ISSN 0956-8549-650

Limits of Arbitrage: The State of the Theory

By

Denis Gromb
Dimitri Vayanos

THE PAUL WOOLLEY CENTRE
WORKING PAPER SERIES NO 9
DISCUSSION PAPER NO 650

DISCUSSION PAPER SERIES

March 2010

Denis Gromb is Professor of Finance at INSEAD, where he teaches Corporate Finance. He graduated from the Ecole Polytechnique, Paris, where he also obtained a PhD in Economics in 1994 while being a visiting student at the London School of Economics. His principal research interests include Corporate Finance, Corporate Governance, Banking, Limited Arbitrage and the Economics of Organizations. His articles have been published in leading academic journals in Economics and Finance. He has previously been on the faculty at MIT's Sloan School of Management (1995-2001) and London Business School (2001-2008) and visited the London School of Economics (2007-2008). He has also taught doctoral level courses at a variety of institutions. Dimitri Vayanos is Professor of Finance at the London School of Economics. He received his undergraduate degree from Ecole Polytechique in Paris and his PhD from MIT in 1993. Prior to joining the LSE, he was faculty member at Stanford and MIT. Vayanos has published in a number of leading economics and finance journals on topics such as: liquidity and asset pricing, limited arbitrage, delegated portfolio management, information in markets and organizations, market microstructure, and behavioural economics. He is the Director of LSE's Paul Woolley Centre for the Study of Capital Market Dysfunctionality, the Director of the CEPR Financial Economics program, and a Research Associate at NBER.

Limits of Arbitrage: The State of the Theory

Denis Gromb*
INSEAD and CEPR

Dimitri Vayanos†
LSE, CEPR and NBER

March 8, 2010‡

Abstract

We survey theoretical developments in the literature on the limits of arbitrage. This literature investigates how costs faced by arbitrageurs can prevent them from eliminating mispricings and providing liquidity to other investors. Research in this area is currently evolving into a broader agenda emphasizing the role of financial institutions and agency frictions for asset prices. This research has the potential to explain so-called "market anomalies" and inform welfare and policy debates about asset markets. We begin with examples of demand shocks that generate mispricings, arguing that they can stem from behavioral or from institutional considerations. We next survey, and nest within a simple model, the following costs faced by arbitrageurs: (i) risk, both fundamental and non-fundamental, (ii) short-selling costs, (iii) leverage and margin constraints, and (iv) constraints on equity capital. We finally discuss implications for welfare and policy, and suggest directions for future research.

Keywords: limits of arbitrage, market anomalies, liquidity, financial constraints, financial institutions, survey

---

*`denis.gromb@insead.edu`

†`d.vayanos@lse.ac.uk`

‡Article prepared for the Second Volume of the Annual Review of Financial Economics. We thank Suleyman Basak, Bruno Biais, Bernard Dumas, Vincent Fardeau, Robin Greenwood, Harald Hau, Peter Kondor, Arvind Krishnamurthy, Anna Pavlova, Christopher Polk, Ana-Maria Santacreu, Yuki Sato and Pierre-Olivier Weill for helpful comments, and the Paul Woolley Centre at the LSE for financial support.

1 INTRODUCTION

Standard models of asset pricing assume a representative agent who participates in all markets costlessly. Equilibrium prices in these models are tied to the representative agent's consumption, which coincides with the aggregate consumption in the economy. The relationship between prices and consumption is summarized in the consumption CAPM, according to which an asset's expected return in excess of the riskfree rate is proportional to the asset's covariance with aggregate consumption. Intuitively, assets that correlate positively with consumption add to the risk borne by the representative agent and must offer high expected return as compensation.

The relationship between risk and expected return predicted by standard models appears to be at odds with a number of stylized facts commonly referred to as market anomalies. Leading anomalies include (i) short-run momentum, the tendency of an asset's recent performance to continue into the near future, (ii) long-run reversal, the tendency of performance over a longer history to revert, (iii) the value effect, the tendency of an asset's ratio of price to accounting measures of value to predict negatively future returns, (iv) the high volatility of asset prices relative to measures of discounted future payoff streams, and (v) post-earnings-announcement drift, the tendency of stocks' earning surprises to predict positively future returns.[1] Reconciling these anomalies with standard models requires explaining variation in asset risk: for example, in the case of short-run momentum, one would have to explain why good recent performance renders an asset riskier and more positively correlated with aggregate consumption. A recent literature pursues explanations along these lines by introducing more general utility functions for the representative agent. Yet, reconciling standard models with all the anomalies, and in a way consistent with their quantitative magnitude remains elusive.

The anomalies listed above concern the predictability of asset returns based on past prices and earnings. An additional set of anomalies concern the relative prices of assets with closely related payoffs. For example, (i) "Siamese-twin" stocks, with claims to almost identical dividend streams, can trade at significantly different prices, (ii) stocks of a parent and a subsidiary company can trade at prices under which the remainder of the parent company's assets has negative value, and (iii) newly issued "on-the-run" bonds can trade at significantly higher prices than older "off-the-run" bonds with almost identical payoffs.[2] Anomalies concerning relative prices have been documented

[1]See, for example, Jegadeesh & Titman (1993) for evidence on short-run momentum, DeBondt & Thaler (1985) for long-run reversal, Fama & French (1992) for the value effect, LeRoy & Porter (1981) & Shiller (1981) for excess volatility, and Bernard & Thomas (1989) for post-earnings-announcement drift. See also the surveys by Fama (1991) and Schwert (2003).

[2]See, for example, Rosenthal & Young (1990) and Dabora & Froot (1999) for evidence on Siamese-twin stocks,

for a more limited set of assets, partly because of the scarcity of asset pairs with closely related payoffs. At the same time, these anomalies are particularly hard to reconcile with standard models. Indeed, while standard models may offer slightly different predictions as to how risk and expected returns are related, they all imply the law of one price, i.e., assets with identical payoffs must trade at the same price. In the previous examples, however, differences in payoffs appear to be insignificant relative to the observed price differences.

Understanding why anomalies exist and are not eliminated requires a careful study of the process of arbitrage: who are the arbitrageurs, what are the constraints and limitations they face, and why arbitrage can fail to bring prices close to the fundamental values implied by standard models. This is the focus of a recent literature on the limits of arbitrage. This article surveys important theoretical developments in that literature, nests them within a simple model, and suggests directions for future research.

Limits of arbitrage are commonly viewed as one of two building blocks needed to explain anomalies. The other building block are demand shocks experienced by investors other than arbitrageurs. Anomalies are commonly interpreted as arising because demand shocks push prices away from fundamental values and arbitrageurs are unable to correct the discrepancies. Such "non-fundamental" shocks to demand are often attributed to investor irrationality. In this sense, research on the limits of arbitrage is part of the behavioral finance agenda to explain anomalies based on investors' psychological biases.[3]

This article departs from the conventional view in two related respects. First, it argues that research on the limits of arbitrage is relevant not only for behavioral explanations of anomalies but also for the broader study of asset pricing. Indeed, psychological biases are not the only source of non-fundamental demand shocks: such shocks can also arise because of institutional frictions relating to contracting and agency, as the examples in the next section show. Research on the limits of arbitrage characterizes how non-fundamental demand shocks, whether behavioral or not, impact prices.

According to the conventional view, non-fundamental demand shocks concern investors other than arbitrageurs, and therefore can be understood independently of the limits of arbitrage. Our second departure is to argue that many non-fundamental demand shocks can be understood jointly with limits of arbitrage within a setting that emphasizes financial institutions and agency. Indeed,

Mitchell et al. (2002) and Lamont & Thaler (2003) for the relative pricing of parent and subsidiary companies, and Amihud & Mendelson (1991), Warga (1992) and Krishnamurthy (2002) for the on-the-run phenomenon.

[3] Behavioral explanations for the anomalies include Barberis et al. (1998), Daniel et al. (1998), Hong & Stein (1999) and Barberis & Shleifer (2003). See also the survey by Barberis & Thaler (2003).

arbitrage is often performed by specialized institutions such as hedge funds and investment banks, and the trading strategies of these institutions are constrained by agency frictions. At the same time, financial institutions and agency frictions are the source of many non-fundamental demand shocks. In this sense, financial institutions do not necessarily correct anomalies, but can also cause them. Research on the limits of arbitrage is currently evolving into a broader agenda emphasizing the role of financial institutions and agency frictions for asset prices. This agenda has the potential to offer a unified explanation of many anomalies.

The emphasis on financial institutions and agency frictions is fruitful for the analysis of welfare and public policy. Crises, including the recent one, show that government intervention can be important for the smooth functioning of financial markets. In standard models, however, there is no scope for such intervention because the equilibrium is Pareto optimal. Research on the limits of arbitrage has the potential to deliver a more useful framework for designing and assessing public policy. Indeed, this research takes a two-tiered view of financial markets: a core of sophisticated arbitrageurs trade against mispricings, and in doing so provide liquidity to a periphery of less sophisticated investors. Under this view, the financial health of arbitrageurs is crucial for the smooth functioning of markets and the provision of liquidity. Understanding how financial health is affected by arbitrageurs' trading decisions, and whether these decisions are socially optimal, can guide public policy.

This article proceeds as follows. Section 2 presents examples of non-fundamental demand shocks, emphasizing that they often stem from institutional considerations. Section 3 surveys important theoretical developments in the literature on the limits of arbitrage, and nests them within a simple model. It emphasizes the following costs faced by arbitrageurs: (i) risk, both fundamental and non-fundamental, (ii) costs of short-selling, (iii) leverage and margin constraints, and (iv) constraints on equity capital.[4] While these are not the only costs faced by real-life arbitrageurs, they are among the most important and have received significant attention in the literature. Besides examining the implications of each type of cost for asset price behavior, Section 3 sketches how these costs can be integrated into richer models that incorporate multiple assets and dynamics. Such models have the potential to address a variety of anomalies—both of the type concerning violations of the law of one price, and of the type concerning return predictability—and are the subject of a rapidly growing literature. Finally, Section 4 discusses implications for welfare and public policy.

[4] Risk is a cost when arbitrageurs are not fully diversified and bear a disproportionate share of the risk of arbitrage trades. Undiversification is related to financial constraints, which we treat separately.

2 DEMAND SHOCKS

Models of limited arbitrage typically assume that some investors experience demand shocks that can drive prices away from fundamental values. This section presents examples of such shocks and their effects on prices. Shocks in the first example are probably best interpreted as arising from behavioral considerations, while in the other examples they more likely arise from institutional frictions.

2.1 Palm-3Com

Lamont & Thaler (2003) study the sale of Palm by its parent company 3Com and the behavior of the two companies' stock prices around this event. On March 2, 2000, 3Com sold 5% of its stake in Palm through an initial public offering (IPO). 3Com also announced that it would spin off its remaining stake in Palm to 3Com shareholders before the end of the year. Under the terms of the spin off, 3Com shareholders would receive 1.525 shares of Palm for each share of 3Com they owned.

The law of one price implies that prior to the spin off, 3Com shares should have been trading at a price exceeding 1.525 times the price of Palm shares. This is because one 3Com share was equivalent to 1.525 shares of Palm plus an equity claim to 3Com's remaining (non-Palm) assets, and the latter claim had non-negative value because of limited liability. The law of one price was, however, violated for a period of approximately two months starting from the IPO. For example, on the day of the IPO, Palm closed at \$95.06 per share, implying a lower bound of $1.525 \times 95.06 = \$145$ for the share price of 3Com. 3Com, however, closed at \$81.81 per share, having dropped from \$104.13 on the previous day. Under these prices the implied value of 3Com's non-Palm assets was -\$22 billion, implying an economically significant violation of the law of one price.

Explaining the mispricing between Palm and 3Com is a challenge to standard models of asset pricing. One must explain, in particular, why some investors were willing to buy Palm shares for \$95.06, while they could acquire them at a lower cost by buying 3Com shares. The most plausible explanations are based on investors' psychological biases and cognitive limitations. Investors buying Palm were possibly not sophisticated enough to appreciate the opportunity of buying it through 3Com. Moreover, because Palm was a manufacturer of a relatively new product (handheld devices), it was possibly associated with the "new economy" to an extent higher than 3Com. This might have led investors overly optimistic about the new economy to be willing to pay a disproportionately high price for Palm.

2.2 Index Effects

Index effects stem from a stock's addition to or deletion from prominent market indices. Starting with Harris & Gurel (1986) and Shleifer (1986), a number of papers document that addition to the Standard and Poor (S&P)'s 500 index raises the price of a stock, while deletion lowers its price. For example, Chen et al. (2004) find that during 1989-2000, a stock's price would increase by an average 5.45% on the day of the announcement that the stock would be added to the index, and a further 3.45% from the announcement day to the day of the actual addition. Conversely, a stock's price would decrease by an average 8.46% on the day of the announcement that the stock would be deleted from the index, and a further 5.97% from the announcement day to the deletion day. These effects are economically significant.

Standard models can account for index effects only if additions and deletions convey information about assets' fundamental values. Even if one is to accept, however, that S&P has an informational advantage relative to the market, it is hard to explain why this advantage (i) grew larger after 1989, which is when index effects became the most significant, and (ii) can be so large to account for the observed index effects. A more plausible explanation is that index additions and deletions trigger changes in the demand by mutual funds. Indeed, passively managed mutual funds track indices mechanically, while actively managed funds benchmark their performance against indices. If, therefore, a stock is added to the S&P500 index, funds that track or are benchmarked against the index are eager to buy the stock, and this can raise the stock's price. The institutional explanation is consistent with the growth of index effects in recent decades since this parallels the growth of institutional investing, index tracking and benchmarking.

Boyer (2007) provides further evidence consistent with the institutional explanation. He focuses on the BARRA value and growth indices, which consist of value and growth stocks, respectively. Unlike the S&P500 index, BARRA indices are constructed using publicly disclosed rules, so additions and deletions do not signal any private information. Boyer finds that "marginal value" stocks, defined as those that just switched from the growth into the value index, comove significantly more with the value than with the growth index, while the opposite is true for "marginal growth" stocks. These effects are hard to explain within standard models because marginal value stocks have very similar characteristics to marginal growth stocks. On the other hand, it is plausible that these effects arise from shifts in demand by mutual funds. For example, inflows into funds tracking the value index trigger purchases of all stocks in that index, and this can raise the prices of these stocks simultaneously. The institutional explanation is further strengthened by Boyer's finding that the effects appear only after 1992, which is when the BARRA indices were introduced.[5]

2.3 Fire Sales by Mutual Funds

Mutual funds respond to outflows by selling stocks in their portfolios. Coval & Stafford (2007) study the behavior of stock prices around sales driven by large outflows. They define fire sales as the sales by the 10% of funds that experience the largest outflows within a given quarter. For each stock, they compute the fraction of average volume generated by fire sales, and they focus on the 10% of stocks for which this fraction is largest. These stocks exhibit a V-shaped price pattern. During the fire-sale quarter and the quarter immediately preceding, their average cumulative abnormal return is -7.9%. This price decline is followed by a recovery: during the year following the fire-sale quarter, the average cumulative abnormal return is 6.1%. This return rises to 9.7% during the 18 months following the fire-sale quarter.[6]

The slow and predictable price recovery is a challenge to standard models. Indeed, standard models can account for an increase in expected return only through an increase in the covariance with aggregate consumption. Explaining why this covariance increases for stocks sold by distressed mutual funds, and why such an increase can account for an annual abnormal return of 6% is difficult. A more plausible explanation is that sales by distressed mutual funds generate price pressure, pushing prices below fundamental values and raising expected returns going forward. This explanation is related to institutional frictions: distressed sales are likely to be triggered by investors who lose confidence in the quality of managers running underperforming funds.

2.4 U.K. Pension Reform and the Term Structure

Demand shocks are likely to also affect prices outside the U.S. and for assets other than stocks, as our last example illustrates. It concerns the impact of the U.K. pension reform on the term structure of interest rates, an episode described in greater detail in Greenwood & Vayanos (2010a). A major objective of U.K. pension reform over the past twenty years has been to ensure the transparency and solvency of pension funds; indeed, the reform was motivated partly by the failure of the Maxwell pension fund in the early 1990s. The reform stipulated that pension funds had to meet a minimum ratio of assets to liabilities. Pension-fund assets, such as stocks and bonds, are publicly traded and

[5] Earlier evidence linking index membership to comovement is in Vijh (1994) and Barberis et al. (2005). Both papers focus on the S&P 500 index.

[6] Related evidence suggesting that fund outflows have large price effects is in Anton & Polk (2008), Jotikasthira et al. (2009), Greenwood & Thesmar (2009) and Lou (2009).

can be valued using market prices. On the other hand, pension-fund liabilities are not traded, and their valuation requires a suitable discount rate. Under the Pensions Act of 2004, this discount rate had to be the yield on long-term inflation-linked government bonds, on the grounds that pension liabilities are long term and indexed to inflation.

Pension funds responded to the reform by buying large amounts of long-term bonds. Indeed, because long-term bonds were providing the discount rate to calculate the value of pension liabilities, they were also the best hedge for these liabilities. Pension-fund purchases had a significant impact on the term structure, especially at the long end. For example, in late 2003, the inflation-indexed bonds maturing in 2016 and 2035 had approximately the same yield. During 2004 and 2005, however, the yield of the 2035 bond fell relative to that of the 2016 bond, with the spread reaching an all-time low of -0.49% in January 2006. At that time, the 2035 and 2055 bonds had yields of 0.72% and 0.48%, respectively, which are extremely low relative to the historical average of 3% of long real rates in the U.K. In accordance with the generally-held view that long yields had decreased because of demand by pension funds, the government agreed in 2005 to issue bonds with maturities of up to 50 years, while also shifting the overall mix of maturities towards the long term.

The inversion at the long end of the U.K. term structure is hard to rationalize within standard representative-agent models. Indeed, in these models the interest rate for maturity T is determined by the willingness of the representative agent to substitute consumption between times 0 and T. Therefore, these models would attribute the drop in the 30-year interest rate to the unlikely scenario that the pension reform signalled a drop in aggregate consumption 30 years into the future. In a similar spirit, the expectations hypothesis of the term structure would attribute the drop in the 2035-2016 yield spread to expectations about short-term interest rates past 2016 decreasing sharply during 2004 and 2005. A more plausible explanation is that the reform triggered high demand for long-term bonds by pension funds, and that generated price pressure.

2.5 Summary

The examples in this section describe a variety of demand shocks that had significant and long-term price effects. A natural question is why arbitrageurs are unable to absorb such shocks and bring prices back to fundamental values. For example, why don't arbitrageurs eliminate the abnormally high expected returns following fire sales by buying the stocks in question? And why were arbitrageurs unable to eliminate the Palm-3Com mispricing by shorting Palm and buying 3Com? Using a simple model, we next examine the constraints faced by arbitrageurs and the implications for asset prices.

3 A SIMPLE MODEL

3.1 Cross-Asset Arbitrage and Intertemporal Arbitrage

We consider an economy where assets are traded in Period 1 and pay off in Period 2. The riskless rate is exogenous and equal to zero. There are two risky assets, A and B, paying off d_A and d_B, respectively. We denote by $\bar{d}_i$ and σ_i, respectively, the mean and standard deviation of d_i, $i = A, B$, and by ρ the correlation between d_A and d_B. For tractability, we assume that d_A and d_B are jointly normal, and modify this assumption in Section 3.4. The price p_B of asset B is exogenous and equal to the asset's expected payoff $\bar{d}_B$. The price p_A of asset A is endogenously determined in equilibrium. Our focus is on how shocks to the demand for asset A affect p_A.

There are two types of agents: outside investors and arbitrageurs. Outside investors' demand for asset A is inelastic and equal to u shares. We refer to u as the demand shock: it is a constant parameter in this section, but becomes a stochastic shock in Section 3.2 where we introduce an initial Period 0. Arbitrageurs are competitive, risk averse and maximize expected utility of wealth W_2 in Period 2. For tractability, we assume that utility is exponential. By possibly reinterpreting the demand shock u as net demand, we normalize the supply of asset A to zero. Under this normalization, the price of asset A in the absence of the demand shock ($u = 0$) equals the asset's expected payoff $\bar{d}_A$.

A demand shock $u \neq 0$ can push the price of asset A away from the expected payoff. Arbitrageurs trade to profit from this discrepancy. Doing so, they also provide liquidity to outside investors. Suppose, for example, that u is positive, in which case outside investors wish to buy asset A. Arbitrageurs provide liquidity because they take the opposite side of this transaction, shorting the asset and limiting the price rise. Liquidity is high when the demand shock's price impact is small: if, for example, u is positive, high liquidity means that the price rise is small.

The model has two interpretations, capturing different but closely related real-life arbitrage situations. In the first "cross-asset arbitrage" interpretation, assets A and B are different and arbitrageurs use asset B to hedge their position in asset A. In the second "intertemporal arbitrage" interpretation, arbitrageurs exploit discrepancies between the prices of the same asset at different points in time. In that interpretation, the arbitrageurs' positions in assets A and B represent trades that arbitrageurs execute in the same asset in different periods. The two interpretations yield the same basic insight: the price effects of demand shocks depend on the risk aversion of arbitrageurs and on the risk that they cannot hedge away.

Cross-Asset Arbitrage

In Period 1, the arbitrageurs choose positions x_A and x_B in assets A and B to maximize expected utility

$$-E_1\left[\exp(-\alpha W_2)\right] \tag{1}$$

subject to the budget constraint

$$W_2 = W_1 + x_A(d_A - p_A) + x_B(d_B - p_B), \tag{2}$$

where α denotes the risk aversion of arbitrageurs and W_1 their wealth in Period 1.[7] Substituting (2) into (1), and using normality and the assumption that $p_B = \bar{d}_B$, we find that arbitrageurs maximize the mean-variance objective

$$x_A(\bar{d}_A - p_A) - \frac{\alpha}{2}\left(x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_Ax_B\rho\sigma_A\sigma_B\right). \tag{3}$$

The optimal demand for asset B is $x_B = -(\rho\sigma_A/\sigma_B)x_A$, i.e., arbitrageurs choose a position in asset B to hedge that in asset A. The optimal demand for asset A is

$$x_A = \frac{\bar{d}_A - p_A}{\alpha\sigma_A^2(1-\rho^2)}. \tag{4}$$

Since asset A is in zero supply, the market-clearing condition is

$$x_A + u = 0. \tag{5}$$

Combining (4) and (5), we find the equilibrium price of asset A:

$$p_A = \bar{d}_A + \alpha\sigma_A^2(1-\rho^2)u. \tag{6}$$

Eq. (6) highlights a number of properties. First, larger demand shocks have a larger price impact ($\partial p_A/\partial u > 0$). Second, a given demand shock u has a larger price impact when arbitrageurs are

[7] Our analysis of cross-asset arbitrage is related to Wurgler & Zhuravskaya (2002). As in our model, they take the price of asset B to be exogenous and equal to the asset's expected payoff. Unlike in our model, however, they restrict arbitrageurs' aggregate dollar investment in assets A and B to be zero.

more risk averse (large α) and asset A has a more uncertain payoff (large σ_A) because in both cases arbitrageurs require more compensation to bear its risk. The shock's impact is also larger when the payoff of asset A is less correlated with that of asset B (small $|\rho|$) because arbitrageurs are less able to hedge their position in asset A using asset B. Thus, assets with higher idiosyncratic risk and fewer substitutes are more sensitive to demand shocks. Note that in the extreme case where assets A and B have identical payoffs ($d_A = d_B$, i.e., $\rho = 1$), the demand shock has no effect and the two assets trade at the same price. Arbitrageurs are able to align the price of asset A fully with that of asset B because they bear no risk in exploiting price discrepancies between the two assets.

Intertemporal Arbitrage

In cross-asset arbitrage, arbitrageurs exploit discrepancies between the prices of two assets at a given point in time. We next consider intertemporal arbitrage, where arbitrageurs exploit discrepancies between the prices of the same asset at different points in time. To interpret our model as one of intertemporal arbitrage, we split Period 1 into two subperiods, and assume that positions in assets A and B represent trades in the same asset in the first and second subperiods, respectively. The model so derived is a simplified version of Grossman & Miller (1988).

We denote by 1 and 1^+, respectively, the first and second subperiod of Period 1, by $d \equiv d_A = d_B$ the common payoff of assets A and B, by $\bar{d}$ and $\bar{d}^+$ the expectation of d as of subperiod 1 and 1^+, respectively, and by σ and σ^+ the respective standard deviation of d. The expectation $\bar{d}^+$ is random as of subperiod 1 if new information arrives between subperiods 1 and 1^+, and has conditional variance $\sigma^2 - (\sigma^+)^2$.[8]

Since in subperiod 1^+ arbitrageurs can trade asset B, and at a price equal to its expected payoff, they are not compensated for bearing risk. Therefore, their aggregate position in subperiod 1^+ is zero, which means that their trade in subperiod 1^+, i.e., in asset B, is the opposite of that in subperiod 1, i.e., in asset A. Using $d_A = d_B = d$, $p_B = \bar{d}^+$ and $x_A = -x_B$, we can write (2) as

$$W_2 = W_1 + x_A(\bar{d}^+ - p_A). \tag{7}$$

Substituting (7) into (1), we find that arbitrageurs choose their position in asset A in subperiod 1 to maximize the mean-variance objective

$$x_A(\bar{d} - p_A) - \frac{\alpha}{2} x_A^2 \left[\sigma^2 - (\sigma^+)^2\right].$$

[8]This follows from $Var(d|1) = E\left[Var(d|1^+)|1\right] + Var\left[E(d|1^+)|1\right]$.

The optimal demand for asset A is

$$x_A = \frac{\bar{d} - p_A}{\alpha\left[\sigma^2 - (\sigma^+)^2\right]}. \tag{8}$$

Combining (8) with the market-clearing condition (5), we find the equilibrium price of asset A:

$$p_A = \bar{d} + \alpha\left[\sigma^2 - (\sigma^+)^2\right] u. \tag{9}$$

The demand shock has a larger price impact when arbitrageurs are more risk averse (large α) and bear more risk. The relevant risk is that between subperiods 1 and 1^+, and is measured by $\sigma^2 - (\sigma^+)^2$. This is because arbitrageurs offset their risky position in subperiod 1 by an opposite position in subperiod 1^+. In the extreme case where subperiods 1 and 1^+ coincide, $\sigma = \sigma^+$ and the demand shock has no effect. As Grossman & Miller emphasize, the time between subperiods 1 and 1^+ can be interpreted as the time it takes for enough risk-bearing capacity to arrive in the market and fully eliminate the effect of the demand shock.[9] The two interpretations of our model correspond to different real-life arbitrageur situations and, accordingly, to different strands of empirical studies. From an asset-pricing theory viewpoint, however, they are isomorphic. From now on, we focus on the cross-asset arbitrage interpretation of the model, and enrich it in ways that illustrate developments in the literature on the limits of arbitrage. The results we derive carry through to the intertemporal arbitrage interpretation.

3.2 Non-Fundamental Risk

In our baseline model, the risk borne by arbitrageurs is "fundamental risk," arising from asset payoffs. An additional type of risk stems from demand shocks when these affect prices. DeLong et al. (DSSW 1990) label this type of risk "noise-trader risk." We use instead the term "non-fundamental risk" to emphasize that while demand shocks may be unrelated to asset payoffs, they can be generated from rational behavior as the examples in Section 2 illustrate. To introduce non-fundamental risk in our model, we assume that the variables $\bar{d}_A$, $\bar{d}_B$ and u, which are constant parameters as of Period 1, are random as of an initial Period 0. Fundamental risk in Period 0 arises because prices in Period 1 depend on $\bar{d}_A$ and $\bar{d}_B$. Non-fundamental risk arises because asset A's price in Period 1 depends on u. We assume that fundamental risk in Period 0 is the same as in

[9] A recent literature derives the slow arrival of new investors from search costs. See, for example, Weill (2007), Duffie et al. (2008), Duffie & Strulovici (2009) and Lagos et al. (2009). See also He and Xiong (2008), who derive capital immobility and segmentation from agency frictions: preventing traders from moving across assets can provide their employer with a better signal of their effort.

Period 1, i.e., the standard deviation of $\bar{d}_i$, $i = A, B$, is σ_i, and the correlation between $\bar{d}_A$ and $\bar{d}_B$ is ρ. We denote the standard deviation of u by σ_u, and assume that u is independent of $\bar{d}_A$ and $\bar{d}_B$.

Eq. (6) implies that the standard deviation, as of Period 0, of asset A's price in Period 1 is

$$\sigma_A \sqrt{1 + \alpha^2 \sigma_A^2 (1 - \rho^2)^2 \sigma_u^2}. \tag{10}$$

Eq. (10) shows that non-fundamental risk σ_u increases asset A's volatility. The effect is through the second term inside the square root, which is the ratio of non-fundamental variance $(\alpha^2 \sigma_A^4 (1-\rho^2)^2 \sigma_u^2)$ to fundamental variance (σ_A^2). This ratio is larger when the demand shock is less predictable (large σ_u) and when a given demand shock u has a larger price impact in Period 1. Consistent with (6), price impact is larger when arbitrageurs are more risk averse (large α), and the payoff of asset A is more uncertain (large σ_A) and less correlated with the payoff of asset B (small $|\rho|$).

Eqs. (6) and $p_B = \bar{d}_B$ imply that the correlation, as of Period 0, between the prices of assets A and B in Period 1 is

$$\frac{\rho}{\sqrt{1 + \alpha^2 \sigma_A^2 (1 - \rho^2)^2 \sigma_u^2}}. \tag{11}$$

Eq. (11) shows that non-fundamental risk lowers the correlation between assets A and B. This is because it increases the volatility of asset A without affecting asset B.

Consider next an arbitrageur who takes a position in assets A and B in Period 0 in response to a demand shock in that period. Since non-fundamental risk increases asset A's volatility and lowers its correlation with asset B, it increases the volatility of the arbitrageur's return in Period 1. This volatility matters when the arbitrageur has a short horizon and must close his position in Period 1. Price volatility caused by demand shocks in Period 1 deters such an arbitrageur from absorbing demand shocks in Period 0. DSSW build on this idea to show that non-fundamental risk can be self-fulfilling. They assume a discrete-time infinite-horizon economy with an exogenous riskless rate r and an asset paying a constant dividend r in each period. Since one share of the second asset yields the same payoff as an investment of one dollar in the first asset, the law of one price implies that the price of the second asset should be one. DSSW show, however, that when arbitrageurs have a one-period horizon an equilibrium exists in which this price is stochastic. Intuitively, if arbitrageurs expect the price to be stochastic, demand shocks have an effect, and this renders the price stochastic.

The stochastic equilibrium of DSSW hinges on a number of assumptions. One is that arbitrageurs have short horizons: if they were infinitely lived, they would enforce the law of one price through buy-and-hold strategies. Short horizons can be viewed as a reduced form for financial constraints, as we show in Sections 3.4 and 3.5. A second critical assumption is that of an infinite horizon: with a finite horizon, the law of one price would hold.[10] Lowenstein & Willard (2006) show that even under these two assumptions, the law of one price would hold in DSSW if interest rates were endogenized, prices were precluded from becoming negative, or borrowing limits were imposed. But while DSSW's result on the failure of the law of one price may not be robust, their broader point that non-fundamental risk is an impediment to arbitrage remains important.

3.3 Short-Selling Costs

In our analysis so far, the only cost arbitrageurs face is risk. Additional costs, however, stem from the way arbitrageur positions are established and financed. Arbitrageurs often establish their positions in the repo market. For example, an arbitrageur wishing to establish a long position in an asset can borrow some of the needed cash by posting the asset as collateral—a repo transaction. Conversely, an arbitrageur wishing to establish a short position in an asset can borrow (to subsequently sell) the asset by posting cash as collateral—a reverse repo transaction. The interest rate earned on the cash, known as the repo rate, can differ across assets and this can be a source of arbitrage costs. For example, shorting an asset that carries a low repo rate relative to other assets is costly because the cash collateral posted by the short-seller earns a below-market interest rate.

Costs involved in establishing and financing positions are often referred to as holding costs. Tuckman & Vila (1992, 1993) introduce exogenous holding costs and show that they prevent arbitrageurs from eliminating mispricings. Arbitrageurs trade against mispricings only when these are large enough to compensate them for the holding costs they incur. Dow & Gorton (1994) show that holding costs can have disproportionately large effects when they are incurred by a sequence of short-horizon arbitrageurs.

Short-selling costs are holding costs associated with short positions. We introduce short-selling costs in our model by assuming that shorting asset A involves a cost c per share. Arbitrageurs maximize the objective

$$x_A(\bar{d}_A - p_A) - \frac{\alpha}{2}\left(x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_Ax_B\rho\sigma_A\sigma_B\right) - c|x_A|1_{\{x_A<0\}}, \tag{12}$$

[10] Our model confirms this: if assets A and B have the same payoff, the correlation ρ is one, and the non-fundamental risk in (10) and (11) disappears.

where $1_{\mathcal{S}}$ is equal to one if condition $\mathcal{S}$ is satisfied and zero otherwise. Eq. (12) is derived from (3) by subtracting the cost $c|x_A|$ of holding a short position x_A in asset A. Solving for the optimal demand for asset A and combining with the market-clearing condition (5), we find that the equilibrium price of asset A is given by

$$p_A = \bar{d}_A + \alpha\sigma_A^2(1-\rho^2)u \quad \text{if} \quad u \leq 0, \tag{13a}$$

$$p_A = \bar{d}_A + \alpha\sigma_A^2(1-\rho^2)u + c \quad \text{if} \quad u > 0. \tag{13b}$$

Short-selling costs affect the price only when the demand shock u is positive because this is when arbitrageurs hold a short position. The price increases by an amount equal to the short-selling cost c, so that arbitrageurs are compensated for incurring c.

Short-selling costs have an effect even when the assets have identical payoffs, and in that case they cause the law of one price to be violated. Setting $\rho = 1$ in (13b), we find $p_A = \bar{d} + c$, where $\bar{d}$ denotes the expected payoff of the two assets ($\bar{d} \equiv \bar{d}_A = \bar{d}_B$). Therefore, when the demand shock u is positive, the price of asset A is higher than the price of the identical-payoff asset B by an amount equal to the short-selling cost c. This analysis maps to the Palm-3Com example: c can be interpreted as the cost of shorting Palm and u as the demand by investors eager to hold Palm over 3Com. D'Avolio (2002) and Lamont & Thaler (2003) report that c was large and prevented arbitrageurs from exploiting the mispricing.

In our model, the short-selling cost c is an exogenous deadweight loss. A number of papers seek to endogenize c based on frictions in the repo market. In Duffie (1996), short-sellers have the choice between two assets with identical payoffs but different exogenous transaction costs. They prefer to short the low transaction cost asset, and their demand to borrow that asset in the repo market lowers the repo rate, driving up the short-selling cost. The friction in the repo market is that asset owners must incur an exogenous transaction cost to lend their asset. Unlike in our model, the short-selling cost c is not a deadweight loss, but accrues to asset owners. Therefore, it is an additional payoff earned from holding the asset and increases the asset price—even in the absence of any positive demand shock u. Krishnamurthy (2002) uses a similar model to show that a price premium arising from short-selling costs can coexist with a premium arising from an asset's superior liquidity.

Duffie et al. (2002) model the repo market as a search market in which asset borrowers and lenders negotiate bilaterally. The search friction generates a short-selling cost, which increases in the demand for short-selling. Vayanos & Weill (2008) show that the combination of a search spot market and a search repo market yields violations of the law of one price—even in the absence of any exogenous differences in transaction costs. Short-sellers concentrate on the more liquid asset, and their activity is what renders the asset more liquid. The more liquid asset carries a price premium both because of its superior liquidity and because high demand for short-selling drives up short-selling costs.

In the extreme case where short-selling costs are infinite, they amount to short-sale constraints, whose implications for asset prices are examined in a number of papers. Miller (1977) shows that when short-sales are not allowed, pessimistic investors are unable to trade and prices reflect the valuation of the most optimistic investors. Harrison & Kreps (1978) show that with multiple trading periods, prices even exceed the valuation of investors who are currently the most optimistic. Indeed, these investors have the option to resell the asset should other investors become more optimistic in future periods. Scheinkman & Xiong (2003) determine asset prices and the value of the resale option in a continuous-time model where differences in beliefs stem from overconfidence. Hong et al. (2006) show that overpricing and the value of the resale option are highest for assets with low float. Diamond & Verrecchia (1987) show that short-sale constraints do not cause overpricing when differences in beliefs stem from private signals rather than an agreement to disagree. Indeed, the market adjusts rationally for the fact that investors with negative private signals are unable to trade. Bai et al. (2006) show that short-sale constraints can cause underpricing because they generate uncertainty about the extent of negative private information. Hong & Stein (2003) show that the occasional release of negative private information can be the source of market crashes.

3.4 Leverage Constraints

In our analysis so far, there is no role for arbitrageur capital: while portfolio decisions and asset prices depend on arbitrageur risk aversion, they are independent of arbitrageur wealth. Wealth does not matter because of the simplifying assumption that arbitrageurs have exponential utility, i.e., their coefficient of absolute risk aversion is independent of wealth. Yet, the capital available to real-life arbitrageurs appears to be an important determinant of their ability to eliminate mispricings and provide liquidity to other investors.

The study of arbitrageur wealth effects is related to that of financial constraints. Indeed, an important theme in the literature on the limits of arbitrage is that arbitrageurs are sophisticated traders, better able to identify mispricings than other, less sophisticated investors. Since capital in the hands of arbitrageurs can earn higher return, other investors can gain by investing their capital with arbitrageurs. If, however, arbitrageurs could access external capital frictionlessly, they would be able to eliminate mispricings, and asset prices and allocations would be as in standard models. Thus, while arbitrageur capital can exceed the arbitrageurs' personal wealth, it appears to be limited by financial constraints.

In this section we study constraints on arbitrageurs' ability to acquire leverage by raising margin debt. We assume that assets A and B have identical payoffs. This isolates the effects of leverage constraints from those of risk: arbitrageurs bear no risk when exploiting a price discrepancy between assets A and B, and such a discrepancy can arise solely because of leverage constraints. To model leverage constraints, we focus on the mechanics of collateral in the repo market, following Geanakoplos (1997; 2003). We also assume that arbitrageurs must collateralize their positions in each asset separately. The model so derived is a simplified version of Gromb & Vayanos (2002; 2010).

Consider an arbitrageur wishing to establish a long position x_i in asset i. The arbitrageur can borrow some of the needed cash by posting asset i as collateral. The borrowed cash is typically less than the market value $x_i p_i$ of the asset collateral; otherwise, a drop in the asset price would cause the value of the collateral to drop below that of the loan. To determine the size of the loan, we assume for simplicity that margin loans have to be riskless and that competitive lenders set the rate equal to the riskless rate (which is zero). Riskless loans are not feasible when the payoff d of the two assets is normal, and we assume instead that d is distributed symmetrically over the bounded support $[\bar{d}-\epsilon, \bar{d}+\epsilon]$, where $\bar{d} \geq \epsilon$. An arbitrageur wishing to establish a long position x_i in asset i can thus borrow a maximum of $x_i(\bar{d}-\epsilon)$, and must pay the remainder

$$x_i \left(p_i - \bar{d} + \epsilon\right) \equiv x_i m_i^+. \tag{14}$$

out of his wealth. The parameter m_i^+ is the margin (or haircut) for a long position in asset i.[11]

Consider next an arbitrageur wishing to establish a short position x_i in asset i. The arbitrageur can borrow (and subsequently sell) asset i posting cash as collateral. The cash collateral typically exceeds the proceeds $|x_i|p_i$ from the sale of the borrowed asset; otherwise, an increase in the asset price would cause the value of the loan to rise above that of the collateral. As for long positions, we assume that margin loans have to be riskless and that competitive asset lenders pay the riskless rate on the cash collateral.[12] An arbitrageur wishing to establish a short position x_i in asset i must

[11] An alternative to requiring margin loans to be riskless is to allow for default and impose an upper bound on its probability. This would yield a constraint of the same form as (14), which would be interpreted as a value-at-risk constraint.

[12] This assumption eliminates the short-selling costs of Section 3.3.

post $|x_i|(\bar{d}+\epsilon)$ units of cash as collateral. Selling the asset yields $|x_i|p_i$ units, and the remainder

$$|x_i|\left(\bar{d}+\epsilon-p_i\right) \equiv |x_i|m_i^-. \tag{15}$$

must be drawn from the arbitrageur's wealth. The parameter m_i^- is the margin (or haircut) for a short position in asset i. Note that the margin requirements m_i^+ and m_i^- are increasing in the parameter ϵ, which is a measure of the volatility of asset i. Indeed, volatility increases the maximum loss that a long or short position can experience.

From (14) and (15), the positions x_A and x_B of an arbitrageur with wealth W_1 must satisfy the following leverage constraint:

$$W_1 \geq \sum_{i=A,B} |x_i|\left(m_i^+ 1_{\{x_i>0\}} + m_i^- 1_{\{x_i<0\}}\right). \tag{16}$$

Arbitrageurs maximize expected utility (1) subject to the budget constraint (2) and the leverage constraint (16).

When can arbitrageurs eliminate price discrepancies between assets A and B, thus enforcing the law of one price? The market-clearing condition (5) requires that arbitrageurs absorb the demand shock u, i.e., take a position $x_A = -u$. If, in addition, assets A and B trade at a price equal to their expected payoff $\bar{d}$, then it is optimal for arbitrageurs not to bear risk and hold an offsetting position $x_B = u$ in asset B. Since for $p_A = p_B = \bar{d}$ the margins m_i^+ and m_1^- are equal to ϵ, the leverage constraint (16) is satisfied if

$$W_1 \geq 2|u|\epsilon. \tag{17}$$

Arbitrageurs can enforce the law of one price if their wealth W_1 is large relative to the demand shock u and the margin requirement ϵ. In that case, the demand shock has no effect on the price of asset A and the market is perfectly liquid. Intuitively, arbitrageurs can provide perfect liquidity when their wealth is large because the leverage constraint is not binding. When, however, arbitrageur wealth is small, the leverage constraint is binding and liquidity is imperfect.

Leverage constraints can give rise to amplification, whereby the effects of an exogenous shock are amplified through changes in arbitrageur positions. Amplification can be derived in our model by assuming that arbitrageurs enter Period 1 with a position from Period 0, and that outside investors' demand is elastic. If, for example, arbitrageurs enter Period 1 with a long position, then a negative demand shock u lowers the price p_A, thus lowering arbitrageur wealth W_1 and tightening the leverage constraint (16). This can force arbitrageurs to liquidate positions, amplifying the price drop. Note that when liquidating their positions, arbitrageurs consume rather than provide liquidity. The liquidity providers are the outside investors, and their demand must be assumed elastic so that they are willing to buy from arbitrageurs.

Leverage constraints and amplification have been studied in macroeconomic settings, starting with Bernanke & Gertler (1989), Shleifer & Vishny (1992) and Kiyotaki & Moore (1997). In these papers, an adverse shock to economic activity depresses collateral values, and this can amplify the drop in activity. Geanakoplos (1997; 2003) defines the concept of a collateral equilibrium, where margin contracts are derived endogenously. He shows that margins are tied to asset volatility as in (14) and (15), and provides examples where margins increase following adverse shocks.

In a financial market context, Gromb & Vayanos (2002) study how leverage constraints affect the ability of arbitrageurs to eliminate mispricings and provide liquidity to outside investors. They show, within a dynamic setting, that liquidity increases in arbitrageur capital and that arbitrageurs can amplify exogenous shocks.[13] Brunnermeier & Pedersen (2009) and Gromb & Vayanos (2009a) extend this analysis to multiple assets in a static and dynamic setting, respectively. With multiple assets, leverage constraints generate not only amplification, but also contagion, whereby shocks to one asset are transmitted to otherwise unrelated assets through changes in arbitrageur positions. Pavlova & Rigobon (2008) derive a contagion result in an international-economy model with portfolio constraints, of which leverage constraints are a special case. Kondor (2009) shows that amplification can arise even in the absence of exogenous shocks, purely as a consequence of arbitrage activity. Indeed, if a price discrepancy between two assets were to remain constant or decrease over time, arbitrageurs would exploit it and reduce it to a level from which it could increase. Garleanu & Pedersen (2009) show that all else equal, assets with lower margin requirements can trade at higher prices. Related results are derived in Cuoco (1997), Basak & Croitoru (2000; 2006) and Geanakoplos (2003). Other dynamic models with leverage constraints include Aiyagari & Gertler (1999), Allen & Gale (2000), Anshuman & Viswanathan (2005), Geanakoplos & Fostel (2008), Adrian et al. (2009), Chabakauri (2009), Danielsson et al. (2009) and Rytchkov (2009).[14]

[13] Weill (2007) derives a link between arbitrageur capital and liquidity provision in a search model. See also Duffie & Strulovici (2009), who use a search model to study the gradual flow of arbitrage capital across trading opportunities.

[14] See also Yuan (2005) for a model in which leverage constraints hamper the revelation of private information, and Grossman & Vila (1992), Liu & Longstaff (2004), Jurek & Yang (2007) and Milbradt (2009) for partial-equilibrium models of portfolio choice by leverage-constrained investors. Amplification and contagion can also be derived in models without explicit leverage constraints but where arbitrageur risk aversion depends on wealth. This is done in Kyle & Xiong (2001) and Xiong (2001), who endow arbitrageurs with logarithmic utility, under which the coefficient of absolute risk aversion decreases in wealth. Following adverse shocks, arbitrageurs reduce their positions because they become more risk averse and not because they hit leverage constraints.

3.5 Constraints on Equity Capital

In addition to constraints on margin debt, arbitrageurs often face constraints in raising equity. For example, the equity of a mutual fund is determined by the flow of investors into the fund, and could be lower than the level at which the fund's profitable investment opportunities are fully exploited. Moreover, poor returns by the fund could trigger outflows by investors, and so render the fund more constrained. Shleifer & Vishny (1997) study the implications of constraints on equity capital for arbitrageurs' ability to exploit mispricings. They show that constraints give rise to amplification, via a mechanism akin to that for leverage constraints. Suppose that arbitrageurs hold long positions in an asset. A negative shock to the asset lowers its price and triggers outflows by investors in arbitrage funds. As a consequence, arbitrageurs are forced to reduce their positions, amplifying the price drop. These results can be derived in our model by assuming that (i) part of arbitrageur wealth belongs to other investors, and is withdrawn following poor returns in Period 1, (ii) arbitrageurs cannot borrow, (iii) arbitrageurs enter Period 1 with a position from Period 0, and (iv) outside investors' demand is elastic.

Shleifer & Vishny show additionally that constraints have an effect not only when they are binding, but also because of the possibility that they might bind in the future. Indeed, suppose that arbitrageurs with ample capital in Period 0 identify an underpriced asset. Investing heavily in that asset exposes them to the risk of large outflows by investors were the underpricing to worsen in Period 1. This would deprive arbitrageurs of capital when they need it the most since this is when the underpricing is the most extreme. As a consequence, arbitrageurs could refrain from investing heavily in the underpriced asset in Period 0, keeping instead some capital in cash. Arbitrageurs' strategy of seeking to match capital with profitable investment opportunities amounts to risk management. Risk management by arbitrageurs is also derived in models with leverage constraints, e.g., Gromb & Vayanos (2002), Liu & Longstaff (2004), Brunnermeier & Pedersen (2009) and Kondor (2009), and is further emphasized in Acharya et al. (2009) and Bolton et al. (2009). Holmstrom & Tirole (2001) explore the implications of risk management by firms for the cross-sectional pricing of assets. They show that assets paying off in states where firms' financial constraints bind are more expensive than assets paying off in other states because they provide capital when it is needed the most.

A number of papers integrate constraints on the equity capital available to arbitrage funds—and especially its dependence on past performance—into dynamic settings. In Vayanos (2004), fund managers face the constraint that their fund will be liquidated following poor performance, and this makes them unwilling to hold illiquid assets at times of high volatility. In He & Krishnamurthy (2008; 2009), the capital available to fund managers cannot exceed a fixed multiple of their personal wealth—a constraint derived from optimal contracting under moral hazard. When managers underperform, the constraint becomes binding and causes volatility and risk premia to increase. Guerrieri & Kondor (2009) derive amplification effects from managers' reputation concerns.[15] In Vayanos & Woolley (2008), investors withdraw capital following underperformance by fund managers because they infer rationally that managers might be inefficient. If, in addition, withdrawals have to be gradual, they generate short-run momentum and long-run reversal in asset returns.[16]

3.6 Summary and Next Steps

Sections 3.1-3.5 show how an array of costs faced by arbitrageurs limit their ability to eliminate mispricings and provide liquidity to outside investors. We next sketch what we view as two important next steps in this research agenda: (i) derive the financial constraints of arbitrageurs within an optimal contracting framework, and (ii) integrate limits-of-arbitrage ideas into richer models that incorporate multiple assets and dynamics and that can be used to address empirical puzzles.

As Sections 3.4 and 3.5 emphasize, arbitrageurs face financial constraints stemming from agency problems with their providers of capital. Most of the papers referenced in these sections do not fully endogenize the constraints. For example, papers on leverage constraints typically rule out equity and papers on equity constraints typically rule out debt. Deriving financial constraints within an optimal contracting framework is an important next step. Indeed, while the constraints studied in the literature take a variety of forms, they tend to generate results in common, e.g., amplification, contagion, and risk management by arbitrageurs. Endogenizing the constraints would help identify whether the common results are driven by a single underlying friction, or whether the constraints are fundamentally different. Identifying the frictions that underlie the constraints would also be useful for policy analysis, as it would clarify what a regulator can and cannot do to alleviate the effects of the constraints.

Starting with Kehoe and Levine (1993), a macroeconomic literature explores the asset pricing implications of financial constraints when these derive from limited commitment by borrowers.

[15]Reputation concerns are also explored in Dasgupta & Prat (2008), Dasgupta et al. (2008), and Malliaris & Yan (2009).

[16]Other papers studying general equilibrium implications of delegated portfolio management include Cuoco & Kaniel (2009), Petajisto (2009), Basak & Pavlova (2010) and Kaniel & Kondor (2010).

Limited commitment is also the source of the constraints in Kiyotaki and Moore (1997), while Holmstrom & Tirole (2001) derive the constraints from moral hazard. Constrained agents in these papers are real-economy firms having access to productive technologies not available to other investors. The interpretation of constrained agents as financial firms is made explicit in He & Krishnamurthy (2008; 2009), who derive constraints on fund managers uniquely able to invest in a subset of traded assets. The constraints derive from moral hazard and contracts are restricted to be static.[17] Extending this line of research to dynamic contracts, while retaining the tractability that is necessary to compute asset prices in general equilibrium, would be an important step forward. A recent literature on optimal dynamic financial contracting in continuous time, e.g., Biais et al. (2007) and DeMarzo & Fishman (2007), could be useful in this respect. The result of that literature that investors punish underperforming managers by scaling down their firms fits with the idea that investors reduce their stakes in underperforming funds.[18]

Our understanding of the costs faced by arbitrageurs has greatly benefited from papers exploring these costs in simple stylized settings. An important next step is to integrate limits-of-arbitrage ideas more squarely into asset pricing theory by developing tractable dynamic multi-asset models that can address empirical puzzles. Work along these lines could take constraints as given and so proceed in parallel with work on optimal contracting—although an important long-term objective is that the two lines of research eventually merge.

A number of papers explore dynamic multi-asset equilibrium settings under the assumption that the only cost faced by arbitrageurs is risk. In Greenwood (2005) and Hau (2009a), arbitrageurs absorb demand shocks of index investors following index redefinitions. In Gabaix et al. (2007), arbitrageurs are uniquely able to hold mortgage-backed securities, and while they can hedge interest-rate risk in the government bond market, they bear prepayment risk. In Garleanu et al. (2009), arbitrageurs absorb demand shocks in the options market, and while they can hedge delta risk in the stock market, they bear jump and volatility risk. In Vayanos & Vila (2009) and Greenwood & Vayanos (2010b), arbitrageurs absorb shocks to the demand and supply of government bonds with specific maturities, and hedge by trading bonds with other maturities. A common theme in these papers is that arbitrageurs transmit shocks to the demand for one asset to other assets, with the effects being largest for assets that covary the most with the original asset. This has implications for phenomena as diverse as index effects in the stock market, the pricing of prepayment risk in

[17] See also Acharya & Viswanathan (2009), who derive constraints from asset substitution, and Hombert and Thesmar (2009) and Stein (2009), who study arbitrageurs' choice between short- and long-term financing.

[18] Bolton & Scharfstein (1990) and Gromb (1994) derive this result in finite-horizon discrete-time settings, and Heinkel & Stoughton (1994) derive the result in a two-period fund-management setting. See also Biais et al. (2010), who show the converse result that investors reward overperforming managers by scaling up their firms.

the mortgage market, the behavior of implied volatility surfaces in the options market, and the behavior of risk premia in the government bond market. Hau (2009b), Jylha & Suominen (2009) and Plantin & Shin (2009) pursue similar themes for the foreign exchange market, and Naranjo (2009) does so for the futures market.

Other papers explore dynamic multi-asset equilibrium settings under the assumption that arbitrageurs face financial constraints. These papers are referenced in the last paragraph of Section 3.4 for the case of leverage constraints and Section 3.5 for the case of constraints on equity capital. They constitute a rapidly growing literature which has the potential to explain a variety of market anomalies based on limits of arbitrage and institutional frictions more broadly.

4 WELFARE AND POLICY

Financial crises, including the recent one, provide a painful reminder that government intervention can be important for the smooth functioning of financial markets. Arguments for intervention often center around the idea that failing financial institutions can cause disruptions in asset markets and the effects can propagate to other institutions. Standard models of asset pricing cannot be used to evaluate the merits of such arguments because they abstract away from financial frictions and institutions. In these models equilibrium is Pareto optimal and there is no scope for intervention.[19] Research on the limits of arbitrage has the potential to deliver a more useful framework for designing and assessing public policy. Indeed, this research emphasizes the role of specialized institutions in providing liquidity in asset markets. Understanding how the financial health of these institutions is affected by their trading decisions, and whether trading decisions are socially optimal, can guide public policy. Despite its relevance, the welfare analysis of asset markets with limited arbitrage is still in its infancy. In this section we survey the existing work and highlight what we view as the main issues, challenges and promises.

We start by clarifying some methodological issues. Several papers study whether constrained arbitrageurs stabilize or destabilize asset prices, i.e., decrease or increase volatility. This is only indirectly a welfare question; assessing welfare by means of a utility-based criterion, such as Pareto, is preferable. Given a welfare criterion, a natural question is whether constraints lower efficiency, i.e., is unconstrained arbitrage better than constrained arbitrage?[20] This question, however, is

[19]The same is true for models in which arbitrageur wealth effects derive from wealth-dependent risk aversion rather than explicit financial constraints (e.g., Kyle & Xiong (2001) and Xiong (2001)).

[20]A related question is whether the presence of arbitrageurs is beneficial, i.e., is constrained arbitrage better than no arbitrage at all? One would expect the two questions to generally have a positive answer: relaxing constraints should be desirable because arbitrageurs provide liquidity.

of limited relevance for assessing policy. Indeed, if arbitrageurs face constraints, one should not assume that a regulator could remove or even relax them. A more suitable criterion is constrained efficiency: can a regulator increase welfare relative to the equilibrium by choosing new allocations that are nevertheless subject to the constraints? At this juncture, the literature has taken two routes, which we discuss next.

The first route is to retain a traditional dynamic equilibrium asset pricing model, but not fully endogenize the constraints. This is done in Gromb & Vayanos (2002), who study how leverage constraints affect arbitrageurs' ability to provide liquidity. The main result in terms of welfare is that equilibrium can fail to be constrained efficient and a reduction in arbitrageur positions can make all agents better off. The intuition is as follows. Following an adverse shock, arbitrageurs incur capital losses and are forced to liquidate positions because their leverage constraints tighten. As a result, they find themselves more constrained and less able to provide liquidity—at a time where liquidity is low and its provision profitable. Ex-ante, arbitrageurs account for this possibility and engage in risk management by keeping some capital in cash to exploit episodes of low liquidity. However, they fail to account for the impact of their liquidations on other arbitrageurs during such episodes. Indeed, liquidation by one arbitrageur hurts other arbitrageurs because it lowers the price at which they can liquidate. This can hurt not only arbitrageurs but also outside investors because of the reduced liquidity that they receive.

The second route has been followed by papers in macroeconomics and international economics, e.g., Caballero & Krishnamurthy (2001), Lorenzoni (2008), Acharya et al. (2009), Hombert (2009), Korinek (2009). These papers derive financial constraints endogenously from contracting frictions in three-period settings. The constraints, however, concern real-economy firms having access to productive technologies, rather than financial firms such as arbitrageurs. Inefficiencies in these papers arise because of a similar mechanism as in the previous paragraph: firms do not account for the impact of their liquidations on the prices at which other firms can liquidate.[21]

The mechanism causing the inefficiencies is a pecuniary externality operating through price changes and the redistribution of wealth that these generate. A redistribution of wealth cannot be Pareto improving when markets are complete because marginal rates of substitution across time and states of nature are identical for all agents. However, when markets are incomplete, as is the case with financial constraints, marginal rates of substitution differ and Pareto improvements are possible. Geanakoplos & Polemarchakis (1986) show this mechanism in a general incomplete-

[21] See also Biais & Mariotti (2009) for a similar inefficiency in a labor economics context, and Nikolov (2009) for a calibration of the inefficiency.

markets setting.[22]

This constrained inefficiency of limited arbitrage opens the door for an analysis of policy (see Gromb & Vayanos (2002; 2009b) for a discussion). Suppose, for example, that arbitrageurs take excessive risk. A regulator might decrease their risk-taking incentives by tightening their financial constraints with a risk-based capital requirement. Alternatively, better risk-management could be enforced directly by taxing arbitrageurs in good times and possibly subsidizing them in bad times, in effect managing part of their resources for them. Subsidies in bad times can be implemented through lender of last resort policies or asset purchase programs.[23] Last, imperfect competition among arbitrageurs might lead them to internalize part of the price effects of their investment decisions, possibly leading them to adopt more socially desirable investment policies. This research agenda has the potential to inform debates on systemic risk, macro-prudential regulation, and lending of last resort, topics that are highly relevant in the context of financial crises.

[22]See also Stiglitz (1982), who shows that in incomplete financial markets, the competitive equilibrium may fail to be constrained efficient.

[23]For a discussion of such policies, see Krishnamurthy (2009) and the references therein. Krishnamurthy argues that such policies are desirable because they can move the market to a more efficient equilibrium.

Literature Cited

Acharya V, Shin H, Yorulmazer T. 2009. *A theory of slow-moving capital and contagion.* Work. Pap., New York Univ.

Acharya V, Viswanathan V. 2009. *Moral hazard, collateral and liquidity.* Work. Pap., New York Univ.

Adrian T, Etula E, Shin H. 2009. *Risk Appetite and Exchange Rates.* Work. Pap., Princeton Univ.

Aiyagari R, Gertler M. 1999. "Overreaction" of asset prices in general equilibrium. *Rev. Econ. Dynam.* 2:3-35.

Allen F, Gale D. 2000. Bubbles and crises. *Econ. J.* 110:236-55.

Amihud Y, Mendelson H. 1991. Liquidity, maturity, and the yield on U.S. treasury securities. *J. Finance* 46:479-86.

Anshuman R, Viswanathan V. 2005. *Costly collateral and liquidity.* Work. Pap., Duke Univ.

Anton M, Polk C. 2008. *Connected stocks.* Work. Pap., London Sch. Econ.

Bai Y, Chang E, Wang J. 2006. *Asset prices under short-sale constraints.* Work. Pap., Mass. Inst. Tech.

Barberis N, Shleifer A. 2003. Style investing. *J. Finan. Econ.* 68:161-99.

Barberis N, Shleifer A, Vishny R. 1998. A model of investor sentiment. *J. Finan. Econ.* 49:307-43.

Barberis N, Shleifer A, Wurgler J. 2005. Comovement. *J. Finan. Econ.* 75:283-317.

Barberis N, Thaler R. 2003. A survey of behavioral finance. In *Handbook of the Economics of Finance,* ed. G Constantinides, M Harris, R Stulz, 1:1051-121. London/New York: Elsevier.

Basak S, Croitoru B. 2000. Equilibrium mispricing in a capital market with portfolio constraints. *Rev. Finan. Stud.* 13:715-48.

Basak S, Croitoru B. 2006. On the role of arbitrageurs in rational markets. *J. Finan. Econ.* 81:143-73.

Basak S, Pavlova A. 2010. *Asset Prices and Institutional Investors.* Work. Pap., London Bus. Sch.

Bernanke B, Gertler M. 1989. Agency costs, net worth, and business fluctuations. *Am. Econ. Rev.* 79:14-31.

Bernard V, Thomas J. 1989. Post-earnings-announcement drift: delayed price response or risk premium? *J. Accounting Res.* 27(Suppl.):1-48.

Biais B, Mariotti T. 2009. Credit, wages, and bankruptcy laws. *J. Europ. Econ. Assoc.* 7:939-73.

Biais B, Mariotti T, Plantin G, Rochet J-C. 2007. Dynamic security design: convergence to continuous time and asset pricing implications. *Rev. Econ. Stud.* 74:345-90.

Biais B, Mariotti T, Rochet J-C, Villeneuve S. 2010. Large risks, limited liability, and dynamic moral hazard. *Econometrica.* 78:73-118.

Bolton P, Santos T, Scheinkman J. 2009. Outside and inside liquidity. Work. Pap., Columbia Univ.

Bolton P, Scharfstein D. 1990. A theory of predation based on agency problems in financial contracting. *Amer. Econ. Rev.* 80:93-106.

Boyer B. 2007. *Comovement among stocks with similar book-to-market ratios.* Work. Pap., Brigham Young Univ.

Brunnermeier M, Pedersen L. 2009. Market liquidity and funding liquidity. *Rev. Finan. Stud.* 22:2201-38.

Caballero R, Krishnamurthy A. 2001. International and domestic collateral constraints in a model of emerging market crises. *J. Monet. Econ.* 48:513-48.

Chabakauri G. 2009. *Asset pricing in general equilibrium with constraints.* Work. Pap., London Sch. Econ.

Chen H, Noronha G, Singal V. 2004. The price response to S&P 500 index additions and deletions: evidence of asymmetry and a new explanation. *J. Finance.* 59:1901-29.

Coval J, Stafford E. 2007. Asset fire sales (and purchases) in equity markets. *J. Finan. Econ.* 86:479-512.

Cuoco D. 1997. Optimal consumption and equilibrium prices with portfolio constraints and stochastic income. *J. Econ. Theory.* 72:33-73.

Cuoco D, Kaniel R. 2009. *Equilibrium prices in the presence of delegated portfolio management.* Work. Pap., Univ. Pennsylvania.

D'Avolio G. 2002. The market for borrowing stock. *J. Finan. Econ.* 66:271-306.

Dabora E, Froot K. 1999. How are stock prices affected by the location of trade? *J. Finan. Econ.* 53:189-216.

Daniel K, Hirshleifer D, Subrahmanyam A. 1998. A theory of overconfidence, self-attribution, and security market under and over-reactions. *J. Finance.* 53:1839-85.

Danielsson J, Shin H, Zigrand J-P. 2009. *Risk appetite and endogenous risk.* Work. Pap., London Sch. Econ.

Dasgupta A, Prat A. 2008. Information aggregation in financial markets with career concerns. *J. Econ. Theory.* 143:83-113

Dasgupta A, Prat A, Verardo M. 2008. *The price impact of institutional herding.* Work. Pap., London Sch. Econ.

DeBondt W, Thaler R. 1985. Does the stock market overreact? *J. Finance.* 40:793-805.

DeLong B, Shleifer A, Summers L, Waldmann R. 1990. Noise trader risk in financial markets. *J. Polit. Economy.* 98:703-38.

DeMarzo P, Fishman M. 2007. Optimal long-term financial contracting. *Rev. Finan. Stud.* 20:2079-128.

Diamond, D, Verrecchia R. 1987. Constraints on short-selling and asset price adjustment to private information. *J. Finan. Econ.* 18:277-311.

Dow J, Gorton G. 1994. Arbitrage chains. *J. Finance.* 49:819-849.

Duffie D. 1996. Special repo rates. *J. Finance.* 51:493-526.

Duffie D, Garleanu N, Pedersen L. 2002. Securities lending, shorting, and pricing. *J. Finan. Econ.* 66:307-39.

Duffie D, Garleanu N, Pedersen L. 2008. Valuation in over-the-counter markets. *Rev. Finan. Stud.* 20:1865-1900.

Duffie D, Strulovici B. 2009. *Capital mobility and asset pricing.* Work. Pap., Stanford Univ.

Fama E. 1991. Efficient capital markets: II. *J. Finance.* 46:1575-617.

Fama E, French K. 1992. The cross-section of expected stock returns. *J. Finance.* 47:427-65.

Fostel A, Geanakoplos J. 2008. Leverage cycles and the anxious economy. *Am. Econ. Rev.* 98:1211-44.

Gabaix X, Krishnamurthy A, Vigneron O. 2007. Limits of arbitrage: theory and evidence from the mortgage-backed securities market. *J. Finance.* 62:557-95.

Garleanu N, Pedersen L. 2009. *Margin-based asset pricing and deviations from the law of one price.* Work. Pap., UC Berkeley.

Garleanu N, Pedersen L, Poteshman A. 2009. Demand-based option pricing. *Rev. Finan. Stud.* In press

Geanakoplos J. 1997. Promises, promises. In *The Economy as an Evolving Complex System II,* ed. B Arthur, S Durlauf, D Lane, 285-320. Reading, MA: Addison-Wesley.

Geanakoplos J. 2003. Liquidity, default and crashes: endogenous contracts in general equilibrium. In *Advances in Economics and Econometrics: Theory and Applications II, Econometric Society Monographs: Eighth World Congress,* ed. M Dewatripont, L Hansen, S Turnovsky, 170-205. Cambridge, UK: Cambridge University Press.

Geanakoplos J, Polemarchakis H. 1986. Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete. In *Uncertainty, Information and Communication, Essays in Honor of Kenneth J. Arrow,* ed. W Heller, R Starr, D Starrett, 3:65-96. Cambridge, UK: Cambridge University Press.

Greenwood R. 2005. Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *J. Finan. Econ.* 75:607-649.

Greenwood R, Thesmar D. 2009. *Stock price fragility.* Work. Pap., Harvard Univ.

Greenwood R, Vayanos D. 2010a. *Bond supply and excess bond returns.* Work. Pap., Harvard Univ.

Greenwood R, Vayanos D. 2010b. Price pressure in the government bond market. *Am. Econ. Rev. P.&P.* In press.

Gromb D. 1994. *Renegotiation in debt contracts.* Work. Pap., INSEAD.

Gromb D, Vayanos D. 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *J. Finan. Econ.* 66:361-407.

Gromb D, Vayanos D. 2009a. *Financially constrained arbitrage and cross-market contagion.* Work. Pap., INSEAD.

Gromb D, Vayanos D. 2009b. *Leverage and liquidity dry-ups: a framework and policy implications.* Work. Pap., INSEAD.

Gromb D, Vayanos D. 2010. A model of financial market liquidity based on arbitrageur capital. *J. Europ. Econ. Assoc. P.&P.* In press.

Grossman S, Miller M. 1988. Liquidity and market structure. *J. Finance.* 43:617-37.

Grossman S, Vila J-L. 1992. Optimal dynamic trading with leverage constraints. *J. Finan. Quant. Anal.* 27:151-168.

Guerreri V, Kondor P. 2009. *Fund managers and defaultable debt.* Work. Pap., Univ. Chicago.

Harris L, Gurel E. 1986. Price and volume effects associated with changes in the S&P 500 list: new evidence for the existence of price pressures. *J. Finance.* 41:815-29.

Harrison M, Kreps D. 1978. Speculative investor behavior in a stock market with heterogeneous expectations. *Quart. J. Econ.* 92:323-36.

Hau H. 2009a. *Global versus local asset pricing: evidence from arbitrage of the MSCI index change.* Work. Pap., INSEAD.

Hau H. 2009b. *The exchange rate effect of multi-currency risk arbitrage.* Work. Pap., INSEAD.

He Z, Krishnamurthy A. 2008. *Intermediary asset pricing.* Work. Pap., Univ. Chicago.

He Z, Krishnamurthy A. 2009. *A model of capital and crises.* Work. Pap., Univ. Chicago.

He Z, Xiong W. 2008. *Multi-market delegated asset management.* Work. Pap., Univ. Chicago.

Heinkel R, Stoughton N. 1994. The dynamics of portfolio management contracts. *Rev. Finan. Stud.* 7:351-87.

Holmstrom B, Tirole J. 2001. LAPM: A liquidity-based asset pricing model. *J. Finance.* 56:1837-67.

Hombert J. 2009. *Solvency shocks, liquidity shocks, and fire sales.* Work. Pap., HEC.

Hombert J, Thesmar D. 2009. *Limits of limits of arbitrage: theory and evidence.* Work. Pap., HEC.

Hong H, Scheinkman J, Xiong W. 2006. Asset float and speculative bubbles. *J. Finance.* 61:1073-1117.

Hong H, Stein J. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *J. Finance.* 54:2143-84.

Hong H, Stein J. 2003. Differences of opinion, short-sales constraints, and market crashes. *Rev. Finan. Stud.* 16:487-525.

Jegadeesh N, Titman S. 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *J. Finance.* 48:65-91.

Jotikasthira P, Lundblad C, Ramadorai T. 2009. *Asset fire sales and purchases and the international transmission of financial shocks.* Work. Pap., Univ. of Oxford.

Jurek J, Yang H. 2007. *Dynamic portfolio selection in arbitrage.* Work. Pap., Princeton Univ.

Jylha P, Suominen M. 2009. Speculative capital and currency carry trades. *J. Finan. Econ.* In press

Kaniel R, Kondor P. 2010. *A delegated Lucas tree.* Work. Pap., Duke Univ.

Kehoe T, Levine D. 1993. Debt-constrained asset markets. *Rev. Econ. Stud.* 60:865-88.

Kiyotaki N, Moore J. 1997. Credit cycles. *J. Polit. Economy.* 105:211-48.

Kondor P. 2009. Risk in dynamic arbitrage: price effects of convergence trading. *J. Finance.* 64:638-58.

Korinek A. 2009. *Systemic risk taking: amplification effects, externalities, and regulatory responses.* Work. Pap., Univ. Maryland.

Krishnamurthy A. 2002. The bond/old-bond spread. *J. Finan. Econ.* 66:463-506.

Krishnamurthy A. 2009. Amplification mechanisms in liquidity crises. *Amer. Econ. J.: Macroecon.* forthcoming.

Kyle A, Xiong W. 2001. Contagion as a wealth effect of financial intermediaries. *J. Finance.* 56:1401-40.

Lamont O, Thaler R. 2003. Can the market add and subtract? Mispricing in tech stock carve-outs. *J. Polit. Economy* 111:227-268.

Lagos R, Rocheteau G, Weill P-O. 2009. *Crises and liquidity in over the counter markets.* Work. Pap., New York Univ.

LeRoy S, Porter R. 1981. The present value relation: tests based on implied variance bounds. *Econometrica.* 49:555-74.

Liu J, Longstaff F. 2004. Losing money on arbitrage: optimal dynamic portfolio choice in markets with arbitrage opportunities. *Rev. Finan. Stud.* 17:611-41.

Lorenzoni G. 2008. Inefficient credit booms. *Rev. Econ. Stud.* 75:809-33.

Loewenstein M, Willard G. 2006. The limits of investor behavior. *J. Finance.* 61:231-58.

Lou D. 2009. *A flow-based explanation for return predictability.* Work. Pap., London Sch. Econ.

Malliaris S, Yan H. 2009. *Nickels versus black swans: reputation, trading strategies and asset prices.* Work. Pap., Yale Univ.

Milbradt K. 2009. *Trading and valuing toxic assets.* Work. Pap., Mass. Inst. Tech.

Miller E. 1977. Risk, uncertainty and divergence of opinion. *J. Finance.* 32:1151-68.

Mitchell M, Pulvino T, Stafford E. 2002. Limited arbitrage in equity markets. *J. Finance.* 57:551-84.

Naranjo L. 2009. *Implied interest rates in a market with frictions.* Work. Pap., ESSEC.

Nikolov K. 2009. *Is private leverage excessive?* Work. Pap., London Sch. Econ.

Pavlova A, Rigobon R. 2008. The role of portfolio constraints in the international propagation of shocks. *Rev. Econ. Stud.* 75:1215-1256.

Petajisto A. 2009. Why do demand curves for stocks slope down? *J. Finan. Quant. Anal.* 44:1013-44.

Plantin G, Shin H. 2009. *Carry trades and speculative dynamics.* Work. Pap., Toulouse Sch. Econ.

Rosenthal L, Young C. 1990. The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC. *J. Finan. Econ.* 26:123-41.

Rytchkov O. 2009. *Dynamic margin constraints.* Work. Pap., Temple Univ.

Scheinkman J, Xiong W. 2003. Overconfidence and speculative bubbles. *J. Polit. Economy.* 111:1183-1219.

Schwert W. 2003. Anomalies and market efficiency. In *Handbook of the Economics of Finance,* ed. G Constantinides, M Harris, R Stulz, 1:939-74. London/New York: Elsevier.

Shiller R. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *Amer. Econ. Rev.* 71:421-36.

Shleifer A. 1986. Do demand curves for stocks slope down? *J. Finance.* 41:579-90.

Shleifer A, Vishny R. 1992. Liquidation values and debt capacity: a market equilibrium approach. *J. Finance.* 47:1343-66.

Shleifer A, Vishny R. 1997. The limits of arbitrage. *J. Finance.* 52:35-55.

Stein J. 2009. Presidential address: sophisticated investors and market efficiency. *J. Finance.* 64:1517-48.

Stiglitz J. 1982. The inefficiency of the stock market equilibrium. *Rev. Econ. Stud.* 49:241-61.

Tuckman B, Vila J-L. 1992. Arbitrage with holding costs: a utility-based approach. *J. Finance.* 47:1283-1302.

Tuckman B, Vila J-L. 1993. *Holding costs and equilibrium arbitrage.* Work. Pap., Mass. Inst. Tech.

Vayanos D. 2004. *Flight to quality, flight to liquidity and the pricing of risk.* Work. Pap., London Sch. Econ.

Vayanos D, Vila J-L. 2009. *A preferred-habitat theory of the term structure of interest rates.* Work. Pap., London Sch. Econ.

Vayanos D, Weill P-O. 2008. A search-based theory of the on-the-run phenomenon. *J. Finance.* 63:1361-98.

Vayanos D, Woolley P. 2008. *An institutional theory of momentum and reversal.* Work. Pap., London Sch. Econ.

Vijh A. 1994. S&P 500 trading strategies and stock betas. *Rev. Finan. Stud.* 7:215-51.

Warga A. 1992. Bond returns, liquidity, and missing data. *J. Finan. Quant. Anal.* 27:605-17.

Weill P-O. 2007. Leaning against the wind. *Rev. Econ. Stud.* 74:1329-54.

Wurgler J, Zhuravskaya E. 2002. Does arbitrage flatten demand curves for stocks? *J. Bus.* 75:583-608.

Xiong W. 2001. Convergence trading with wealth effects. *J. Finan. Econ.* 62:247-292.

Yuan K. 2005. Asymmetric price movements and borrowing constraints: a REE model of crisis, contagion, and confusion. *J. Finance.* 60:379-411.