-based methods, is in accommodating unobserved state variables. Generally, conventional Euler Equation-based estimation methods have difficulties accommodating unobserved state variables.[1] Our approach also

[1]This is because the estimating moment conditions are derived from the rational expectations implication that deviations between predicted and observed actions are orthogonal to any information available at time t, which includes all state variables which affect an agent's period t choice. Therefore, to form the sample analogs of these orthogonality conditions, the econometrician needs to know the value of all the state variables at times t and $t+1$. See Pakes (1994) (pp. 188–189) for a more thorough discussion.

accommodates dynamic optimization models in which agents' choices are both continuous and discrete, for which conventional Euler-Equation methods are either not applicable or difficult.

This paper is also related to a recent literature on the identification and estimation of dynamic game models (e.g., Pesendorfer and Schmidt-Dengler (2003), Aguirregabiria and Mira (2003), Berry, Ostrovsky, and Pakes (2004), Bajari, Benkard, and Levin (2004)). While we do not focus on dynamic games here, one contribution that we make is the consideration of situations where agents have both continuous action spaces and continuous state spaces.

The plan of the paper is as follows. In the next section, we present a single-agent dynamic optimization problem and state our model assumptions. We describe our two-step estimation approach in Section 3. In Section 4, we illustrate our methodology by estimating a dynamic model of trading behavior in monthly exchanges operated by provincial regulatory agencies in Ontario, Canada to allocate milk production quotas across milk farmers. We conclude in section 5.

2 Empirical framework

Consider the following dynamic optimization problem of an agent i:

$$\max_{\{q_{it}\}_t} E\left[\sum_{t=0}^{\infty} \beta^t U\left(x_{it}, s_{it}, q_{it}; \theta\right) \mid \{q_{it}\}_t\right] \tag{1}$$

subject to

$$x_{it+1} = x_{it} + q_{it}; \quad s_{it} \sim F_s(\cdot; \gamma). \tag{2}$$

For example, consider an investment model where x_{it} can be interpreted as a stock and the control q_{it} as investment, or incremental additions to the stock which can be purchased at some fixed price. (For convenience, we will sometimes refer to x as the "stock" and q as "investment" in this paper, in reference to this example.)

$U(\cdots; \theta)$ is a per-period utility function, parameterized by the parameter vector θ. The per-period utility depends on the current stock x_{it} and an idiosyncratic variable s_{it}, which is known to agent i before he makes his choice of q_{it}. x_{it} is the state variable of the dynamic model, and s_{it} represents shocks to payoffs and/or preferences. Throughout the paper, we use the term "shock" to refer to the s_{it}, which we assume is observed by the optimizing agent at the time she makes her period t decision, but not by the econometrician.[2] Moreover, we will sometimes use the alternate terminology "unobserved state variable" (which was also used in Pakes (1994)) to refer to the s_{it} variables. While the state variable x_{it} is observed, the shock s_{it} is not. Therefore, the presence of the unobserved shock s_{it} induces randomness in the observed choices of the control q.

The shock distribution takes a parametric form $F_s(\cdot;\gamma)$, which is known up to the parameters γ, which are to be estimated. We assume:

Assumption 1 *The shocks* $s_{it} \sim F_s(\cdot;\gamma)$ *are* i.i.d. *across both agents* i *and periods* t.

While this independence assumption rules out the important case of serial correlation in the unobserved shocks over time (arising perhaps from unobserved agent-specific fixed effects), it is a common assumption made in the literature on estimation of dynamic models. On the other hand, it is straightforward to extend the *i.i.d.* assumption to one where where heterogeneity in the distribution of the shock s_{it} across agents and time is explicitly parameterized to depend on observed conditioning covariates.

Assuming stationarity, the agent's optimal policy function can be expressed as the maximizer of Bellman's equation: for $t = 1, 2, 3, \dots,$

$$\begin{aligned} q\left(x_{it}, s_{it}; \theta, \gamma\right) &= \operatorname{argmax}_q \left\{ U\left(x_{it}, s_{it}, q; \theta\right) + \beta \mathcal{E}_{x_{t+1}, s_{t+1} | x_t, s_t, q} V\left(x_{t+1}, s_{t+1}; \theta, \gamma\right) \right\} \\ &= \operatorname{argmax}_q \left\{ U\left(x_{it}, s_{it}, q; \theta\right) + \beta \mathcal{E}_{s_{t+1} | x_t, s_t, q} V\left(x_t + q_t, s_{t+1}; \theta, \gamma\right) \right\} \end{aligned} \tag{3}$$

where:

$$V\left(x_{t+1}, s_{t+1}; \theta, \gamma\right) \equiv \max_{\{q_{i\tau}\}_\tau} \mathcal{E}\left[\sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} U\left(x_{i\tau}, s_{i\tau}, q_\tau; \theta\right) \mid \{q_{i\tau}\}_\tau, x_{t+1}, s_{t+1} \right]. \tag{4}$$

In what follows, we simplify notation by defining

$$\mathcal{V}\left(x_{it} + q_{it}; \theta, \gamma\right) \equiv \int V\left(x_{it} + q_{it}, s; \theta, \gamma\right) F_s(ds; \gamma),$$

the ex ante value function at time t, where the expectation is over s_{t+1}, the future realization of the shock.

[2] This usage differs from the macroeconomic literature, where a "shock" is often unobserved by both the econometrician as well as the optimizing agent when she makes her decision.

2.1 Monotonicity and Quantile Invariance

We assume that the policy functions are monotonic in the unobserved state variable, conditional on a particular value for the observed state variable.

Assumption 2 *The policy functions $q(x_t, s_t; \theta, \gamma)$ are nondecreasing in s_t, conditional on x_t.*

Corollary 1 *Given Assumption 1, a sufficient condition for Assumption 2 is that U is supermodular in (q, s), for all x.*

Proof: The optimal policy q is given by

$$\text{argmax}_q \bar{U}(x, s, q) \equiv \{U(x, s, q; \theta) + \beta \mathcal{V}_{t+1}(x + q; \theta, \gamma)\}. \tag{5}$$

In order for $q(s, x; \theta, \gamma)$ to be non-decreasing in s given x, we require $\bar{U}(x, s, q; \theta)$ to be supermodular in (q, s), for all x. This is equivalent to supermodularity of $U(x, s, q; \theta)$ in (q, s) given x, because the expected continuation value function $\mathcal{V}(x + q; \theta, \gamma)$ does not depend on s, from Assumption 1.

An important implication of Assumption 2 is **quantile invariance**: conditional on x_t, the τ-th quantile of q conditional on x_t is $q(x_t, s_\tau; \theta, \gamma)$, where s_τ is the τ-th quantile of $F_s(\cdot)$. This implication of monotonicity was also exploited by Matzkin (2003) in her nonparametric estimation methodology for non-additive (in the error term) random functions.

The independence assumption that the distribution function F_s does not depend on x allows us to accommodate situations (such as atoms in $F(q|x)$) where we only have *weak* monotonicity of q in s, given x, so long as, for every quantile $\tau \in [0, 1]$, there exists an x for which $(F)^{-1}_{q|x}(\tau|x)$ is a singleton. This allows the investment decision to be a mixed discrete-continuous choice variable, with a point mass at zero (indicating no investment). This accommodates models of non-convex adjustment costs (cf. Eberly (1994)), and is appropriate for the empirical illustration we consider below.[3]

3 Estimation approach

The parameters we wish to estimate are θ and γ, respectively the utility function and shock distribution parameters. To simplify notation, we assume that our data are a balanced

[3]Elsewhere (Hong and Shum (2004)), we consider the case where the control function is continuously differentiable, in which case a stochastic Euler Equation approach could be employed for estimation.

panel: $\{q_{it}, x_{it}\}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$. This is not critical, as our estimator also applies to cases where the number of cross-sectional observations differs across time periods.

From the data, we can estimate the empirical distribution of q given x for each x. Denote each element of this family of distributions (indexed by x) by $\hat{F}(q|x)$. Therefore, $\hat{F}(q_{it}|x_{it})$ denotes the estimated conditional probability of $q \leq q_{it}$, conditional on the observed state variable being equal to x_{it}.

Since the conditioning variable x is continuous, we employ a kernel estimator for these conditional CDFs:

$$\hat{F}(q|x) = \frac{\frac{1}{T}\frac{1}{N}\frac{1}{h}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{1}\left(q_{it} \leq q\right) K\left(\frac{x - x_{it}}{h_N}\right)}{\frac{1}{T}\frac{1}{N}\frac{1}{h}\sum_{t=1}^{T}\sum_{i=1}^{N} K\left(\frac{x - x_{it}}{h_N}\right)}$$

where $K(\cdot)$ is a kernel weighting function and h_N is a bandwidth sequence. In computing the empirical CDF's, we employ all the observations, including those for which $q = 0$ (i.e.,, for which the agent remained at a corner solution and investment is zero).

We make the following assumptions on the kernel function:

Assumption 3 *1. $K(\cdot)$ is a r-th order kernel (with $r \geq 2$) function: (i) $\int K(u)\,du = 1$; (ii) $\int u^{\xi} K(u)\,du = 0$ for $\xi = 1, \ldots, r-1$; and (iii) $\int u^{r} K(u)\,du < \infty$.*

2. As $N \to \infty$, the bandwidth sequence (i) $h \to 0$; (ii) $\frac{Nh}{\log N} \to \infty$; and (iii) $\sqrt{N}h^{r} \to 0$.

Furthermore, we also require smoothness assumptions on the per-period utility function:

Assumption 4 *The functions $U(x, s, q; \theta)$ and $\frac{\partial}{\partial\theta}U(x, s, q; \theta)$ have continuous partial derivatives in (x, s, q) of order $r+1$ (where r is the order of the kernel from the previous step). The expectations of all derivatives of order up to $r+1$ exist.*

Condition 3(iii) above, along with the higher-order assumption on the kernel, are standard conditions for eliminating the asymptotic bias in the kernel estimates. Assumption 4 ensures that the asymptotic bias of the limit pairwise-differencing estimating function (described below) can be approximated up to the r-th order (as in Powell, Stock, and Stoker (1989)). Next, we describe our proposed two-step estimation approach.

3.1 First step: A Pairwise-differencing of First-order conditions

We can derive estimates of γ, the parameters of the shock distribution, as well as a subset of the parameters θ in the utility function, by exploiting the first-order condition of the maximization problem in Eq. (3).[4] This step exploits the state-contingent nature of optimal decision-making which implies that, conditional on the observed state variables, the variation in observed choices across agents must be due to randomness in the unobserved state variables across agents.

First, the deterministic accumulation nature of stock evolution process implies that for an agent i who invests a positive amount $q_{it} > 0$, we can rewrite her maximization problem as

$$q_t\left(x_{it}, s_{it}; \theta, \gamma\right) = \text{argmax}_q \left\{U\left(x_{it}, s_{it}, q; \theta\right) + \beta \mathcal{V}_{t+1}\left(x_{it} + q; \theta, \gamma\right)\right\}. \tag{6}$$

The first-order condition for this problem is

$$U_3\left(x_{it}, s_{it}, q_{it}; \theta\right) + \beta \mathcal{V}_t'\left(x_{it} + q_{it}; \theta, \gamma\right) = 0 \tag{7}$$

where $U_3(\cdots)$ refers to the derivative of $U(\cdots)$ with respect to its third argument. For any pair of agents i and j in period t such that $x_{it} + q_{it} = x_{jt} + q_{jt}$,

$$\mathcal{V}_t'\left(x_{it} + q_{it}; \theta, \gamma\right) = \mathcal{V}_t'\left(x_{jt} + q_{jt}; \theta, \gamma\right).$$

Hence we can condition on such pairs of agents in order to control for the unknown form of the expected value function.

Second, from the quantile invariance Assumption 2 and the assumption that s is distributed independently of x, we know that any individual i with a (q_{it}, x_{it}) pair must have received a shock s_{it} equal to $F_s^{-1}\left(\hat{F}(q_{it}|x_{it}); \gamma\right)$, the $\hat{F}(q_{it}|x_{it})$-th quantile of the shock distribution. This suggests that the cross-sectional variation in q given x for a collection of quantiles allows us to recover the corresponding quantiles of F_s, and hence estimate the γ parameters.

The considerations above immediately suggest a pairwise difference estimator for θ_1 and γ. Consider a pair of individuals i and j in period t with the same $x_{it} + q_{it} = x_{jt} + q_{jt}$. If we subtract the first-order conditions for these two observations, we obtain

$$\left\{U_3\left(x_{it}, s_{\hat{F}(q_{it}|x_{it})}, q_{it}; \theta_1\right) - U_3\left(x_{jt}, s_{\hat{F}(q_{jt}|x_{jt})}, q_{jt}; \theta_1\right)\right\} = 0, \tag{8}$$

[4] Recently, Berry and Pakes (2000) also exploit the first-order condition to derive estimates of structural parameters for models of *multi-agent* dynamic games. While we restrict our attention to single-agent problems, we focus on accommodating unobserved state variables, which are not present in the models considered by Berry and Pakes.

where $s_\tau \equiv F_s^{-1}(\tau;\gamma)$, the τ-th quantile of F_s.

Let θ_1 denotes the subset of the parameters θ which enter Eq. (8). Precisely, θ_1 is the subset of the parameters θ which are not eliminated by either (i) taking the derivative of the utility function U with respect to its third argument; (ii) taking the difference of the utility function derivative U_3 between any two individuals. The remaining parameters $\theta_2 \equiv \{\theta \setminus \theta_1\}$ will be estimated in the second step of our procedure.

The pairwise-difference estimator of (θ_1, γ) takes the following form:

$$\min_{\theta_1,\gamma} \frac{1}{N^2}\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{1}{h_{1,N}}K\left(\frac{(x_{it}+q_{it})-(x_{jt}+q_{jt})}{h_{1,N}}\right)\cdot\right.\right. \tag{9}$$
$$\left.\left.\mathbf{1}(q_{it}\neq 0)\cdot\mathbf{1}(q_{jt}\neq 0)\cdot\left[U_3(x_{it},\hat{s}_{it},q_{it};\theta_1)-U_3(x_{jt},\hat{s}_{jt},q_{jt};\theta_1)\right]^2\right\}\right]$$

where $\hat{s}_{it} \equiv \hat{F}_s^{-1}\left(\hat{F}(q_{it}|x_{it};\gamma);\mathcal{S}\right)$. The kernel function $K(\cdot)$ and bandwidth sequence $\{h_{1,N}\}$ obey Assumption 3 above. Moreover, in computing the objective function (9) above, we only include observations with non-zero investment ($q \neq 0$) because only for these observations is the first-order condition (7) satisfied.

Given an estimate $\hat{\gamma}$ of the parameters in the shock distribution function, we can immediately derive an estimate of the optimal policy function

$$\tilde{q}_t(x,s) \equiv \hat{F}_{q|x}^{-1}\left(\hat{F}_s(s;\hat{\gamma})\right),\ \forall s. \tag{10}$$

Our estimate of the period t investment choice q_t at a given state (x,s) is just the $F_s(s;\hat{\gamma})$-th quantile of $\hat{F}(q|x)$, the empirical conditional distribution of q given x.

Asymptotic theory for first-step Ahn and Powell (1993) and Honore and Powell (1994) pioneered the use of pairwise-differencing methods in econometrics. The objective function (9) resembles a weighted least squares objective, where each pair of observations is weighted by a kernel function which takes on small values when certain features of the pair of observations are very far apart. The asymptotic normality of our first-step estimates of θ_1 and γ is given in the following theorem, the proof of which is in the Appendix, Section A.1:

Theorem 1 *Let $\hat{\psi} \equiv \left(\hat{\theta}_1, \hat{\gamma}\right)$, the maximizers to (9). Then, given Assumptions 2—4, as well as assumptions 5 and 6 in the appendix,*

$$\sqrt{N}\left(\hat{\psi}-\psi\right) \xrightarrow{d} N\left(0, A^{-1}\Omega A^{-1}\right)$$

as $N \to \infty$, where A and Ω are defined in Eqs. (23) and (29) in the Appendix.

3.1.1 Remarks

Our econometric framework is parametric, in the sense that both the utility function and shock distributions are assumed to be of known parametric form. In principle, the shock distribution F_s can be given a very flexible parametric form. In our empirical work below, we also consider a flexible piecewise-linear specification for F_s, as follows:

Let $s_k \equiv F_s^{-1}(\tau_k)$ denote the τ_k-th quantile of the shock distribution F_s. Let κ denote the total number of quantiles to be estimated (and the corresponding quantile values by $\tau_1 < \tau_2 < \cdots < \tau_\kappa$). For any fixed κ, we approximate the distribution of the shocks F_s via a piece-wise linear function tied down at the origin as well as the κ points $\{s_k, \tau_k\}_{k=1}^{\kappa}$. That is, we approximate the inverse CDF of F_s as

$$\hat{F}_s^{-1}(\tau;\mathcal{S}) \equiv \begin{cases} \tau\frac{s_1}{\tau_1} & \text{if } \tau \in [0,\tau_1] \\ s_{i-1} + (\tau - \tau_{i-1})\frac{s_i - s_{i-1}}{\tau_i - \tau_{i-1}} & \text{if } \tau \in (\tau_{i-1}, \tau_i], \;\; i = 2, \ldots, \kappa - 1. \\ s_{\kappa-1} + (\tau - \tau_{\kappa-1})\frac{s_\kappa - s_{\kappa-1}}{\tau_\kappa - \tau_{\kappa-1}} & \text{if } \tau \in (\tau_{\kappa-1}, 1]. \end{cases} \tag{11}$$

In principle, by letting $\kappa \to \infty$ as $N \to \infty$, we could derive a semiparametric estimate of the shock distribution F_s. The identifiability of the shock distribution F_s stands in contrast to the literature on dynamic discrete choice models and dynamic discrete games (eg. Rust (1987), Pesendorfer and Schmidt-Dengler (2003)), which typically assumes that the distribution of the unobserved errors F_s is completely known, because it is not identified from the discrete actions. The continuity of the action spaces gives more identifying information to F_s. However, in deriving the asymptotic theory for our estimator, we focus on the case where F_s is parameterized by a fixed, finite-dimensional vector γ. The asymptotic characterization of the semiparametric case is technically involved, and we leave it for future work.

Second, we note that the validity of this pairwise-differencing step imposes several requirements on the economic model. First, for every $\tau \in (0,1)$, there must exist a pair of individuals (i,j) such that (i) $x_i + q_i = x_j + q_j$; (ii) $F(q_i|x_i) = F(q_j|x_j) = \tau$; and (iii) $x_i \neq x_j$. Second, we require that the (i,j)-differenced first-order condition is a nontrivial equation:

$$0 = U_3(x_i, s(\tau), q_i; \theta_1) - U_3(x_j, s(\tau), q_j; \theta_1). \tag{12}$$

If these two conditions are met, then Eq. (12) can be used to solve for θ_1 and $s(\tau) \equiv F_s^{-1}(\tau)$.

To focus on this, let us consider several examples. First, consider the linear-quadratic case, where the policy function will be linear in its arguments: $q = a + bx + cs$. (i) and (ii) imply that, for a given s, we must find $x_1 \neq x_2$ so that $x_1 + a + bx_1 + cs = x_2 + a + bx_2 + cs$, which can clearly only be satisfied if and only if $x_1 = x_2$. Hence, we require the policy function to be nonlinear in x. For example, for the case $q = (x+s)^2$, the points (x_1, x_2) would work, where $x_1 = \left[-(2s+1) \pm \sqrt{4s + 2 + 4\left(x_2 + (x_2+s)^2\right)}\right]/2$, as long as s does not become too negative (in our empirical work, we assume that $s > 0$).

Finally, we note that the deterministic accumulation assumption plays an important role in allowing us to control for the expected value function term by conditioning on the sum $(x+q)$; if accumulation was stochastic, then (in general) we would have to condition on q and x separately, leaving no more variation in our data to identify the parameters.[5] Generally, however, the deterministic accumulation assumption can be replaced by a weaker assumption that x_{t+1} is some scalar-valued deterministic (and potentially parameterized) function of x_t and q_t: $x_{t+1} = l\left(x_t, q_t; \zeta\right)$, for instance: the non-randomness of x_{t+1} conditional on x_t and q_t is the crucial assumption. The particular form $l\left(x_t, q_t; \zeta\right) = x_t + q_t$ used above arises naturally in our empirical illustration later.

3.2 Second step

Not all model parameters can be identified from the first step pairwise-differencing approach. In the second step, we use the first-order condition again to derive moment restrictions to estimate parameters in θ which were not in the subset θ_1 estimated in the first step. Recall that $\theta_2 \equiv \{\theta \setminus \theta_1\}$ denotes the set of parameters which were not estimated in the first step.

Given $\hat{\gamma}$ and $\hat{\theta}_1$, (respectively) the shock distribution parameters and the subset of the utility function parameters which were estimated in the first step, define the first-order condition for observation (i,t) where the investment level $q_{it} > 0$ as follows:

$$0 = h_{it}(x_{it}, q_{it}; \hat{\theta}_1, \hat{\gamma}, \theta_2) \equiv U_3\left(x_{it}, s_t(q_{it}, \hat{\gamma}), q_{it}; \hat{\theta}_1, \theta_2\right) + \beta \mathcal{V}'\left(x_{it} + q_{it}; \hat{\theta}_1, \hat{\gamma}, \theta_2\right) \tag{13}$$

where $s_t(q_{it}, \hat{\gamma}) \equiv F_s^{-1}\left(\hat{F}\left(q_{it}|x_{it}\right); \hat{\gamma}\right)$. In what follows, we will use $\hat{F}_s$ as shorthand for $F_s(\cdot; \hat{\gamma})$.

To begin, we assume that we have been able to compute the expected value function $\mathcal{V}(\cdots)$ (we delay discussion of how this can be done until later). Due to estimation error from estimating θ_1, γ and $F(q|x)$ in the first step, the first order condition $h_{it}(x_{it}, q_{it}; \hat{\theta}_1, \hat{\gamma}, \theta_2)$

[5]In this sense, the role of the deterministic accumulation assumption is equivalent to the exclusion restriction in identifying selection models (cf. Heckman (1990), p. 315), which allows the researcher to condition on certain covariates in order to control for the selection probability, leaving the excluded (from the selection equation) variable to sweep out the variation in the dependent variable and identify the parameters.

need not be identically zero, even at the true parameter vector θ_0. Therefore, we estimate θ_2 via a minimum distance procedure:[6]

$$\hat{\theta}_2 = \text{argmin}_{\theta_2} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\left(q_{it} \neq 0\right) \cdot \left[h_{it}(x_{it}, q_{it}; \hat{\theta}_1, \hat{\gamma}, \theta_2)\right]^2 .$$

As before, we only include observations with non-zero investment ($q \neq 0$) in the objective function.

Computing the expected value function For computing the expected value function $(\cdot; \theta, \gamma)$, we have several options. First, given our estimates of $\hat{F}_s$ from the first step, we could compute $\mathcal{V}(\cdot; \theta)$ directly by standard numerical dynamic programming methods, as described in Rust (1996) and Judd (1998). Numerical dynamic programming for continuous control problems can be difficult, since it involves solving for the optimal policy function $q(x, s)$ at every point (x, s) in the state space.

If one wished to avoid numerical computation of the dynamic programming problem, we propose an alternative which is attractive when the datasets available to the researcher are large (as in the dataset we consider later), is to compute the value function by a forward integration procedure in the spirit of Hotz and Miller (1993). This procedure exploits the representation of the value function at time t as the expected discounted sum of future utilities (cf. Eq. (4)) rather than the more familiar recursive representation via the Bellman equation (cf. Eq. (3)) which underlies numeric dynamic programming algorithms. Hotz and Miller (1993) recognize that, given enough data, and a particular parametric form of the per-period utility function $U(\cdots; \theta)$, the expectation over future states in equation (4) can be replaced by forward integration over the observed conditional probabilities $\hat{F}(q|x)$ (cf. equation (3.12) in Hotz and Miller (1993)).

Using Assumption 1 above, this approach can be used in the case where agent i's control variable is continuous. More precisely, we approximate the agent's expected value function at a particular point $(x_t + q_t)$ and particular parameters (θ, γ) as:

$$\mathcal{V}\left(x_t + q_t; \hat{F}_s, \theta\right) = \sum_{z=t+1}^{T} \left[\beta^{s-t} \int \int \cdots \int U\left(x_z, \tilde{s}(q_z), q_z; \theta\right) \prod_{z'=t+1}^{z} \hat{F}\left(dq_{z'} | x_{z'}\right)\right] \tag{14}$$

[6]The choice of a square norm is somewhat arbitrary; other norms, such as absolute deviation, may also be used. Furthermore, weighting schemes could be introduced to improve the efficiency of the estimation procedure. We have not considered these alternative possibilities.

where $\tilde{s}(q_z)$ is the $\hat{F}(q_z|x_z)$-th quantile of the distribution $F_s(\cdot;\hat{\gamma})$. Here T ($>> t$) is a sufficiently large number so that the discounted utility beyond period T is essentially zero, when discounted back to period t. Hence, given a particular parameter vector (θ,γ), $\mathcal{V}(x_t+q_t;\theta,\gamma)$ can be calculated based simply on our nonparametric estimates of $\hat{F}(q_{z'}|x_{z'})$. Assumption 1 permits us to recover the correspondence between $\tilde{s}(q_z)$ and q_z.

In practice, we simulate $h_{it}(x_{it},s_{it};\hat{\theta}_1,\hat{\gamma},\theta_2)$ based upon the techniques described in Hotz, Miller, Sanders, and Smith (1994). Using the shock distribution parameters $\hat{\gamma}$ estimated from the first step, $\mathcal{V}(x_{it}+q_{it};\theta,\gamma)$ can be simulated as

$$\mathcal{V}_t(x_{it}+q_{it};\theta,\gamma) \approx \left[\sum_{z=t+1}^{T} \beta^{z-t} U\left(x_{iz}^l, s_{iz}^l, q_z\left(s_{iz}^l, x_{iz}^l\right);\theta\right)\right] \tag{15}$$

for each (i,z) where[7]

- $s_{iz}^l \sim \hat{F}_s(\cdot;\hat{\gamma})$, *i.i.d.* across i, z
- $x_{iz}^l = x_{iz-1}^l + q\left(s_{iz-1}^l, x_{iz-1}^l\right)$, $z=t+1,\ldots,T$; $x_{it+1}^l = x_{it+1}$, $\forall l$.
- $q\left(s_{iz}^l, x_{iz}^l\right) = \hat{F}^{-1}_{q|x_{iz}^l}\left(F_s\left(s_{iz}^l;\hat{\gamma}\right)|\cdot\right)$.

Finally, in order to implement the second-step estimator, we must also compute the derivative of the expected value function. This is most easily approximated by a numeric finite-difference:

$$\mathcal{V}'(x_{it}+q_{it};\hat{F}_s,\theta) \approx \frac{\mathcal{V}(x_{it}+q_{it}+\Delta;\hat{F}_s,\theta)-\mathcal{V}(x_{it}+q_{it};\hat{F}_s,\theta)}{\Delta}$$

for Δ small.

Consequently, we estimate θ_2 using the following simulated minimum distance estimator:

$$\hat{\theta}_2 = \operatorname{argmin}_{\theta_2} \frac{1}{NT}\frac{1}{L}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L}\left[h_{it}(x_{it}^l, q_{it}^l;\hat{\theta}_1,\hat{\gamma},\theta_2)\right]^2. \tag{16}$$

In the following theorem, we present the limit distribution for the second-step estimator $\hat{\theta}_2$. The proof is in the appendix, section A.2.

[7] To further facilitate the simulation procedure, we can draw a u_{it}^l, which is a $U[0,1]$ random variable, for each trader i, period t, and simulation l, and recover the shock $s_{it}^l = \hat{F}_s^{-1}(u_{it}^l)$, and the quantity $q\left(s_{it}^l, x_{it}^l\right) = \hat{F}^{-1}_{q|x_{it}^l}\left(u_{it}^l\right)$ by inverting the estimated shock and conditional quantity CDFs at u_{it}^l.

Theorem 2 *Given Assumptions 2 and 3, as well as assumptions 7 and 8 in the appendix,*

$$\sqrt{N}\left(\hat{\theta}_2 - \theta_2\right) \xrightarrow{d} N\left(0, \bar{A}^{-1}\bar{\Omega}\bar{A}^{-1}\right) \quad \textit{for} \quad \bar{\Omega} = Ev^*\left(z_l, \psi, \theta_2\right) v^*\left(z_l, \psi, \theta_2\right)'$$

as $N \to \infty$, where $\bar{A}$ and $v^*(\cdots)$ are defined in Eqs. (34) and (36) of the appendix.

In deriving the asymptotic behavior of the second-stage estimator, we ignore the approximation error in computing the expected value function (as well as its derivative). For numeric dynamic programming methods, which usually are based on iterative function approximation algorithms, this generally requires that the accuracy of the function approximation (as measured in terms of the order of an approximating polynomial, or number of knot points in an approximating spline) increase as $N \to \infty$. For the simulation-based approximation, we require that the number of simulation draws L increases quickly enough as $N \to \infty$ so that variation due to the simulation itself is small enough and does not affect the asymptotic variance. From Gourieroux and Monfort (1996), a sufficient condition for the asymptotic variance to be unaffected from simulation error is that $L/\sqrt{N} \to \infty$.

In principle, given the parametric assumptions on $F_s(\cdot;\gamma)$, the parameters θ and γ could be jointly estimated in the second step, without requiring the pairwise-differencing first-step. However, by estimating θ_1 and γ in the first step, we reduce the number of parameters which must be estimated in the second step. Since the second step potentially involves numeric dynamic programming in order to recover the value function, reducing the dimensionality of the parameter space also reduces significantly the number of times that the value function must be computed, therefore reducing the computational burden. Such a "two-step" approach was also taken in Rust's (1987) dynamic discrete-choice model of bus engine replacement, in which the parameters describing the mileage Markov transition matrix was estimated in a first-step to reduce the computational burden in the second-step, which involved value function iteration.

4 Empirical Illustration: Markets for Milk Production Quota

As an illustration of our methodology, we estimate a dynamic trading model of the milk production quota market. In Ontario, Canada, milk production is controlled via production quotas which grant holders the right to produce a certain quantity of milk per year. Since 1980, in the province of Ontario these quota have been traded among dairy farmers in monthly double auctions administered by the Dairy Farmers of Ontario (DFO) (cf. Biggs (1990)). This paper analyzes data from the eleven auctions between September 1997 and July 1998. Our goal is to estimate the parameters of agents' utility functions, and the distribution of the unobserved state variables, using the two-step pairwise-differencing methodology described earlier.

Each quota exchange is a double auction market. All producers who wish to sell quota submit offers to the exchange indicating that they have a certain volume of quota for sale and at a certain minimum price per unit. Producers who wish to buy quota submit bids to the exchange indicating that they would like to buy a certain volume of quota and that they are willing to pay a specific maximum price per unit. Units are traded at a market clearing price (MCP) at which the total quantity demanded (approximately) equals the total quantity supplied.

In order to fit the milk-quota trading market into our dynamic framework, we consider a dynamic, forward-looking model of the quota demand/supply process, in which each individual trader faces a dynamic optimization problem where the market clearing price p_t is treated as a state variable.

The timing of the game is as follows. At the beginning of month t, trader i owns x_{it} units of production quota. She experiences a shock s_{it} and must decide the amount of quota q_{it} to trade at any price p_t. The amount actually transacted would be $q\left(x_{it}, s_{it}, p_t^*\right)$, where p_t^* denotes the realized market-clearing price for period t. Generally, the optimal amount is given by a function $q\left(x_{it}, s_{it}, p_t\right)$ which takes values in $(-\infty, \infty)$. For positive values of $q\left(\cdots\right)$, this can be interpreted as a demand function, and when negative it can be interpreted as a supply function.

Given these considerations, we model each trader i as choosing a sequence $\{q_{it}\}_t$ to maximize the expected discounted present value of its utility from its milk quota trading operations:

$$\max_{\{q_{it}\}_t} \mathcal{E}_{0|\{q_{it}\}_t} \sum_{t=0}^{\infty} \beta^t U\left(x_{it}, s_{it}, q_{it}, p_t^*; \theta\right) \tag{17}$$

subject to

$$x_{it+1} = x_{it} + q_{it}; \quad s_{it+1} \sim F_s; \quad F(p_0^*, p_1^* \dots). \tag{18}$$

The expectation $\mathcal{E}_{0|\{q_{it}\}_t}$ is over the sequences of x_{it}, p_t^* and s_{it} induced by trader i's chosen sequence $\{q_{it}\}_t$ as well as her beliefs about the stochastic process of market clearing prices $\{p_t^*\}_t$. Given her beliefs about the evolution of prices, each trader i's optimal policy function $q_{it} = q\left(x_{it}, s_{it}, p_t^*\right)$ can be recursively characterized via Bellman's equation:

$$q\left(x_{it}, p_t^*, s_{it}\right) = \operatorname{argmax}_q U\left(x_{it}, s_{it}, q_{it}, p_t^*; \theta\right) + \beta \mathcal{V}_t\left(x_{it} + q_{it}\right) \tag{19}$$

where

$$\mathcal{V}_t\left(x_{it}+q_{it}\right) \equiv \mathcal{E}_{p^*_{t+1},s_{it+1}|p^*_t,s_{it}} V\left(x_{it}+q_{it}, s_{it+1}, p^*_{t+1}\right).$$

A dynamic competitive rational expectations equilibrium of a continuum market is characterized by perfect foresight on the part of traders about the sequence of market-clearing prices, even though at the individual trader level there is uncertainty about the shocks received by other traders. Subsequently, equilibrium strategies in this market can be characterized as optimal policies of a dynamic optimization problem solved by each trader individually. See section A.3 in the Appendix for more discussion of the equilibrium.

The most significant change we make in the empirical model is to adapt it for a nonstationary problem, which arises because in the perfect foresight equilibrium path, the policy function in each period depends explicitly on the particular period in question due to the price that period. We accommodate these features by estimating a finite horizon model with $T = 11$ (since we only observe eleven months of data) but allow the terminal value of the problem to depend on $x_{12} = x_{11} + q_{11}$, the stock that a given trader has after the first eleven months. More specifically, we assume this terminal value for trader i to be a flexible polynomial in x_{i12}.[8]

Our assumptions that the observed trades and prices are from the dynamic equilibrium path of a trading market with a continuum of traders seems appropriate given the large numbers of traders participating in each auction and, moreover, is borne out in part by the data, as we explain in section A.4 in the Appendix. Nevertheless, these are strong assumptions which rules out two features — strategic behavior and price uncertainty — which may be important in this market. Therefore, we stress that the empirical example here mainly illustrates the use of our estimator in practice.

4.1 Estimation Results

Section A.4 in the Appendix contains a full description of our dataset. For our results, we assume a CARA form for the utility function:

$$U(w_{it}) = -\frac{1}{r}\exp\left(-rw_{it}\right)$$

[8]The perfect foresight assumption is convenient because it reduces the dimensionality of the state space of the dynamic problem. If prices evolved stochastically (from the agent's point of view), the we would also need to estimate the transition probabilities for price, which would make the task of computing the value function (by either backward recursion or forward integration) much more difficult.

and the following linear specification for trader i's period t payoff:

$$w_{it} = x_{it} \cdot s_{it} - p_t \cdot q_{it} - K \cdot \mathbf{1}\left(q_{it} \neq 0\right).$$

The per-period payoffs for each trader are as follows. Each period, trader i receives some profits $x_{it} \cdot s_{it}$ from producing and selling milk under its current stock of quota, but pays an amount $p_t \cdot q\left(x_{it}, s_{it}, p_t\right)$ to acquire additional quota. Furthermore, she incurs a fixed adjustment cost K which is associated with any non-zero transaction of quota (and the magnitude of which is not dependent on the amount of quota transacted): this would accommodate not only bidding costs but also general fixed costs associated with expanding/contracting the scale of milk production (and is required to rationalize the large number of zeros, as evidenced in Figure 1). Hence, this implies that the FOCs (Eq. (7) and (13)) only hold for those observations (i, t) for which $q_{it} \neq 0$, so that only these observations will be used in the first step of the estimation procedure. Given this specification, s_{it} can be interpreted as stochastic production shocks which affect a trader's profits from his milk production.

[Figure 1 about here.]

While we have derived the asymptotic covariance matrix for our estimator in Theorems 1 and 2 above, it is fairly tedious in practice to compute it. Therefore, in the empirical implementation, we obtained standard errors for our estimates using a bootstrap re-sampling procedure: for each specification, we re-sampled (with replacement) sequences from the dataset, and re-estimated the model for each re-sampled dataset. The reported bootstrap confidence intervals are therefore the empirical quantiles of the distribution of parameter estimates obtained in this fashion. We employed 50 bootstrap resamples in computing each set of standard errors.

Log-normal shock distribution parameterization First, we present results from a tightly parameterized model, assuming a log-normal specification for F_s, whereby $\log s \sim N\left(\mu, \sigma^2\right)$. The parameter estimates are shown in Table 1.

[Table 1 about here.]

These magnitudes imply that the mean shock is 26.687, which can be interpreted as the monthly return from a unit of quota (in 1986 Canadian $'000). At a price of about $11,000 per unit of quota, these magnitudes imply that a producer would "recoup" her investment in less than half a month ($=\frac{11{,}000}{26{,}687}$): this seems quite an unrealistically small figure. The estimates of K and r indicate, respectively, very small adjustment costs (only 26 cents) and a very low level of risk aversion. In the top graph of Figure 2, we present our estimate of the implied policy function $q_t(x, s, p)$ for the log-normal distribution results. The policy function is estimated using Eq. (10) above.

[Figure 2 about here.]

Piecewise-linear shock distribution parameterization Second, we present results using a more flexible piecewise-linear form for the shock distribution F_s, as described in Eq. (11) above. We jointly estimated the 0.15, 0.25, 0.5, 0.75, and 0.85 quantiles for F_s. The estimated CDF is graphed in Figure 3. The median shock is estimated to be about 1.1, implying (using the same reasoning as in the previous paragraph) that a trader recoups his investment in about 10 ($=\frac{11{,}000}{1.1}$) months: this appears more realistic than the estimate obtained from the log-normal parameterization, reported above.[9]

[Figure 3 about here.]

[Table 2 about here.]

In the bottom graph of Figure 2, we present our estimate of the implied policy function $q_t(x, s, p)$ for the F_s (with linear interpolation) estimated in the first step (and plotted in Figure 3). The estimate of K implies that the magnitude of fixed adjustment costs are \$65.70. The estimate of r, the coefficient of absolute risk aversion, is 0.0189, remains very small.

5 Conclusions and Extensions

In this paper, we proposed a new two-step pairwise-differencing procedure for structural estimation of a dynamic optimization model with unobserved state variables. To our knowledge, our estimator represents the first application of pairwise-difference methods, which

[9] We also considered a specification allowing F_s to depend on covariates which vary across periods t by estimating a different distribution of s_{it} for each time period t. However, we found that the covariates had little effect, and left the results virtually unchanged. Therefore, we do not report those results.

have primarily been used in cross-sectional contexts (cf. Honore and Powell (1994)), to structural dynamic optimization problems.

The most restrictive assumption made in this paper is that the unobserved state variables are independent across time. In accommodating serial correlation, we would have to consider carefully the problem of initial conditions which, in turn, is very closely related to the issue of unobserved individual-specific heterogeneity (cf. Heckman (1981)). In future work, we plan to explore extensions to our procedure to handle these issues.

The estimation procedure only accommodates univariate unobserved state variables in agents' policy functions. This rules out multi-agent models in which the unobserved state variables of *all* the agents enter into each agent's policy function, as in the dynamic oligopoly model considered by Berry and Pakes (2000) where one firm's optimal investment is affected by the productivity state of every firm in the market, and all of these productivities are unobservable to the econometrician. It will be interesting to investigate in future work whether monotonicity and quantile invariance can be useful in these situations.

References

Aguirregabiria, V., and P. Mira (2003): "Sequential Estimation of Dynamic Discrete Games," Working Paper, Boston University.

Ahn, H., and J. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

Ait-Sahalia, Y. (1996): "The Delta Method for Nonparametric Kernel Functionals," Department of Economics, Princeton University.

Bajari, P., and L. Benkard (2001): "Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," mimeo., Stanford University and NBER.

Bajari, P., L. Benkard, and J. Levin (2004): "Estimating Dynamic Models of Imperfect Competition," NBER Working Paper, #10432.

Berry, S., M. Ostrovsky, and A. Pakes (2004): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry Exit Examples)," NBER Working Paper, #10506.

Berry, S., and A. Pakes (2000): "Estimation from the Optimality Conditions for Dynamic Controls," mimeo., Yale University.

Biggs, E. (1990): *The Challenge of Achievement: The Ontario Milk Marketing Board's first 25 years of operation.* The Ontario Milk Marketing Board.

Eberly, J. (1994): "Adjustment of Consumers' Durables Stocks: Evidence from Automobile Purchases," *Journal of Political Economy*, 102, 403–436.

GOURIEROUX, C., AND A. MONFORT (1996): *Simulation-Based Econometric Methods*. Oxford University Press.

HECKMAN, J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data*, ed. by C. Manski, and D. McFadden. MIT Press.

——— (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313–318.

HONG, H., AND M. SHUM (2004): "Pairwise-Difference Estimation of Dynamic Models with Stochastic Euler Equations," Working Paper, Duke University.

HONORE, B., AND J. POWELL (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64, 241–278.

HOPENHAYN, H. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.

HOTZ, J., AND R. MILLER (1993): "Conditional Choice Probabilties and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

HOTZ, J., R. MILLER, S. SANDERS, AND J. SMITH (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

JOVANOVIC, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50, 649–670.

JUDD, K. (1998): *Numerical Methods in Economics*. MIT Press.

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

MATZKIN, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1376.

OLLEY, S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

PAKES, A. (1994): "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in *Advances in Econometrics: Proceedings of the 1990 Meetings of the Econometric Society*, ed. by C. Sims, and J.-J. Laffont, vol. 2, pp. 171–259. Cambridge University Press.

PAKES, A., AND M. SIMPSON (1989): "Patent Renewal Data," *Brookings Papers on Economic Activity: Microeconomics*, pp. 331–403.

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003): "Identification and Estimation of Dynamic Games," NBER Working Paper, #9726.

POWELL, J., J. STOCK, AND T. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1996): "Numerical Dynamic Programming in Economics," in *Handbook of Computational Economics*, ed. by D. K. H. Amman, and J. Rust. North Holland.

Taber, C. (2000): "Semiparametric identification and heterogeneity in discrete-choice dynamic programming models," *Journal of Econometrics*, 96, 201–229.

A Limit distribution for multi-step estimator

Throughout, we focus on the case of a single cross-section in a fixed period t for clarity of presentation; the argument can be easily adapted to a panel setting. Hence, in this section we consider estimators which minimize cross-sectional versions of the objective functions (9) and (16) in the main text. Specifically, we consider a first-step estimator

$$\hat{\psi} \equiv \left(\hat{\theta}_1, \hat{\gamma}\right) = \underset{\theta_1, s_1, \dots, s_K}{\operatorname{argmin}} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g_n(z_i, z_j; \psi) \equiv$$

$$\frac{1}{N^2} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \frac{1}{h} K\left(\frac{(x_{it} + q_{it}) - (x_{jt} + q_{jt})}{h} \right) \left[U_3\left(x_{it}, \hat{s}_{it}, q_{it}; \theta_1\right) - U_3\left(x_{jt}, \hat{s}_{jt}, q_{jt}; \theta_1\right) \right]^2 \right\} \right]$$

where $z_i \equiv (x_{it}, q_{it}, x_{it+1})$, $z_j \equiv (x_{jt}, q_{jt}, x_{jt+1})$. Similarly, the second-step estimator we focus on is

$$\hat{\theta}_2 = \operatorname{argmin}_{\theta_2} \frac{1}{N} \sum_{i=1}^{N} \left(h_{it}(x_{it}, q_{it}; \hat{\theta}_1, \hat{\gamma}, \theta_2) \right)^2 .$$

Throughout, we assume that kernel function $K(\cdot)$ is a r-th order kernel satisfying Assumption 2 in the main text.

A.1 Proof of Theorem 1 (asymptotic normality for first-step estimator)

In addition to the assumptions given in the main text, we also make the following auxiliary regularity conditions:

Assumption 5 *Let $P \equiv dim(\psi) = dim(\gamma) + dim(\theta_1)$, and*

$$G_0(\psi) \equiv E_{x,q} E_{x',q'} \left[\mathbf{1}\left(x + q = x' + q'\right) \left[U_3\left(x, F_s\left(F(q|x); \gamma\right), q, \theta_1\right) - U_3\left(x', F_s\left(F(q'|x'); \gamma\right), q', \theta_1\right) \right]^2 \right],$$

the limit objective function of the first-step estimator.

1. *$\psi \in \Psi$, a compact subset of $\mathbf{R}^P$, and true value $\psi_0 \in int(\Psi)$.*
2. *$G_0(\psi)$ is uniquely maximized at ψ_0.*
3. *$g_n(z_i, z_j; \psi)$ is a continuous function of $\psi \in \Psi$ with probability 1*
4. *$|g_n(z_i, z_j; \psi)| < \bar{g}(z_i, z_j)$ for all $\psi \in \Psi$, for some function $\bar{g}(\cdots)$ with $E\left[\bar{g}(\cdots)\right] < \infty$.*

Equation (9) implies that the first step estimator solves the estimating equation:

$$\begin{aligned} W_n(\psi) &\equiv \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{h} K\left(\frac{x_{it+1} - x_{jt+1}}{h} \right) \hat{m}\left(z_i, z_j, \hat{\psi}\right) \frac{\partial}{\partial \psi} \left[\hat{m}\left(z_i, z_j, \hat{\psi}\right) \right] \\ &\equiv \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} r(z_i, z_j, \hat{\psi}) = 0 \end{aligned}$$

where

$$\begin{aligned} \hat{m}\left(z_i,z_j,\psi\right) \equiv & U_3\left(x_{it},F_s^{-1}\left(\hat{F}\left(q_{it}|x_{it}\right);\gamma\right),q_{it};\theta_1\right)-U_3\left(x_{jt},F_s^{-1}\left(\hat{F}\left(q_{jt}|x_{jt}\right);\gamma\right),q_{jt};\theta_1\right), \\ m\left(z_i,z_j,\psi\right) \equiv & U_3\left(x_{it},F_s^{-1}\left(F\left(q_{it}|x_{it}\right);\gamma\right),q_{it};\theta_1\right)-U_3\left(x_{jt},F_s^{-1}\left(F\left(q_{jt}|x_{jt}\right);\gamma\right),q_{it};\theta_1\right). \end{aligned} \tag{20}$$

We make some additional smoothness and differentiability assumptions:

Assumption 6 *Define $\tilde{v}(z_i,\psi) \equiv E\left[r(z_i,z_j,\psi)|z_i\right]$ and $\lambda(\psi) \equiv E\tilde{v}(z_i,\psi)$.*

1. *Stochastic equicontinuity: $\forall\ \delta_n \to 0$, $\sqrt{n}\left[sup_{||\psi-\psi_0||<d_n}||W_n(\psi)-W_n(\psi_0)-\lambda(\psi)||\right] \xrightarrow{p} 0$.*
2. *$\lambda(\psi_0)=0$ and is differentiable at ψ_0, with nonsingular Jacobian matrix A.*
3. *The expectation $E\left[||r(z_i,z_j,\psi)||^2\right]$ exists and is finite.*

These assumptions resemble those required for Theorem 2 in Honore and Powell (1994).

For a fixed i, the following approximation can be derived:

$$\begin{aligned} & \hat{F}\left(q_{it}|x_{it}\right)-F_s\left(s_{it}\right) \\ \approx & \frac{1}{f\left(x_{it}\right)}\left[\frac{1}{Nh}\sum_{l=1}^{N}\mathbf{1}\left(q_{lt}<q_{it}\right)K\left(\frac{x_{lt}-x_{it}}{h}\right)-f\left(x_{it}\right)F_s\left(s_{it}\right)\right]-\frac{F_s\left(s_{it}\right)}{f\left(x_{it}\right)}\left[\frac{1}{Nh}\sum_{l=1}^{N}K\left(\frac{x_{lt}-x_{it}}{h}\right)-f\left(x_{it}\right)\right] \\ = & \frac{1}{f\left(x_{it}\right)}\left[\frac{1}{Nh}\sum_{l=1}^{N}\mathbf{1}\left(q_{lt}<q_{it}\right)K\left(\frac{x_{lt}-x_{it}}{h}\right)\right]-\frac{F_s\left(s_{it}\right)}{f\left(x_{it}\right)}\left[\frac{1}{Nh}\sum_{l=1}^{N}K\left(\frac{x_{lt}-x_{it}}{h}\right)\right]. \end{aligned} \tag{21}$$

Using a standard first order Taylor expansion argument, we can approximate the estimator by

$$\sqrt{N}\left(\hat{\psi}-\psi\right)=A_n^{-1}\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\hat{m}\left(z_i,z_j,\psi\right)\frac{\partial}{\partial\psi}\left[\hat{m}\left(z_i,z_j,\psi\right)\right] \tag{22}$$

where

$$A_n \equiv \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\frac{\partial}{\partial\psi}\left(\hat{m}\left(z_i,z_j,\psi^*\right)\frac{\partial}{\partial\psi}\left[\hat{m}\left(z_i,z_j,\psi^*\right)\right]\right)$$

is the Jacobian term which can be re-written as

$$\begin{aligned} & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\frac{\partial}{\partial\psi}\left(\left[m\left(z_i,z_j,\psi\right)\right]\frac{\partial}{\partial\psi}\left[m\left(z_i,z_j,\psi\right)\right]\right)+o_p(1) \\ = & E_{z_j}E_{z_i}\left[\frac{\partial}{\partial\psi}\left(\left[m\left(z_i,z_j,\psi\right)\right]\frac{\partial}{\partial\psi}\left[m\left(z_i,z_j,\psi\right)\right]\right)\middle|x_{it+1}=x_{jt+1}\right]+o_p\left(1\right) \\ = & E_{z_j}E_{z_i}\left[\frac{\partial}{\partial\psi}\left[m\left(z_i,z_j,\psi\right)\right]\frac{\partial}{\partial\psi}\left[m\left(z_i,z_j,\psi\right)\right]\middle|x_{it+1}=x_{jt+1}\right]+o_p\left(1\right) \\ \equiv & A+o_p(1), \end{aligned} \tag{23}$$

where A is the same matrix as stated in Assumption 6(2). (ψ^* is a set of intermediate values between ψ and $\hat{\psi}$.) The first equality in (23) follows from the uniform convergence of the nonparametric kernel estimate of the conditional expectation. The second equality in (23) follows since, by assumption, $m(z_i, z_j, \psi) \equiv 0$ for all z_i, z_j such that $x_{it+1} = x_{jt+1}$,[10] so that

$$E_{z_j} E_{z_i} \left[m(z_i, z_j, \psi) \frac{\partial^2}{\partial \psi \partial \psi'} m(z_i, z_j, \psi) \middle| x_{it+1} = x_{jt+1} \right] \equiv 0. \tag{24}$$

Next, we address the terms which appear behind the double summation in (22). Define

$$\begin{aligned} \hat{w}(z_i, z_j, \psi) &\equiv \hat{m}(z_i, z_j, \psi) \frac{\partial}{\partial \psi} \left[\hat{m}(z_i, z_j, \psi) \right], \\ w(z_i, z_j, \psi) &\equiv m(z_i, z_j, \psi) \frac{\partial}{\partial \psi} \left[m(z_i, z_j, \psi) \right]. \end{aligned} \tag{25}$$

Note that $\hat{w}(z_i, z_j, \psi)$ can be approximated by the first order linearization

$$\begin{aligned} & w(z_i, z_j, \psi) + \frac{\partial}{\partial s_{it}} w(z_i, z_j, \psi) \left(\hat{F}(q_{it}|x_{it}) - F_s(s_{it}) \right) + \frac{\partial}{\partial s_{jt}} w(z_i, z_j, \psi) \left(\hat{F}(q_{jt}|x_{jt}) - F_s(s_{jt}) \right) + o_p(1) \\ =& w(z_i, z_j, \psi) \\ & + \frac{\partial}{\partial s_{it}} w(z_i, z_j, \psi) \frac{1}{f_s(s_{it})} \left[\frac{1}{f(x_{it})} \frac{1}{Nh} \sum_{l=1}^{N} \mathbf{1}(q_{lt} < q_{it}) K\left(\frac{x_{lt} - x_{it}}{h} \right) - \frac{F_s(s_{it})}{f(x_{it})} \frac{1}{Nh} \sum_{l=1}^{N} K\left(\frac{x_{lt} - x_{it}}{h} \right) \right] \\ & + \frac{\partial}{\partial s_{jt}} w(z_i, z_j, \psi) \frac{1}{f_s(s_{jt})} \left[\frac{1}{f(x_{jt})} \frac{1}{Nh} \sum_{l=1}^{N} \mathbf{1}(q_{lt} < q_{jt}) K\left(\frac{x_{lt} - x_{jt}}{h} \right) - \frac{F_s(s_{jt})}{f(x_{jt})} \frac{1}{Nh} \sum_{l=1}^{N} K\left(\frac{x_{lt} - x_{jt}}{h} \right) \right] + o_p(1) \end{aligned}$$

where the second equality follows by substituting in equation (21) above.

Hence, we can approximate the linear term in equation (22) by a U-statistic representation:

$$\begin{aligned} \sqrt{N}\left(\hat{\psi} - \psi \right) =& A_n^{-1} \Bigg\{ \frac{1}{\sqrt{N} N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{h} K\left(\frac{x_{it+1} - x_{jt+1}}{h} \right) w(z_i, z_j, \psi) \\ & + \frac{1}{\sqrt{N} N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \frac{1}{h} K\left(\frac{x_{it+1} - x_{jt+1}}{h} \right) v(z_i, z_j, z_l; \psi) + o_p(1) \Bigg\} \end{aligned} \tag{26}$$

where

$$\begin{aligned} v(z_i, z_j, z_l; \psi) =& \frac{\partial}{\partial s_{it}} w(z_i, z_j, \psi) \frac{1}{f_s(s_{it})} \frac{1}{h} K\left(\frac{x_{lt} - x_{it}}{h} \right) \frac{1}{f(x_{it})} \left[\mathbf{1}(q_{lt} < q_{it}) - F_s(s_{it}) \right] \\ & + \frac{\partial}{\partial s_{jt}} w(z_i, z_j, \psi) \frac{1}{f_s(s_{jt})} \frac{1}{h} K\left(\frac{x_{lt} - x_{jt}}{h} \right) \frac{1}{f(x_{jt})} \left[\mathbf{1}(q_{lt} < q_{jt}) - F_s(s_{jt}) \right]. \end{aligned}$$

Given our assumptions on the kernel and bandwidth sequence (Assumption 2 in the main text), the bias terms in the nonparametric kernel estimation are asymptotically negligible and the conditions

[10] The intuition behind this result is that conditional on the event that $x_{it+1} = x_{jt+1}$, the estimating equation is an identity for the model we consider.

for Lemma 3.1 in Powell, Stock, and Stoker (1989) hold. Hence, we can invoke the projection representation of (26). For the first term within the curly brackets in (26), we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)w\left(z_i,z_j,\psi\right)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)w\left(z_i,z_j,\psi\right)\Big|z_i\right)+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)w\left(z_i,z_j,\psi\right)\Big|z_j\right)+o_p\left(1\right)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E_{z_j}\left[w\left(z_i,z_j,\psi\right)\Big|x_{j+1}=x_{i+1}\right]f\left(x_{it+1}\right)+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}E_{z_i}\left[w\left(z_i,z_j,\psi\right)\Big|x_{i+1}=x_{j+1}\right]f\left(x_{jt+1}\right)+o_p\left(1\right)\\
=&o_p\left(1\right)
\end{aligned}$$

Both terms in the above display vanish asymptotically for the same reasoning that leads to (24). Therefore, if $F\left(q_{it}|x_{it}\right)$ were known, the pairwise differencing step does not introduce any additional variation to the parameter estimate: the nonparametric estimates of $F_{q|x}$ produce all the first order variation for the estimates of the ψ parameters. This is reflected in the non-negligible limit for the second term of equation (26):

$$\begin{aligned}
&\frac{1}{\sqrt{N}N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)v\left(z_i,z_j,z_l;\psi\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)v\left(z_i,z_j,z_l;\psi\right)\Big|z_l\right)+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)v\left(z_i,z_j,z_l;\psi\right)\Big|z_j\right)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)v\left(z_i,z_j,z_l;\psi\right)\Big|z_i\right)+o_p\left(1\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E\left(\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)v\left(z_i,z_j,z_l;\psi\right)\Big|z_l\right)+o_p\left(1\right)\equiv\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\tilde{v}\left(z_l,\psi\right)+o_p\left(1\right)
\end{aligned}$$

where the second inequality follows from Assumption 3, and the $\tilde{v}(\cdots)$ function in the final expression is the same as that specified in Assumption 6. After tedious but straightforward calculations:

$$\begin{aligned}
\tilde{v}\left(z_l,\psi\right)=&E_{z_i}\left[E_{z_j}\left(\frac{\partial}{\partial s_{it}}w\left(z_i,z_j,\psi\right)\Big|x_{jt+1}=x_{it+1}\right)\frac{f\left(x_{it+1}\right)}{f_s\left(s_{it}\right)}\left(\mathbf{1}\left(q_{lt}<q_{it}\right)-F_s\left(s_{it}\right)\right)\Big|x_{it}=x_{lt}\right]\\
&+E_{z_j}\left[E_{z_i}\left(\frac{\partial}{\partial s_{jt}}w\left(z_i,z_j,\psi\right)\Big|x_{it+1}=x_{jt+1}\right)\frac{f\left(x_{jt+1}\right)}{f_s\left(s_{jt}\right)}\left(\mathbf{1}\left(q_{lt}<q_{jt}\right)-F_s\left(s_{jt}\right)\right)\Big|x_{jt}=x_{lt}\right].
\end{aligned}\tag{27}$$

Therefore, we conclude that

$$\sqrt{N}\left(\hat{\psi}-\psi\right)\xrightarrow{d}N\left(0,A^{-1}\Omega A^{-1}\right)\tag{28}$$

where

$$\Omega = E\tilde{v}\left(z_l, \psi\right)\tilde{v}\left(z_l, \psi\right)' \tag{29}$$

and A is defined as in (23) above. ■

In principle, the asymptotic variance can be consistently estimated. We can estimate A by

$$\hat{A} \equiv \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{x_{jt+1}-x_{it+1}}{h}\right)\frac{\partial}{\partial\psi}\hat{m}\left(z_i, z_j, \hat{\psi}\right)\frac{\partial}{\partial\psi}\hat{m}\left(z_i, z_j, \hat{\psi}\right)' \tag{30}$$

and Ω by

$$\hat{\Omega} \equiv \frac{1}{N}\sum_{l=1}^{N}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\hat{v}\left(z_i, z_j, z_l, \hat{\psi}\right)\right]\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\hat{v}\left(z_i, z_j, z_l, \hat{\psi}\right)\right]'$$

for

$$\begin{aligned}\hat{v}\left(z_i, z_j, z_l; \hat{\psi}\right) =& \frac{\partial}{\partial s_{it}}\hat{w}\left(z_i, z_j, \hat{\psi}\right)\frac{1}{f_s\left(s_{it};\hat{\gamma}\right)}\frac{1}{h}K\left(\frac{x_{lt}-x_{it}}{h}\right)\frac{1}{\hat{f}\left(x_{it}\right)}\left[\mathbf{1}\left(q_{lt}<q_{it}\right)-F_s\left(s_{it};\hat{\gamma}\right)\right]\\ &+\frac{\partial}{\partial s_{jt}}\hat{w}\left(z_i, z_j, \hat{\psi}\right)\frac{1}{f_s\left(s_{jt};\hat{\gamma}\right)}\frac{1}{h}K\left(\frac{x_{lt}-x_{jt}}{h}\right)\frac{1}{\hat{f}\left(x_{jt}\right)}\left[\mathbf{1}\left(q_{lt}<q_{jt}\right)-F_s\left(s_{jt};\hat{\gamma}\right)\right].\end{aligned}$$

In the above, $\hat{f}\left(x_{it}\right)$ denotes the usual kernel estimate: $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it}-x_{jt}}{h}\right)$ and, similarly, $\hat{f}\left(x_{jt}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{x_{it}-x_{jt}}{h}\right)$.

A.2 Proof of Theorem 2 (asymptotic normality for second-step estimator)

In this proof, We abstract away from approximation error in computing (or simulating) the value function. If the value function is simulated, this requires that a sufficiently large number of simulation draws to compute the value function so that the variation due to the simulation itself is small enough and does not affect the asymptotic variance.

We begin with some conditions analogous to Assumption 5 for the first-step estimator:

Assumption 7 *Let $P_2 \equiv dim(\theta_2)$, and*

$$H_0(\theta_2) \equiv E\left[h(x, q; \hat{F}_s, \hat{\theta}_1, \theta_2)\right]^2,$$

the limit objective function of the second-step estimator.

1. *$\theta_2 \in \Theta_2$, a compact subset of $\mathbf{R}^{P_2}$, and true value $\theta_2^0 \in int(\Theta_2)$.*
2. *$H_0(\theta_2)$ is uniquely maximized at θ_2^0.*
3. *$h(\cdots, \theta_2)$ is a continuous function of $\theta_2 \in \Theta_2$ with probability 1*

4. $|h(\cdots,\theta_2)| < \bar{\bar{g}}(\cdots)$ *for all* $\theta_2 \in \Theta_2$, *for some function* $\bar{\bar{g}}(\cdots)$ *with* $E\left[\bar{\bar{g}}(z_i,z_j)\right] < \infty$.

For a sequence of *i.i.d.* random variables $\tau_{iz}^l \sim U(0,1)$, we can express the forward value function as an expectation with respect to the τ_{iz}^l's. For each $x_{iz}^l = x_{iz-1}^l + \hat{F^z}^{-1}_{q|x_{iz-1}^l}\left(\tau_{iz}^l\right)$:

$$\bar{V}\left(x_{it}+q_{it};\hat{\psi},\theta_2\right) \equiv E_\tau \sum_{z=t+1}^{T} \beta^{z-t} U\left(x_{iz}^l, F_s^{-1}\left(\tau_{iz}^l;\hat{\gamma}\right), \hat{F^z}^{-1}_{q|x_{iz}^l}\left(\tau_{iz}^l\right);\hat{\theta}_1,\theta_2\right). \tag{31}$$

Given this expression, we note that our second step estimator $\hat{\theta}_2$ solves the sample score function:

$$J_N(\theta_2) \equiv \frac{1}{N}\sum_{i=1}^{N} \bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}\left(\cdot|:\right), \hat{\psi};\theta_2\right) = 0, \tag{32}$$

where the arguments of $\hat{F}_q^{-1}(\cdot|:)$ are related to the terms τ_{iz}^l and x_{iz}^l in equation (31), and

$$\begin{aligned}
&\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}\left(\cdot|:\right), \hat{\psi};\theta_2\right) \\
&\equiv \left[U_3\left(x_{it}, F_s^{-1}\left(\hat{F}\left(q_{it}|x_{it}\right);\hat{\gamma}\right), q_{it};\hat{\theta}_1,\theta_2\right) + \beta\bar{V}'\left(x_{it}+q_{it};\hat{\psi},\theta_2\right)\right] * \\
&\qquad \frac{\partial}{\partial\theta_2}\left[U_3\left(x_{it}, F_s^{-1}\left(\hat{F}\left(q_{it}|x_{it}\right);\hat{\gamma}\right), q_{it};\hat{\theta}_1,\theta_2\right) + \beta\bar{V}'\left(x_{it}+q_{it};\hat{\psi},\theta_2\right)\right]
\end{aligned}$$

with $\bar{V}(\cdots)$ defined as in equation (31) above. The inclusion of $\hat{F}_q^{-1}(\cdot|:)$ as an argument in $\bar{h}(\cdots)$ recognizes that $\bar{V}(\cdots)$ depends on the entire set of functions $\hat{F}_q^{-1}(\cdot|:) \equiv \left\{\hat{F}_{q|:}^{-1}(\cdot)\right\}_{t=1}^{T}$, not just on any one of these functions evaluated at a particular point.

Analogously to Assumption 6, we make some additional smoothness and differentiability assumptions on the $\bar{h}(\cdots)$ function:

Assumption 8 *Define* $\mu(\theta_2) \equiv E\bar{h}(\cdots,\theta_2)$.

1. *Stochastic equicontinuity of the sample score function: for any sequence* $\delta_n \to 0$, $\sqrt{N}\left[sup_{||\theta_2-\theta_2^0||<d_n}||J_N(\theta_2) - J_N(\theta_2^0) - \mu(\theta_2)||\right] \xrightarrow{p} 0$.
2. *Central limit theorem for the sample score function:* $\sqrt{N}J_N(\theta_2) \xrightarrow{d} N(0,\bar{\Omega})$.
3. $\mu(\theta_2) = 0$ *at* θ_2^0 *and is differentiable at* θ_2^0, *with nonsingular Jacobian matrix* $\bar{A}$.

The estimated conditional quantile of q given x through inverting the kernel density estimate of the conditional CDF can be approximated linearly by

$$\begin{aligned}
&\hat{F}_{q|x}^{-1}(\tau) - F_{q|x}^{-1}(\tau) = \frac{1}{f_{q|x}\left(F_{q|x}^{-1}(\tau)\right)}\left[\hat{F}\left(F_{q|x}^{-1}(\tau)\,|x\right) - \tau\right] + o_p\left(\frac{1}{\sqrt{N}}\right) \\
=&\frac{1}{N}\sum_{l=1}^{N}\frac{1}{f_{q|x}\left(F_{q|x}^{-1}(\tau)\right)}\frac{1}{f(x)}\left[\mathbf{1}\left(q_{lt} \leq F_{q|x}^{-1}(\tau)\right) - \tau\right]\frac{1}{h}K\left(\frac{x_{lt}-x}{h}\right) + o_p\left(\frac{1}{\sqrt{N}}\right) \qquad (33) \\
\equiv&\frac{1}{N}\sum_{l=1}^{N}G_h\left(q_{lt}, x_{lt}, \tau, x\right) + o_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Taking a Taylor expansion of (32) around $\hat{\theta}_2$, one obtains

$$\begin{aligned}
0 &= \frac{1}{N}\sum_i \bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \hat{\theta}_2\right) \\
&= \frac{1}{N}\sum_i \bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \theta_2\right) \\
&\qquad + \left(\hat{\theta}_2 - \theta_2\right)\frac{1}{N}\sum_i \frac{\partial}{\partial\theta_2}\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \theta_2\right) + o_p(1).
\end{aligned}$$

Applying a standard law of large numbers to the Jacobian term in the above expression:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_2}\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \hat{\theta}_2\right) \\
&\xrightarrow{p}\bar{A} \equiv E\frac{\partial}{\partial\theta_2}\bar{h}\left(x_{it}, q_{it}, F\left(q_{it}|x_{it}\right), F_q^{-1}(\cdot|\,:), \psi, \theta_2\right).
\end{aligned} \qquad (34)$$

with $\bar{A}$ the same matrix as specified in Assumption 8. Hence,

$$\sqrt{N}\left(\hat{\theta}_2 - \theta_2\right) = \bar{A}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \theta_2\right) + o_p\left(\frac{1}{\sqrt{N}}\right). \qquad (35)$$

The recursive use of the nonparametric estimates $F_q^{-1}(\cdot|\,:)$ in the construction of the expected value function above (in equation (31)) makes it rather tedious to derive explicit analytic expressions for the asymptotic linear representation of the nonlinear (in $\hat{F}_q^{-1}(\cdot|\,:)$) functional $\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \theta_2\right)$. Hence, in the following we will use $g\left(x_{it}, q_{it}, F\left(q_{it}|x_{it}\right), \left(\hat{F}_q^{-1}(\cdot|\,:) - F_q^{-1}(\cdot|\,:)\right), \psi, \theta_2\right)$ to denote a linear (in $\hat{F}_q^{-1}(\cdot|\,:)$) functional, without explicitly writing out its lengthy analytic formula.

Given this notation, we derive the asymptotically linear representation of $\bar{h}\left(x_{it}, q_{it}, \hat{F}\left(q_{it}|x_{it}\right), \hat{F}_q^{-1}(\cdot|\,:), \hat{\psi}, \theta_2\right)$

as

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{h}\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),\left(\hat{F}_{q}^{-1}\left(\cdot|:\right)\right),\hat{\psi},\theta_2\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{h}\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),F_{q}^{-1}\left(\cdot|:\right),\psi,\theta_2\right)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}g\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),\hat{F}_{q}^{-1}\left(\cdot|:\right)-F_{q}^{-1}\left(\cdot|:\right),\psi,\theta_2\right)+B\sqrt{N}\left(\hat{\psi}-\psi\right)\\
&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\frac{\partial}{\partial s}U_3\left(x_{it},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)}\left[\hat{F}\left(q_{it}|x_{it}\right)-F\left(q_{it}|x_{it}\right)\right]+o_p\left(1\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\frac{1}{N}\sum_{i=1}^{N}g\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),G_h\left(q_{lt},x_{lt},\cdot,:\right),\psi,\theta_2\right)+B\sqrt{N}\left(\hat{\psi}-\psi\right)\\
&\quad+\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{x_{lt}-x_{it}}{h}\right)\frac{\frac{\partial}{\partial s}U_3\left(x_{it},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)f\left(x_{it}\right)}\left[\mathbf{1}\left(q_{lt}\leq q_{it}\right)-F\left(q_{it}|x_{it}\right)\right]+o_p\left(1\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E\left[g\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),F\left(q_{it}|x_{it}\right),G\left(q_{lt},x_{lt},\cdot,:\right),\psi,\theta_2\right)\Big|q_{lt},x_{lt}\right]+B\sqrt{N}\left(\hat{\psi}-\psi\right)\\
&\quad\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E_{z_i}\delta\left(x_{it}-x_{lt}\right)\frac{\frac{\partial}{\partial s}U_3\left(x_{it},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)f\left(x_{it}\right)}\left[\mathbf{1}\left(q_{lt}\leq q_{it}\right)-F\left(q_{it}|x_{it}\right)\right]+o_p\left(1\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E_{z_i}\left[g\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),G\left(q_{lt},x_{lt},\cdot,:\right),\psi,\theta_2\right)\right]+B\sqrt{N}\left(\hat{\psi}-\psi\right)\\
&\quad\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E_{z_i}\left[\frac{\frac{\partial}{\partial s}U_3\left(x_{lt},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)}\left(\mathbf{1}\left(q_{lt}\leq q_{it}\right)-F\left(q_{it}|x_{lt}\right)\right)\Big|x_{it}=x_{lt}\right]+o_p\left(1\right).
\end{aligned}
$$

In the above display $B\equiv E\frac{\partial}{\partial\psi}\bar{h}\left(x_{it},q_{it},F\left(q_{it}|x_{it}\right),F_{q}^{-1}\left(\cdot|:\right),\psi,\theta_2\right)$, and $G_h\equiv\left\{G_h,t=1,\ldots,T\right\}$, where each G_h is defined as in equation (33) above. Furthermore, the collection of functions

$$G\left(q_{lt},x_{lt},\cdot,:\right)=\left\{G\left(q_{lt},x_{lt},\cdot,:\right),t=1,\ldots,T\right\}$$

denotes the deterministic limit of the collection G_h where, for each $t=1,\ldots,T$:

$$G\left(q_{lt},x_{lt},\cdot,\cdot\right)\equiv\frac{1}{f_{q|:}\left(F_q\left(\cdot|:\right)\right)}\frac{1}{f\left(:\right)}\left[\mathbf{1}\left(q_{lt}\leq F_{q|:}^{-1}\left(\cdot\right)\right)-\cdot\right]\delta\left(x_{lt}-:\right)$$

where $\delta\left(\cdot\right)$ denotes the generalized Dirac function (cf. Ait-Sahalia (1996)):

$$\int^{x}\delta\left(u\right)du=0\text{ for }x<0\quad\text{and}\quad\int^{x}\delta\left(u\right)du=1\text{ for x}\geq 0.$$

Next, we use the modeling assumption that $\bar{h}\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),F_{q}^{-1}\left(\cdot|:\right),\psi,\theta_2\right)\equiv 0$ in order to summarize the above analysis as

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{h}\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),\hat{F}_q^{-1}\left(\cdot|\cdot\right),\hat{\psi},\theta_2\right)\\
=&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E_{z_i}\left[g\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),G\left(q_{lt},x_{lt},\cdot,:\right),\psi,\theta_2\right)\right]+B\sqrt{N}\left(\hat{\psi}-\psi\right)\\
&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}E_{z_i}\left[\frac{\frac{\partial}{\partial s}U_3\left(x_{lt},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)}\left(\mathbf{1}\left(q_{lt}\leq q_{it}\right)-F\left(q_{it}|x_{lt}\right)\right)\Big|x_{it}=x_{lt}\right]+o_p\left(1\right)
\end{aligned}$$

If we plug the linear representation of $\sqrt{N}\left(\hat{\psi}-\psi\right)$ (equation (28) above) into the linear representation for $\sqrt{N}\left(\hat{\theta}_2-\theta_2\right)$ (equation (35) above), we conclude that

$$\sqrt{N}\left(\hat{\theta}_2-\theta_2\right)\xrightarrow{d}N\left(0,\bar{A}^{-1}\bar{\Omega}\bar{A}^{-1}\right)\quad\text{for}\quad\bar{\Omega}=Ev^*\left(z_l,\psi,\theta_2\right)v^*\left(z_l,\psi,\theta_2\right)'$$

where

$$\begin{aligned}
v^*\left(z_l,\psi,\theta_2\right)\equiv&E_{z_i}\left[g\left(x_{it},q_{it},G\left(q_{lt},x_{lt},\cdot,:\right),\psi,\theta_2\right)\right]+BA^{-1}\tilde{v}\left(z_l,\psi\right)\\
&+E_{z_i}\left[\frac{\frac{\partial}{\partial s}U_3\left(x_{lt},s_{it},q_{it};\theta\right)}{f\left(s_{it}\right)}\left(\mathbf{1}\left(q_{lt}\leq q_{it}\right)-F\left(q_{it}|x_{lt}\right)\right)\Big|x_{it}=x_{lt}\right]
\end{aligned}\tag{36}$$

and A and $\tilde{v}\left(z_l,\psi\right)$ are defined in Eqs. (23) and (27). Moreover $\bar{\Omega}$ is the matrix specified in Assumption 8. ■

Again, each of terms in the asymptotic distribution can, in principle, be consistently estimated. A can be estimated by (30), $\tilde{v}\left(z_l,\psi\right)$ by

$$\hat{v}\left(z_l,\hat{\psi}\right)\equiv\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{h}K\left(\frac{x_{it+1}-x_{jt+1}}{h}\right)\hat{v}\left(z_i,z_j,z_l,\hat{\psi}\right).$$

B can be consistently estimated by

$$\hat{B}\equiv\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\psi}\bar{h}\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),\hat{F}_q^{-1}\left(\cdot|:\right),\hat{\psi},\hat{\theta}_2\right).$$

$\bar{A}$ can be consistently estimated by

$$\hat{\bar{A}}\equiv\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_2}\bar{h}\left(x_{it},q_{it},\hat{F}\left(q_{it}|x_{it}\right),\hat{F}_q^{-1}\left(\cdot|:\right),\hat{\psi},\hat{\theta}_2\right).$$

Finally, we estimate $\bar{\Omega}$ by $\hat{\bar{\Omega}}\equiv\frac{1}{N}\sum_{i=1}^{N}\hat{v}^*\left(z_l,\hat{\psi},\hat{\theta}_2\right)\hat{v}^*\left(z_l,\hat{\psi},\hat{\theta}_2\right)'$, for

$$\hat{v}^*\left(z_l,\hat{\psi},\hat{\theta}_2\right)=\hat{v}_2\left(z_l,\hat{\psi},\hat{\theta}_2\right)+\hat{B}\hat{A}^{-1}\hat{v}\left(z_l,\hat{\psi}\right)+\hat{v}_3\left(z_l,\hat{\psi},\hat{\theta}_2\right)$$

where

$$\hat{v}_2\left(z_l, \hat{\psi}, \hat{\theta}_2\right) \equiv \frac{1}{N}\sum_{i=1}^{N} g\left(x_{it}, q_{it}, G_h\left(q_{lt}, x_{lt}, \cdot, :\right), \hat{\psi}, \hat{\theta}_2\right)$$

$$\hat{v}_3\left(z_l, \hat{\psi}, \hat{\theta}_2\right) \equiv \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{x_{it}-x_{lt}}{h}\right) \frac{\frac{\partial}{\partial s} U_3\left(x_{it}, \hat{F}_s^{-1}\left(\hat{F}\left(q_{it}|x_{it}\right); \hat{\gamma}\right), q_{it}; \hat{\theta}\right)}{\hat{f}\left(s_{it}; \hat{\gamma}\right) \hat{f}\left(x_{it}\right)} \left[\mathbf{1}\left(q_{lt} \leq q_{it}\right) - \hat{F}\left(q_{it}|x_{it}\right)\right].$$

B Additional description of dynamic market equilibrium

We employ a model of dynamic competitive equilibrium in a continuum market composed of individually atomistic traders, which bears some similarities to industry equilibrium models considered in Jovanovic (1982) and Hopenhayn (1992). The price-taking nature of equilibrium implies that $F_i(p_0^*, p_1^* \ldots)$ is identical for all traders i (i.e., that any individual trader's choice $\{q_{it}\}_t$ has no effect on the sequence of prices which result). The evolution of prices derives from consideration of the requirement that they clear the market each period: given policies $q\left(x_{it}, s_{it}, p_t\right)$, $\forall i$,

$$p_t : \int\int q\left(x, s, p_t\right) \mathcal{J}_t(dx)\mathcal{H}_t(ds) = 0\ , \quad \forall t \tag{37}$$

where $\mathcal{J}_t(\cdot)$ and $\mathcal{H}_t(\cdot)$ denote, respectively, the distribution of quota stocks and shocks in the cross-section of traders during period t. From equation (37), it is clear that the evolution of prices is completely determined by the sequences $\{\mathcal{J}_t\}$ and $\{\mathcal{H}_t\}$, which are deterministic, given any initial distribution of stocks $\mathcal{J}_0$ and shocks $\mathcal{H}_0$. In particular, $\mathcal{H}_t(s)$ coincides with the marginal distribution of s_{it} in the cross-section of traders. Given our *i.i.d.* assumption on the shock distribution, it is immediate that

$$\mathcal{H}_t(s) = F_s\left(s\right),\ \forall t. \tag{38}$$

Similarly, the cross-sectional distribution of stocks $\mathcal{J}_t(x)$ is induced by $\mathcal{J}_{t-1}(x)$, p_{t-1}, and $\mathcal{H}_{t-1}(s)$ where

$$\mathcal{J}_t(x) = \int\int \mathbf{1}\left(z + q(z, s, p_{t-1}) \leq x\right) \mathcal{H}_{t-1}(ds)\mathcal{J}_{t-1}(dz) \tag{39}$$

with $\mathcal{J}_0(z)$ given.

Given any initial stock distribution $\mathcal{J}_0$, the sequences $\{\mathcal{J}_t\}$ and $\{\mathcal{H}_t\}$ are both deterministic, and evolve according to (38) and (39). Therefore, by the market clearing conditions (37), the sequence $\{p_t\}_t$ is also deterministic. This leads to the implication that, in competitive equilibrium in this market, all traders will have perfect foresight about the evolution of prices.

In summary, a perfect foresight equilibrium of this market consists of deterministic sequences $\{\mathcal{J}_t\}$, $\{\mathcal{H}_t\}$, and $\{p_t\}_t$ such that (1) prices clear the market each period (satisfying (37)); (2) each trader selects individually optimal trades (according to (19)); and (3) $\mathcal{H}_t$ and $\mathcal{J}_t$ evolve according to (38) and (39), given a particular $\mathcal{J}_0$.

The estimation procedure described above utilizes the equilibrium model only insofar as it implies that, at the individual trader level, quantities are chosen as if in a dynamic optimization problem with perfect foresight about the sequence of market-clearing prices. However, we do not employ the market clearing conditions (37) in our estimation — therefore, they could potentially serve as overidentifying restrictions which could be used to gauge the fit of our behavioral model.

B.1 Data: summary statistics

In this section, we present some summary statistics on the data. The trading unit for quota is expressed in kilograms of butterfat, and one kilogram of quota purchased on the exchange allows a producer to ship one kilogram of butterfat per day, in perpetuity, for as long as the unit of quota is held.[11] Over the eleven exchanges, we observe the bids placed by 2,574 distinct producers. For each trader, we have data on her total quota stock in September 1997 (the first month in our sample), as well as her purchases/sales of quota in each subsequent month, which we used to construct her total quota for each month. Figure 1 is a histogram of quantities traded by each (trader/month). Note that about 85% of the observations have zero trade.

In Table 3, we present some information on each individual exchange. Column 4 in Table 3 shows that a large number of sellers and buyers participate in each exchange, which suggests that there may not be much scope for strategic behavior, which we have not accommodated in our empirical model.

[Table 3 about here.]

However, columns 5 and 6 of table 3 also indicate that each bidder's chance of getting their order filled (i.e., submitting selling bids below the MCP, or submitting buying bids above the MCP) also varied between exchanges. For sellers, the success rate was highest in late 1997 and late summer 1998. In our framework, we model all unsuccessful traders as submitting a zero bid at the market clearing price.

[11] Prior to September 1997, a unit of quota conferred on its owner the right to produce milk containing one kilogram of butterfat *per year*. In September 1997, however, the trading unit for quota was re-defined in kilograms of butterfat per day.

Figure 1: Histogram of quantity trade per trader/month

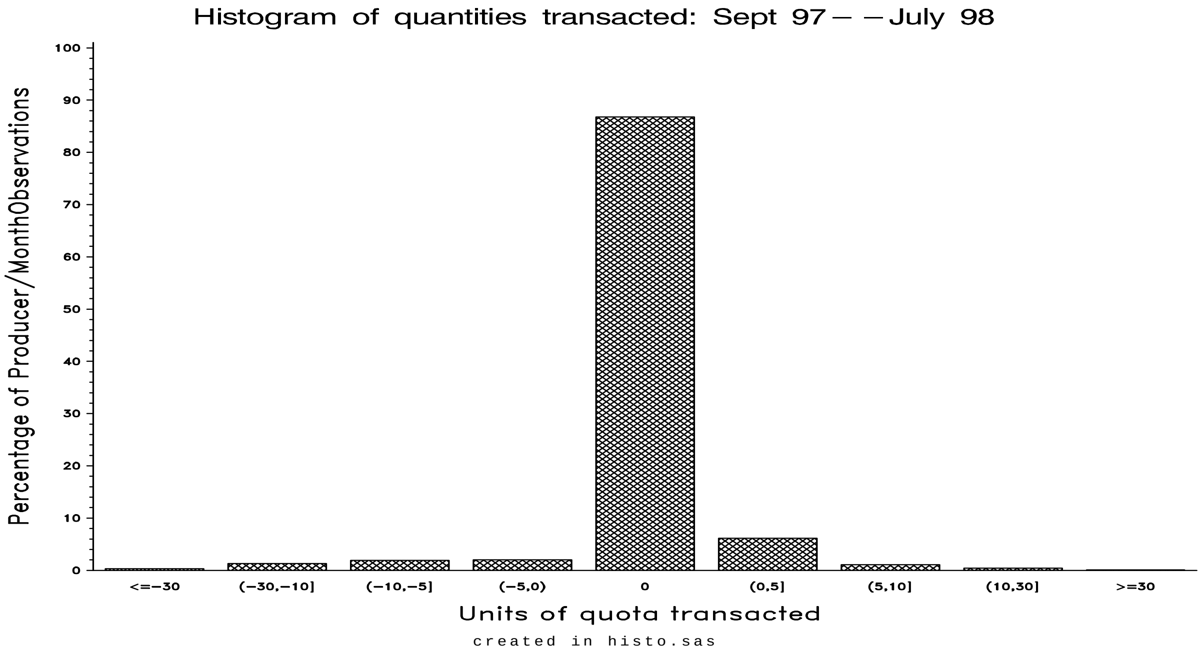
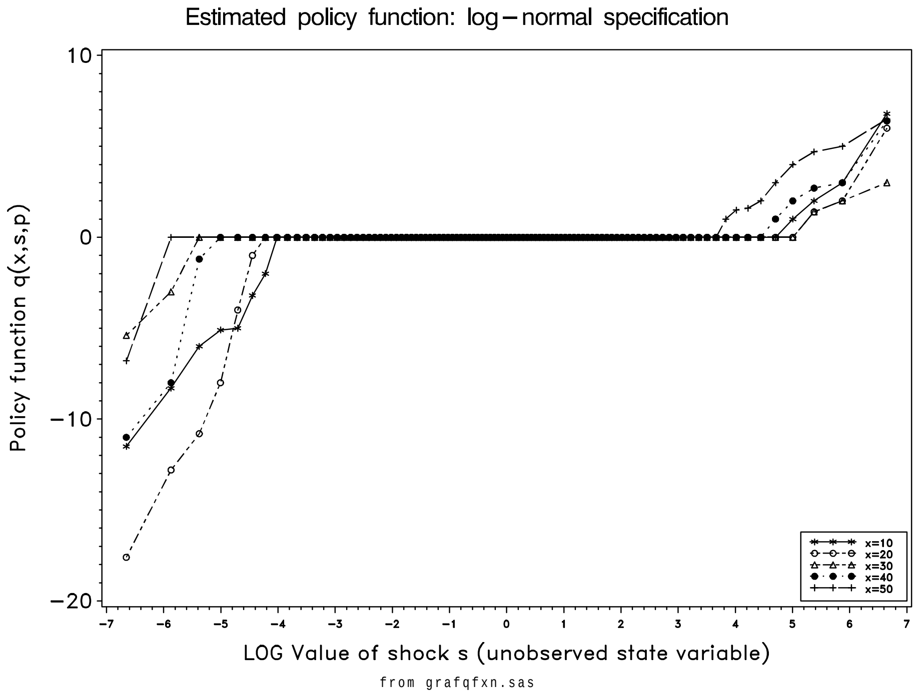


Figure 2: Estimated Policy Functions



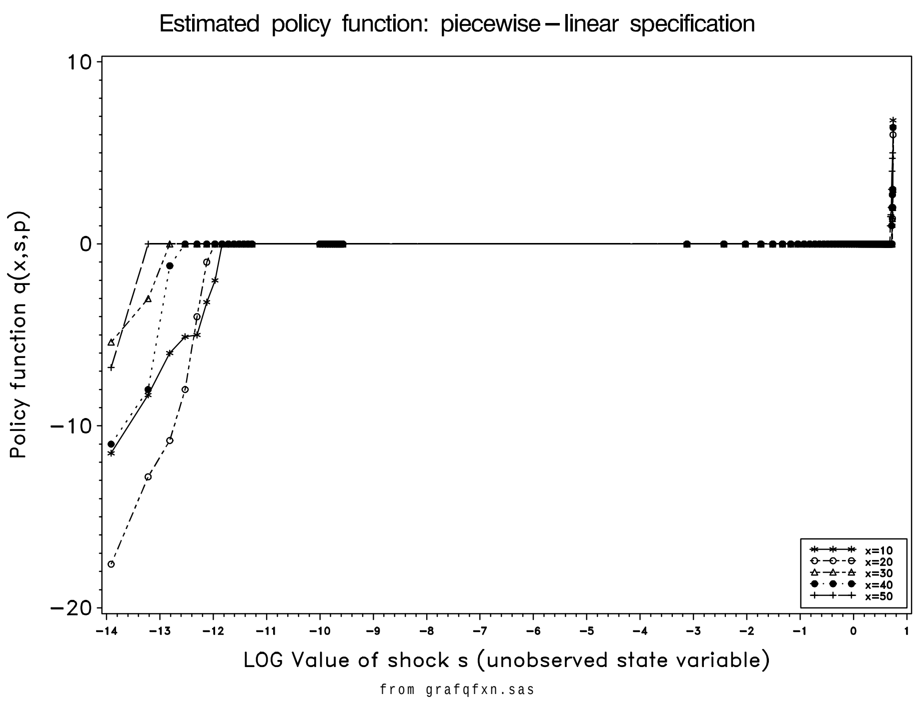
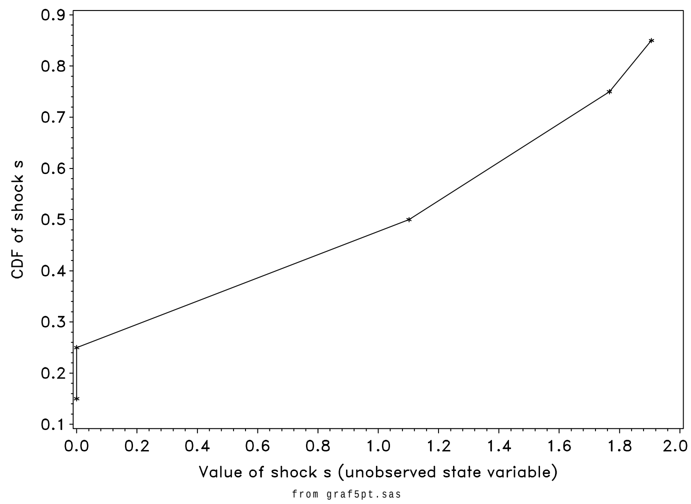


Figure 3: Estimated CDF of shock s
Estimated using equation (9).



Five quantiles were estimated: 0.15, 0.25, 0.5, 0.75, 0.85.

Table 1: Parameter estimates: log-normal specification for F_s
$\log s \sim N\left(\mu, \sigma^2\right)$

	Estimate	Bootstrap confidence intervals:[a]			
		10%	40%	60%	90%
K	0.00026	0.00010	0.00020	0.00026	0.00060
r	0.03199	0.00579	0.02935	0.03121	0.03374
μ	-0.7956	-0.7956	-0.7956	-0.7956	-0.6706
σ	2.8591	2.2830	2.8591	2.8591	2.8591

[a]Empirical quantiles of distribution of parameter estimates obtained via (nonparametric) bootstrap resampling procedure.

Table 2: Parameter estimates: flexible piecewise-linear specification for F_s

	Estimate	Bootstrap confidence intervals:[a]			
		10%	40%	60%	90%
K	0.0657	0.0443	0.0619	0.0774	0.1081
r	0.0189	0.0134	0.0165	0.0197	0.0270

Fourth-order polynomial approximation employed for terminal value (cf. end of section 4).

[a]Empirical quantiles of distribution of parameter estimates obtained via (nonparametric) bootstrap resampling procedure. These confidence bands do not reflect estimation error due to the first-step pairwise-difference estimates of F_s.

Table 3: Summary statistics for each quota exchange

Year	Month	#bids	#traders	#sellers[a] (%success)	#buyers (% success)	MCP[b]
1997	9	726	509	219 (63.4)	290 (59.7)	15999.00
1997	10	523	396	248 (57.2)	148 (84.5)	15250.00
1997	11	603	471	253 (84.6)	218 (82.6)	15025.00
1997	12	538	419	163 (94.4)	256 (46.5)	15510.00
1998	1	529	428	126 (91.2)	302 (39.1)	16150.00
1998	2	780	579	182 (85.7)	397 (53.9)	16360.00
1998	3	648	532	214 (93.0)	318 (75.8)	16501.00
1998	4	579	447	212 (27.4)	235 (94.0)	15499.00
1998	5	725	575	247 (52.2)	328 (98.5)	14500.00
1998	6	838	625	178 (86.5)	447 (72.5)	14500.25
1998	7	547	446	105 (88.6)	341 (44.0)	15025.00

[a] Note that #buyers + #sellers = #traders. %success denotes % of sellers who sold in the exchange, i.e., who submitted bids at or below the MCP.

[b] Canadian dollars per kilogram of butterfat per day.

PARTICIPANTS

From Queen's

Nazim Belhocine

Allen Head

Sharif Kahn

Beverly Lapham

Jeremy Lise

Huw Lloyd-Ellllis

James MacKinnon

Saleem Nechi

Yu Ren

Stephen Tapp

James Thompson

Casey Warman

From Elsewhere

Jason Allen, Bank of Canada

Eugene Choo, University of Toronto

Susumu Imai, Concordia University

Philippe Morin, University of Toronto

May 12, 2006 (Friday)

Logon

Quantitative Economics

May 12-13

NIEPA V: Workshop on Numerically Intensive Economic Policy Analysis

Papers will be presented in Dunning 27.

E-poster session will be held in the Dunning Conference Room.

Location: Dunning 27 and other locations

View All Dates

Posted by: Chris Ferrall, 6-Apr-06 12:37pm ET

Quantitative Economics

8:45am - 9:45am

Samuel Danthine, UQAM

Bargaining Frictions and Hours Worked

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 Noon ET

Quantitative Economics

10:00am - 11:00am

Samuel Kortum, Minnesota

An Anatomy of International Trade: Evidence from French Firms

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:02pm ET

Quantitative Economics

11:00am - Noon

Elena Pastorino, Iowa

Career Dynamics Under Uncertainty: Estimating the Value of Firm Experimentation

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:07pm ET

Quantitative Economics

Noon - 1:45pm

E-Poster Session

Ahu Gemici, Pennsylvania, *Understanding Family Migration and Employment Decisions*

Jean-Francois Houde, Queen's, *Identification and 2-step Estimation of DDC Models with Unobserved Heterogeneity*

Kazuko Kano, British Columbia, *Menu Costs, Strategic Interactions, and Retail Price Movements*

Natalia Mishagina, Queen's, *TBA*

Travis Ng, Toronto, *Outsourcing by risk-pooling*

Joel Rodrigue, Queen's, *Imported Productivity Gains and Domestic Firms*

Mari Sakudo, Pennsylvania, *Co-residence and intergenerational transfers*

Katsumi Shimotsu, Queen's, *Nested Pseudo-likelihood Estimation and Bootstrap-based Inference for Structural Discrete Markov Decision Models*

Jun Zhang, Queen's, *Nonparametric Identification and Estimation of Contests*

Location: Dunning Conference Room

Posted by: Chris Ferrall, 13-Apr-06 12:25pm ET

Quantitative Economics

1:45pm - 2:45am

Carlos J. Serrano, Toronto

The Market for Intellectual Property

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:11pm ET

Quantitative Economics

3:00pm - 4:00pm

Matthew Shum, Johns Hopkins

Pairwise-Difference Estimation of a Dynamic Optimization Model

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:21pm ET

Quantitative Economics

May 12-13

NIEPA V: Workshop on Numerically Intensive Economic Policy Analysis

Papers will be presented in Dunning 27.

E-poster session will be held in the Dunning Conference Room.

Location: Dunning 27 and other locations

View All Dates

Posted by: Chris Ferrall, 6-Apr-06 12:37pm ET

Quantitative Economics

9:30am - 10:30am

Rasmus Lentz, Wisconsin

An Empirical Model of Growth through Product Innovation

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:13pm ET

Quantitative Economics

10:45am - 11:45am

Carrie Lee, Queen's

Wealth Shocks, Birth Cycles and Wage Trends: A Dynamic Equilibrium Model of Fertility, Wages and Labor Supply in the 20th Century

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:15pm ET

Quantitative Economics

1:00pm - 2:00pm

Victor Aguirregabiria, Boston

A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:20pm ET

Quantitative Economics
2:00pm - 3:00pm
Suqin Ge, Minnesota

Women's College Choice: How Much Does Marriage Matter?

NIEPA V

Location: Dunning 27

Posted by: Chris Ferrall, 13-Apr-06 12:17pm ET

Claim Form for Expenses
Numerical Intensive Economic Policy Analysis
May 12-13, 2006

Name: ______________________________

Organization: ______________________________

Address: ______________________________

____

____

Phone: ______________________ FAX: ______________________

E-Mail: ______________________________

Expenses

	Cdn. $	U.S. $
Travel (plane, train, car, etc.):	$ __________	$ __________
Accommodation:	$ __________	$ __________
Taxis:	$ __________	$ __________
Food (excluding meals provided at the conference):	$ __________	$ __________
Other (please specify):	$ __________	$ __________
Total Expenses	**$ __________**	**$ __________**

*Please note that U.S. addresses will be reimbursed in U.S. dollars unless otherwise requested.

Claimant's Signature: ______________________ **Date:** ______________

Please note: Queen's University pays 40 cents per kilometer for those who drive. Meal allowances are: $10.00 for breakfast, $13.00 for lunch and $27 for dinner. If submitting an air fare ticket, please include the invoice from the travel agent, or fax for electronic tickets, as well as the boarding passes. **Please send original receipts.**

For reimbursement, send your original receipts to: Mrs. Sharon Sullivan
John Deutsch Institute
Dunning Hall, Room 216C
Kingston, Ontario K7L 3N6

If you have any questions, please contact Sharon Sullivan at (613) 533-2294,
e-mail: sullivas@qed.econ.queensu.ca

A - Hochelaga Inn
24 Sydenham St. South, 877-933-9433 ; 613-549-5534

B - Dunning Hall
corner of Union & University, (Rm 27) 533-2250

C - Grizzly Grill Thursday after 8:00 PM
395 Princess Street 544-7566

D - Pan Chancho, Friday 6:30 PM
44 Princess Street 544-7790

E - Chris's house, Saturday evening,
123 Patrick St. (at Pine St.) 545-9835

Path on the map is along Clergy Street, through the park and then down the hill on Patrick St.

ACCOMMODATION AND TRAVEL

Accommodation Rooms have been set aside for Thursday-Saturday night under "NIEPA Workshop" at: The Hochelaga Inn. Please make your own reservation by calling either 1-877-933-9433, or by e-mail stay@hochelagainn.com. The hotel is a 10 minute walk to Dunning Hall and a 10 minute walk to downtown locations. Breakfast is included, as well as parking and wireless internet.

Travel The most important travel note is that *flying to Kingston is only possible and affordable if you* ***fly to Toronto on Air Canada****. This is often* ***not*** *the cheapest and most convenient way to get here.*

Getting Close to Kingston For U.S. travelers, two common alternatives to flying through Toronto are:

- Fly to Syracuse and rent a car (about 2 hours driving time notwithstanding Yahoo Map estimate below)
- Fly to Montreal and take the train to Kingston (about 3 hours including transit to Dorval station).

In general, Kingston is 2 to 3 hours by ground from four major airports.

Direction/Airport	Get to Kingston by
West: Toronto (Pearson)	[Air] [Rail] [Car] [Bus] [[Door-to-Door Shuttle]
North: Ottawa	[Rail] [Car] [Door-to-Door Shuttle]
East: Montreal (Dorval/Trudeau)	[Rail] [Car] [Door-to-Door Shuttle]
South: Syracuse	[Car] [Door-to-Door Shuttle]

Getting to Kingston

- *Rail - Via Train service to Kingston:* *schedule and booking*

  When traveling by rail, connections to/from airports:
  - *Toronto:* 30 minute taxi/shuttle to/from Union Station downtown
    booking a door-to-door shuttle is often better
  - *Ottawa:* 10 minute taxi to/from Central Station downtown
  - *Montreal:* 10 minute shuttle bus or taxi to/from Dorval Station (NOT Central Station)
    If the train schedule matches up well your flight, this is the best way to get to Kingston.

  When reading Via PDF schedules remember that "x7" in a column means "excluding Sunday".

- *Car Rental - Driving Times and Directions from Airports:*

  NB: Actual speed on Highway 401 is 120kmh not the posted 100kmh, so travel times are less than the calculated times.
  [Toronto] [Ottawa] [Montreal] [Syracuse]

- *Bus*

  Runs directly from Pearson Airport Toronto to Queen's (10 minutes walk to the hotel) 3 times a day. Check the schedules. If the schedule works well for your flights then taking the bus instead of a door-to-door shuttle in one direction would be appreciated.

- *Door-to-door Shuttles Services*

  If you book a shuttle please let me know so that we can try to split the fixed cost with other travelers.

  - Alcorn Shuttle
  - ExecuTrans (e-mail or 613-384-8412)

Home | Rates & Rooms | History & Pictures | Location
Packages | Conference | Reservations | Local Attractions

Welcome to the Hochelaga Inn!

Welcome to the Hochelaga Inn! When you are in Kingston, be sure to stay in the Limestone City's finest historic bed and breakfast. The Hochelaga Inn is one of Kingston's best and most recognizable historic inns. We pride ourselves on giving our guests first class service and comfort in an idyllic setting. We are situated in Kingston's historic Sydenham ward, just steps from Queen's University, and Kingston's beautiful waterfront. The Hochelaga Inn boasts round the clock front desk service, complimentary parking and breakfast, and is in close proximity to all of Kingston's major downtown attractions. Our friendly, accommodating staff will do everything to make certain that your stay in Kingston, whether it be for business or pleasure, is a memorable one. If you choose the historic Hochelaga Inn, your are choosing the intimacy and luxury of a privately owned bed and breakfast without sacrificing the convenience of a large hotel.

Website by: **Sugarman Design**

Discover Canada Map

- Ways to save
- General information
- Special offers
- Advance purchases
- Flexi Fares
- Commuter Pass
- Corridorpass
- Canrailpass
- North America Rail Pass
- BizPak
- Group travel
- AirConnect to Dorval
- Air Miles
- Plan your journey
- Schedules
- Fares and Reservations
- Our stations
- Your baggage
- Your special needs
- Gift certificates
- Aboard the train
- Meals on board
- Vacation packages
- Travel credit
- Info. for non-Canadians
- Questions, comments and other information
- Gift Shop, Souvenirs
- Our trains
- All our trains
- Rockies and Pacific
- Prairies and northern Manitoba
- Ontario
- Québec
- Atlantic region
- Photo albums
- Our classes

Yahoo! Canada Mail

Yahoo! Canada Maps

Welcome, **ferrallc**
[Sign Out, My Account]

Search the Web | Search

Maps Home - Help

Yahoo! Driving Directions

Maps | **Driving Directions**

MY YAHOO!

You are currently using **Netscape 4.0**. For the best experience using Yahoo! Canada Maps we recommend that you upgrade to the latest version of **Netscape**. Get latest version now

Starting from: A Lester B. Pearson International Airport YYZ, Toronto, ON Save Address

Arriving at: B 24 Sydenham St, Kingston, ON Save Address

Distance: 271.3 km **Approximate Travel Time:** 2 hours 40 mins

Get Reverse Directions

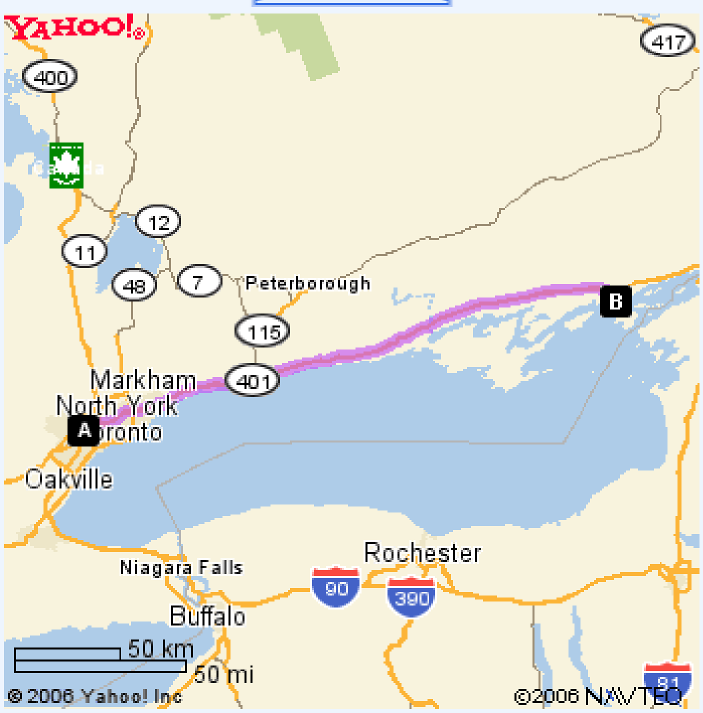
Text Only | Printable Version | Email Directions

Your Full Route

Zoom In - 1st.3city567prov.910- Zoom out

NW N NE

W E



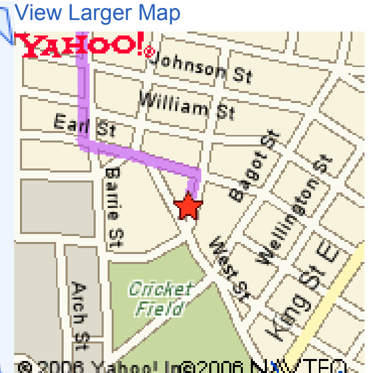
SW S SE

Click on Map to Zoom in & Recentre

Click on Map to Recentre only

Your Destination

View Larger Map



Directions

Show Turn by Turn Maps

1. Start going toward the **AIRPORT EXIT** on Terminal Access Road **1** - go **0.8** km
2. Take ramp onto **ON-409** - go **5.3** km
3. **ON-409** becomes MacDonald-Cartier Fwy - go **0.3** km
4. MacDonald-Cartier Fwy becomes **ON-401** - go **259.5** km
5. Take exit #617/Division St. toward Kingston/Westport - go **0.6** km

6. Turn right on Division St - go **3.9** km
7. Turn left on Princess St[ON-2] - go **0.2** km
8. Turn right on Barrie St - go **0.3** km
9. Turn left on Earl St - go **0.3** km
10. Turn right on Sydenham St - go **< 0.1** km
11. Arrive at **24 SYDENHAM ST, KINGSTON**

When using any driving directions or map, it's a good idea to do a reality check and make sure the route still exists, be awre of construction, and follow traffic safety precautions. Please use this only as an aid in planning.

Get New Driving Directions

A **Enter starting address or select from My Locations**

My Locations Edit

-- My Locations --

Address

YYZ

City, Province, or Postal Code

Toronto, ON

Country

Canada

B **Enter destination address or select from My Locations**

My Locations Edit

-- My Locations --

Address

24 Sydenham St

City, Province, or Postal Code

Kingston, ON

Country

Canada

Get Directions

Maps | **Driving Directions** | **Yellow Pages**

Yahoo! Canada   Mail

Search the Web   Search

YAHOO! CANADA Maps

Welcome, **ferrallc**
[Sign Out, My Account]

Maps Home - Help

Yahoo! Driving Directions

Maps | **Driving Directions**   MY YAHOO!

You are currently using **Netscape 4.0**. For the best experience using Yahoo! Canada Maps we recommend that you upgrade to the latest version of **Netscape**. Get latest version now

Starting from: A Macdonald Cartier International Airport YOW, Ottawa, ON Save Address

Arriving at: B 24 Sydenham St, Kingston, ON Save Address

Distance: 187.0 km **Approximate Travel Time:** 2 hours 7 mins   Get Reverse Directions

Text Only | Printable Version | Email Directions

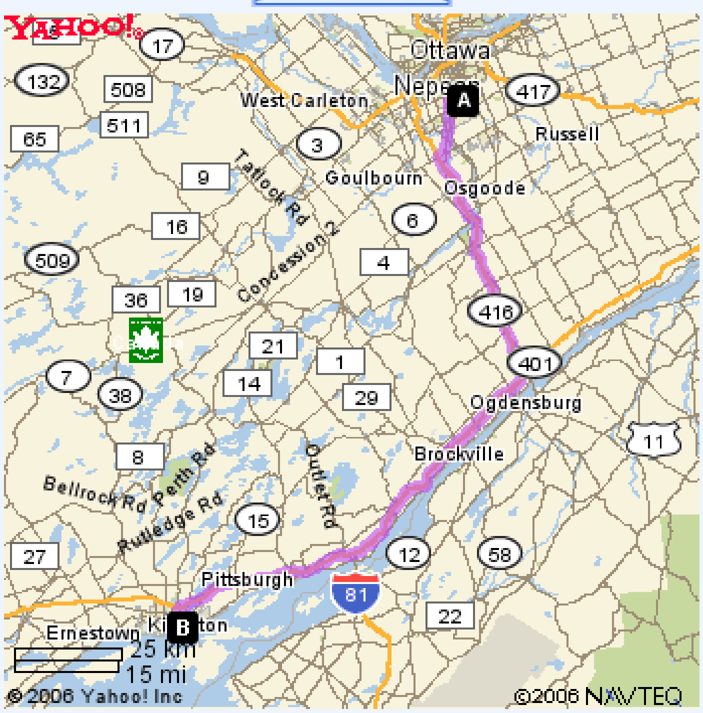
Your Full Route

Zoom In - 1st.3city567prov.910- Zoom out

N

NW   NE

W   E



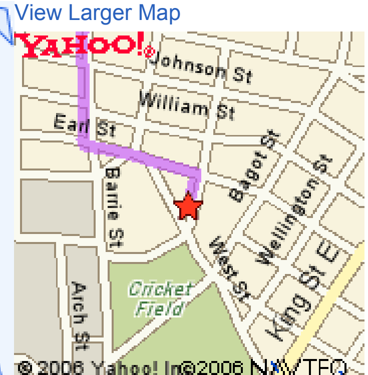
SW   S   SE

Click on Map to Zoom in & Recentre

Click on Map to Recentre only

Your Destination

View Larger Map



Directions

Show Turn by Turn Maps

1. Start going toward the **AIRPORT EXIT** on Airport Pky - go **0.8** km
2. Continue on Ramp - go **0.3** km
3. Continue on Regional Road **107** - go **2.3** km
4. Turn left on Regional Road **32** - go **3.1** km
5. Turn left on Regional Road **73[ON-16]** - go **13.2** km

6. Turn right on Regional Road **8** - go **1.9** km
7. Take ramp onto **ON-416** - go **55.3** km
8. Take (L) ramp onto **ON-401** - go **104.9** km
9. Take exit #617/Division St. toward Kingston/Westport - go **0.3** km
10. Turn right on Division St - go **4.2** km
11. Turn left on Princess St[ON-2] - go **0.2** km
12. Turn right on Barrie St - go **0.3** km
13. Turn left on Earl St - go **0.3** km
14. Turn right on Sydenham St - go **< 0.1** km
15. Arrive at **24 SYDENHAM ST, KINGSTON**

When using any driving directions or map, it's a good idea to do a reality check and make sure the route still exists, be awre of construction, and follow traffic safety precautions. Please use this only as an aid in planning.

Get New Driving Directions

A Enter starting address or select from My Locations	**B** Enter destination address or select from My Locations
My Locations Edit	My Locations Edit
-- My Locations --	-- My Locations --
Address	Address
YOW	24 Sydenham St
City, Province, or Postal Code	City, Province, or Postal Code
Ottawa, ON	Kingston, ON
Country	Country
Canada	Canada

Get Directions

Maps | **Driving Directions** | **Yellow Pages**

You are currently using **Netscape 4.0**. For the best experience using Yahoo! Canada Maps we recommend that you upgrade to the latest version of **Netscape**. Get latest version now

Starting from: A Dorval International Airport YUL, Montreal, QC Save Address

Arriving at: B 24 Sydenham St, Kingston, ON Save Address

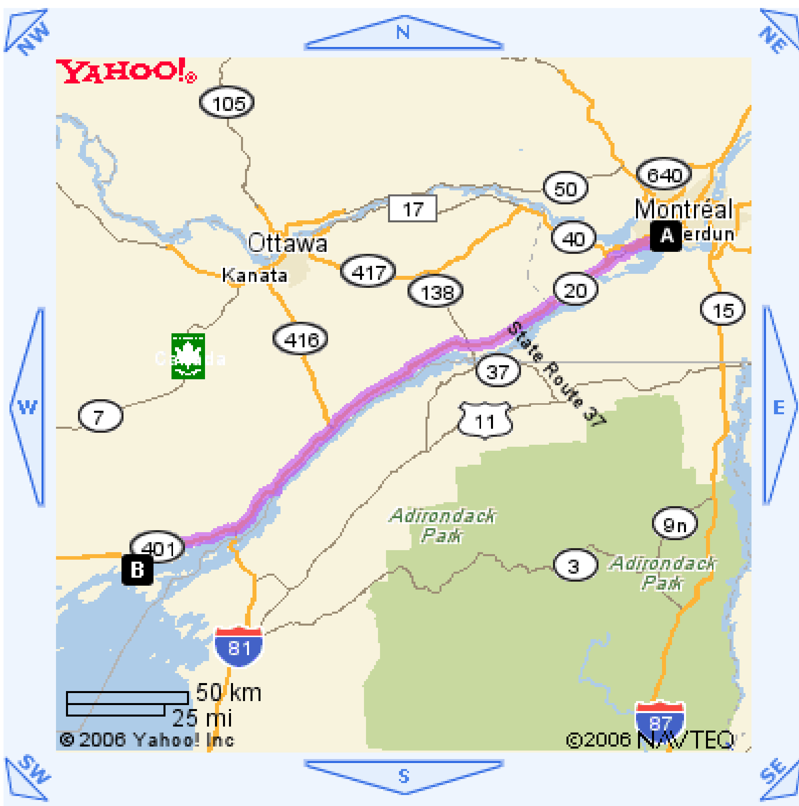
Distance: 272.7 km **Approximate Travel Time:** 2 hours 47 mins

Get Reverse Directions

Text Only | Printable Version | Email Directions

Your Full Route

Zoom In - 1st.3city567prov.910- Zoom out

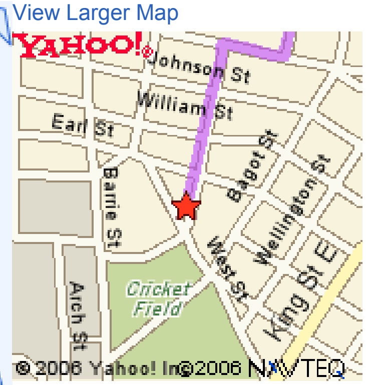


Click on Map to Zoom in & Recentre

Click on Map to Recentre only

Your Destination

View Larger Map



Directions

Show Turn by Turn Maps

1. Start going toward the **AIRPORT EXIT** on Connecting Road - go **< 0.1** km
2. Continue on Roméo-Vachon Rue - go **1.6** km
3. Take ramp onto **QC-520** - go **0.5** km
4. Take ramp onto Montréal-Toronto Boul - go **0.6** km

5. Take L ramp onto **QC-20** - go **55.5** km
6. **QC-20** becomes **ON-401** - go **207.9** km
7. Take exit **#619** toward Montreal St./Kingston/Battersea - go **0.5** km
8. Turn right on Battersea Rd - go **0.3** km
9. Battersea Rd becomes Montreal St - go **5.3** km
10. Turn right on Brock St - go **0.2** km
11. Turn left on Sydenham St - go **0.3** km
12. Arrive at **24 SYDENHAM ST, KINGSTON**

When using any driving directions or map, it's a good idea to do a reality check and make sure the route still exists, be awre of construction, and follow traffic safety precautions. Please use this only as an aid in planning.

Get New Driving Directions

A **Enter starting address or select from My Locations**

My Locations Edit

-- My Locations --

Address

YUL

City, Province, or Postal Code

Montreal, QC

Country

Canada

B **Enter destination address or select from My Locations**

My Locations Edit

-- My Locations --

Address

24 Sydenham St

City, Province, or Postal Code

Kingston, ON

Country

Canada

Get Directions

Yahoo! Canada Mail

YAHOO! CANADA Maps

Welcome, **ferrallc**
[Sign Out, My Account]

Search the Web | Search

Maps Home - Help

Yahoo! Driving Directions

Maps | Driving Directions

MY YAHOO!

You are currently using **Netscape 4.0**. For the best experience using Yahoo! Canada Maps we recommend that you upgrade to the latest version of **Netscape**. Get latest version now

Starting from: A Syracuse Hancock International Airport SYR, Syracuse, NY Save Address

Arriving at: B 24 Sydenham St, Kingston, ON Save Address

Distance: 205.3 km **Approximate Travel Time:** 2 hours 3 mins

Get Reverse Directions

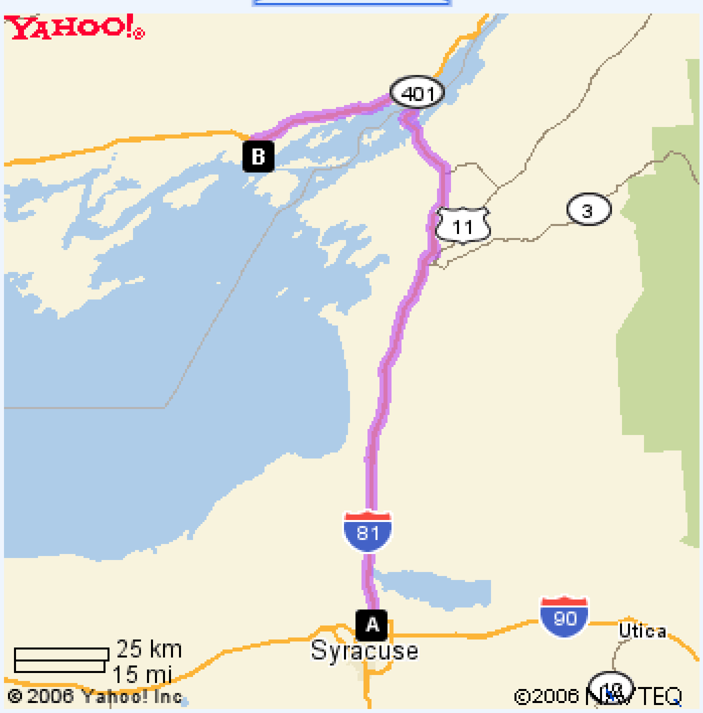
Text Only | Printable Version | Email Directions

Your Full Route

Zoom In - 1st.3city567prov.910- Zoom out

NW N NE

W E



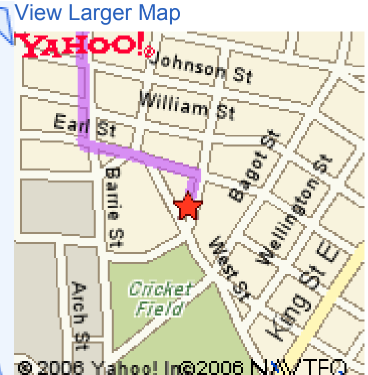
SW S SE

Click on Map to Zoom in & Recentre

Click on Map to Recentre only

Your Destination

View Larger Map



Directions

Show Turn by Turn Maps

1. Start going toward the **AIRPORT EXIT** on Airport Blvd - go **1.4** km
2. Take ramp onto **I-81** - go **150.0** km
3. **I-81** becomes **ON-137** - go **3.9** km
4. Take L ramp onto **ON-401** - go **44.7** km

5. Take exit #617/Division St. toward Kingston/Westport - go **0.3** km
6. Turn right on Division St - go **4.2** km
7. Turn left on Princess St[ON-2] - go **0.2** km
8. Turn right on Barrie St - go **0.3** km
9. Turn left on Earl St - go **0.3** km
10. Turn right on Sydenham St - go **< 0.1** km
11. Arrive at **24 SYDENHAM ST, KINGSTON**

When using any driving directions or map, it's a good idea to do a reality check and make sure the route still exists, be awre of construction, and follow traffic safety precautions. Please use this only as an aid in planning.

Get New Driving Directions

A Enter starting address or select from My Locations	**B** Enter destination address or select from My Locations
My Locations Edit	My Locations Edit
-- My Locations --	-- My Locations --
Address	Address
SYR	24 Sydenham St
City, Province, or Postal Code	City, Province, or Postal Code
Syracuse, NY	Kingston, ON
Country	Country
United States	Canada

Get Directions

Maps | Driving Directions | Yellow Pages

Schedule Details

Toronto Airport, ON to Kingston, ON

(Check for reverse route: Kingston, ON to Toronto Airport, ON)

New Search

Operator Name: **Trentway-Wagar Inc.**
Schedule Details: **Route 2711**
Effective From June 26, 2005 Onward

- **BUY TICKETS ONLINE!**

- View Complete Schedule in PDF
- Holiday Schedule Information
- Important Notes
- Where to Buy / Stop Locations
- Fares

IF YOU ARE DEPARTING FROM ONE OF THE LOCATIONS BELOW YOU WILL NEED TO CALL COACH CANADA AT 1-800-461-7661 TO MAKE A PICKUP RESERVATION.
- Bowmanville
- Cobourg
- Colborne
- Grafton
- Marysville
- Newcastle
- Odessa
- Port Hope

	From: Toronto Airport, ON			→	To: Kingston, ON		
Select Day of Travel Daily	Airport Terminal 3	Airport Terminal 2	Airport Terminal 1		Bus Terminal	Queens University	Route/Trip
Daily	2:30p	2:35p	2:40p	→	6:20p	6:30p	9511
Daily	2:30p	2:35p	2:40p	→	6:20p		9511/762*
Daily	7:00p	7:05p	7:10p	→	10:50p	11:00p	9512
Daily	7:00p	7:05p	7:10p	→	10:50p		9512/782*
Daily	9:30p	9:35p	9:40p	→	1:20a[t]	1:30a[t]	9503

**** Please ask the driver if a transfer is necessary. Transfers may cost more than direct service.***

[t] *Indicates next day.*

We will make every effort to operate the above schedules. However, delays may occur from time to time due to traffic congestion and other unforeseen circumstances. Always allow ample time for situations beyond our control.

ALCORN SHUTTLE & TOURS

Major Airports, Personal & Corporate Travel
561 Armstrong Rd. #215 Kingston, ON K7K 4N8

Tel: **613-547-5412**
Toll Free: **1-888-547-5412**

Email: alcorn@kos.net & & Web Site:: www.alcornlimo.com

Astro Van	Single	Couple	Extra Person	Whole Van	Private Sedan
Toronto	$200.00*	$225.00	$20.00	$300.00	$300.00
Syracuse	$180.00	$200.00	$20.00	$250.00	$250.00
Ottawa	$180.00	$200.00	$20.00	$250.00	$250.00
Dorval	$225.00	$250.00	$20.00	$300.00	$300.00
Mirabel	$250.00	$270.00	$20.00	$325.00	$325.00
Charter Rate	$50.00/hr				

Door To Door Service 2 Hours Before Flight Times.
Luxury Sedan + Mini Vans with TV/DVD
Waiting time: $15.00/hour (Gratuity not included)
***Discount seats available ($100.00/person, Call for details)**

All private Daytrips Available: Niagra Falls, Wonderland, Casinos, Shopping and Special Toronto or Ottawa Tours.

Airport Meeting Points

- **Toronto: Pre-arranged Limo Pickup Area**
- **Syracuse: Bottom of Arrival Escalator**
- **Montreal: Arrival Gates**
- **Ottawa: Arrival Gates**

VISA, Mastercard, Cheques, or Cash
Fully Insured and O.H.T.B approved since 1995

LOCALLY OWNED COMPANY REGISTERED WITH THE GREAT TORONTO AIRPORT AUTHORITY,
FULLY INSURED & O.H.T.B. APPROVED SINCE 1995

CALL YOUR RESERVATION: 613-547-5412
E-MAIL YOUR RESERVATION: alcorn@kos.net

FAX YOUR RESERVATIONS: 547-0638
TEXT YOUR RESERVATIONS: 583-2473

Kingston's Original Shuttle Service since 1995