Lecture Notes

Information Choice and Its Foundations

Thorsten V. Koeppl

Queen's University

Fall Term 2014

1 Basic Set-up

1.1 Economy

We start of with a standard Dixit-Stiglitz type model of monopolistic competition a la Blanchard and Kiyotaki (1989).

Demand Block

- Aggregate Demand:
$$1 = \beta E_t \left[(1 + R_{t+1}) \frac{P_t C_t}{P_{t+1} C_{t+1}} \right] \tag{1.1}$$

- Demand for Individual Goods:
$$C_{it} = \left(\frac{P_{it}}{P_t} \right)^{-\gamma} \tag{1.2}$$

- Price Index
$$P_t = \left(\int_0^1 P_{it}^{1-\gamma} di \right)^{1/(1-\gamma)} \tag{1.3}$$
where $\gamma > 1$ is the elasticity of substitution between individual goods

- Labour-Leisure Choice
$$C_t L_{it}^{1/\psi} = \frac{W_{it}}{P_t} \tag{1.4}$$
where ψ is the Frisch elasticity of labour supply

Supply Block

- Production Function
$$Y_{it} = A_{it} L_{it} \tag{1.5}$$
where $A_{it} = \exp z_{it}$ are productivity (supply) shocks

- Expected (Real) Profits
$$\Pi_{it} = E_{it} \left[(1 + \tau) \frac{P_{it}}{P_t} Y_{it} - \frac{W_{it}}{P_t} L_{it} \right] \tag{1.6}$$
where τ is a sales tax that is rebated (lump-sum) to households

We will place the analysis in the context of monetary policy. To do so, we simply assume an (exogenous) process for M_t so that some quantity theory equation holds

$$M_t = Y_t P_t \tag{1.7}$$

Two issues:

1) What is the role of the expectation E_{it}?

2) How are expectations E_{it} formed?

We do not abandon rationality nor do we fuzz around with preferences. Indeed, both of these would change the expectations operator in the Euler equation rather than the one on the firm's problem.

1.2 Full information

Firms do not have an intertemporal problem. Hence, they can maximize profits state-by-state by setting prices taking demand $Y_{it} = f(P_{it})$ as given

$$(1+\tau)\left(Y_{it} + P_{it}\partial Y_{it}/\partial P_{it}\right) - \partial Y_{it}/\partial P_{it} W_{it}/A_{it} = 0 \tag{1.8}$$

or

$$P_{it} = \frac{1}{1+\tau}\left(\frac{1}{\epsilon_p(Y_{it})+1}\right)\frac{W_{it}}{A_{it}} \tag{1.9}$$

where $\epsilon_p(Y_{it})$ is the price elasticity of demand which is $-1/\gamma$.

Prices are set as a mark-up over (nominal) marginal costs.[1] Log-linearizing, using market clearing, the first-order condition on labour and the demand equation we obtain

$$p_{it} = p_t + \mu + \alpha(y_t - a_{it}) \tag{1.10}$$

The parameter $\alpha \in (0,1)$ determines how elastic the firm's prices are with respect to aggregate output relative to its own productivity.

[1] Deviations from the aggregate price is set as a mark-up over real marginal costs.

Result: Assuming that $p_t = \int p_{it} di$, we get that aggregate output (supply) is independent of m_t. In other words, monetary policy (demand) shocks do not matter, since we have a vertical aggregate supply curve.

Why? Integrate over p_{it} implies that p_t drops out.

1.3 Incomplete Information

Assume now that firms *cannot observe their demand* Y_{it}. Profits are

$$E_{it} = \left[(1+\tau) Y_t \left(\frac{P_{it}}{P_t} \right)^{1-\gamma} - \left(\frac{W_{it}}{A_{it} P_t} \right) Y_t \left(\frac{P_{it}}{P_t} \right)^{-\gamma} \right] \tag{1.11}$$

Problem: FOC is non-linear in P_t. Hence, we cannot get an exact solution to the FOC for P_{it}.

Hence, work with "Certainty-Equivalence". What is it? (i) quadratic approximation to the objective function yields *linear* decision rules; (ii) log-linearize the first-order condition:[2]

$$p_{it} = E_{it}[p_t + \alpha(y_t - a_{it})] \tag{1.12}$$

Now the information set of the firm starts to matter and we need to have a theory of how firms form their expectations. This can give rise to "frictions" that cause monetary policy (or demand shocks) to matter.[3]

Three possibilities:

1. Only observe own price: standard Lucas-Phelps economy with Bayesian updating

2. Sticky information: only a fraction λ of firms obtains new information on p

[2] We set $\mu = 0$.

[3] We can view this as a way to justify "sticky" prices beyond ad-hoc time-dependent pricing a la Calvo. Traditionally, we have relied on state-dependent price adjustments with menu costs. However, prices seem to be rather flexible, since incentives to adjust are large precisely when the individual price is out-of-whack a lot. With informational friction, it is fully rational for firms not to adjust their prices.

3. Noisy information: we obtain a game of strategic complementarities as before

What are the micro foundations for these frictions?

2 Delayed Information

Assume now monetary policy follows

$$m_t = m_{t-1} + v_t \tag{2.1}$$

where $v_t \sim \mathcal{N}(0, \sigma^2)$ is interpreted as a *shock to velocity.*

Assume further that only a fraction λ observes this shock today, while all others observe the shock only tomorrow, but have information on all variables from the previous period.

We got

$$p_t = \int p_{it} di = \int E_{it}[p_t + \alpha y_t] di - \alpha \int a_{it} di = \int E_{it}[\alpha m_t + (1-\alpha) p_t] di \tag{2.2}$$

where we have normalized $A = 1$ and assume that there are no technology shocks.

When receiving information (fraction λ), we have that m_t is measurable w.r.t. E_{it}. When not (fraction $1-\lambda$), we have $E_{it} = E_{t-1}$.

Represented as forecast errors on the price level this gives us

$$(1-(1-\alpha)\lambda)(p_t - E_{t-1}[p_t]) = \alpha\lambda(m_t - E_{t-1}[m_t]) + \alpha E_{t-1}[m_t - p_t]. \tag{2.3}$$

We have $E_{t-1}[m_t - p_t] = 0$. Why? Take expectations w.r.t. E_{t-1}.

This results in an expectations augmented Phillips curve that is – partially – backward looking.

$$p_t = \xi v_t + m_{t-1} \tag{2.4}$$

$$y_t = (1-\xi) v_t \tag{2.5}$$

with $\xi \in (0,1)$, where $\partial\xi/\partial\alpha > 0$.

Also: $\alpha = \frac{\psi+1}{\psi+\gamma}$ so that the elasticities of substitution and labour supply matter.

<u>Remark:</u> One can introduce persistence by assuming that the arrival of new information is a stochastic process for individual firms. With probability λ a firm gets then new information so that its expectations operator is E_t. Otherwise, it is E_{t-j} where it has gotten information j periods ago.

A firm's information is then geometrically distribution so that we have that a mass $\lambda(1-\lambda)^j$ of firms has the expectations operator E_{t-j}. Then, the price level is a fixed point to the equation

$$p_t = \sum_{j=0}^{\infty} \lambda(1-\lambda)^j E_{t-j}[\alpha m_t + (1-\alpha)p_t]. \tag{2.6}$$

3 Noisy Information

Assume that firms receive a private signal $z_{it} = m_t + \epsilon_{it}$, where $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2/\tau)$. The parameter $\tau \in [0,\infty)$ measures the informativeness of the signal. Again, we assume that with a one-period delay, everyone learns the signal perfectly.

Now, higher-order beliefs matter again. Why? It matters for demand what a firm believes of other islands and what those islands believe about the firm's belief, etc.

This implies that p_t has to solve

$$p_t = \sum_{j=1}^{\infty} \alpha(1-\alpha)^{j-1} \bar{E}_t^j(m_t) \tag{3.1}$$

where $\bar{E}_t^j = \int E_{it}[\int E_{it}[\dots \quad j \text{ times} \quad \dots]di]di$.

Since both variables are normally distributed with relative (!) precision 1 and τ respectively we again have that the belief is a weighted average of the prior and the signal itself

$$E_{it} = E_t[n_t | z_{it} = n_t + \epsilon_{it}] = E_{t-1}(m_t) + \left(\frac{\tau}{1+\tau}\right)(z_{it} - E_{t-1}[m_t]) \tag{3.2}$$

Again, prices and output are given by

$$p_t = \xi^0 v_t + m_{t-1} \tag{3.3}$$

$$y_t = (1-\xi^0) v_t \tag{3.4}$$

with $\xi^0 = \frac{\alpha\tau}{1+\alpha\tau} \in (0,1)$.

Note that α matters here. The lower α, the stronger the strategic complementarities.

<u>Remark:</u> To introduce persistence, one can assume that the signals accumulate over time without full revelation. The problem is then very hard to solve. The way to go is to guess and verify a log-linear price rule and then to rely on Kalman filtering as the optimal learning process/signal extraction over time.

4 Microfoundations I: Information Costs

4.1 Set-up

- state s_t follows a known first-order Markov process
- $u^\tau = (u_t, u_{t+\tau})$ innovations to this process
- $s_{t+\tau} = \Psi(s_t, u^\tau)$ is the transition process

Firms make decisions when to update and obtain a new observation of s.

- $D(i) = \mathbb{N}_0 \to \mathbb{R}$ are planning dates
- $d(i) = D(i) - D(i-1)$ are periods of inattentiveness
- $F_{D(i)}$ for $t \in [D(i), D(i+1))$ is the information set at t

Key: firms decisions must be measurable w.r.t. $F_{D(i)}$

4.2 Updating Problem

Value function for the firm

$$V(s) = \sup_d \int_0^d e^{-rt}\Pi(s,t)dt + e^{-rd}E[-K(s_d) + V(s_d)|s] \tag{4.1}$$

where $s_d = \Psi(s, u^d)$.

The problem is thus *recursively* time-dependent. The optimal adjustment interval depends on the state s, when the firm has obtain information for the last time.[4]

The solution $(V(s), d(s))$ is described by the FOC and the envelope conditions

$$0 = e^{-rd}\Pi(s,d) - re^{-rd}E[-K(s_d) + V(s_d)|s] - e^{-rd}E[K(s_d) + V(s_d)|s]\frac{\partial \Psi}{\partial d} \tag{4.2}$$

$$V_s(s) = \int_0^d e^{-rt}\Pi_s(s,t)dt + e^{-rd}E[-K_s(s_d) + V_s(s_d)|s]\Psi_s(s, u^d) \tag{4.3}$$

Intuition: Flow value of planning $(\Pi(s,d))$ is keeping the old plan chosen at s. Benefit from updating at d is given by the flow value of having gotten new information minus the costs of doing so plus the costs of postponing for doing so (saving marginal changes in costs K, but given up marginal changes in V).

4.3 Example – Isoelastic Demand

Demand is given by $D(\epsilon_t, P_{it}) = \epsilon_t P_{it}^{-\theta}$ where ϵ_t is an iid shock.

Marginal costs s_t follows an (independent) geometric Brownian motion so that $ds_t = \sigma s_t W_t$ where W_t is a Brownian motion.

Costs of updating are a fraction κ of profits.

Note that the firm's problem is *independent* of choosing d and whether to update information or not. Hence,

$$\max_{P_{it}} E[DP_{it} - sP_{it}] \tag{4.4}$$

[4] It is not an optimal stopping time problem which is dependent on the *current, observed state*.

with FOC given by

$$P_{it} = \frac{\theta}{\theta - 1} E[s_t] \tag{4.5}$$

Hence, expected profits are constant during periods of inattentiveness and given by

$$\Pi(s, t) = \Xi E[s_t]^{1-\theta} = \Xi s^{1-\theta} \tag{4.6}$$

with expected costs of updating being equal to

$$E[\kappa \Xi s_d^{1-\theta}] \tag{4.7}$$

Solution:

1) Guess that the value function must satisfy $V = As^{1-\theta}$.

2) Use the fact that $E_t[s_d^{1-\theta}] = s^{1-\theta} e^{bd}$ where $b = 0.5\sigma^2\theta(\theta - 1)$ to obtain

$$A = \max_d \left[\frac{\Xi(1 - e^{-rd})}{r} + e^{(b-r)d}(-\kappa\Xi + A) \right] \tag{4.8}$$

3) Solve to obtain (for small κ) approximately

$$d^* = \sqrt{\frac{4\kappa}{\sigma^2\theta(\theta - 1)}} \tag{4.9}$$

Hence: information is updated more slowly when costs increase, volatility of MC falls and demand is less price elastic.

4.4 Example – Phillips Curve

Remark: One can show under some assumptions that the stationary equilibrium distribution of inattentiveness – the length d – among producers is exponentially distributed with some parameter ρ. Hence, at any point in time, a fraction $\rho e^{-\rho x}$ have not planned for x periods and at every instant the share of firms planning is ρ.

The model consists of

- stochastic costs of adjustments κ_i
- exogenous process for monetary policy $m_t = p_t + y_t$
- optimal price setting $p_{it} = E_D[p_t + \alpha(y_t - y_t^n)]$ where y_t^n is the costless information steady state

Denote the distribution of time among firms since the last adjustment of prices by $D(i)$. The price level p_t is a fixed point of

$$p_t = \int p_{it} dD(i) = \rho \int_{-\infty}^{t} e^{\rho(t-i)} E_i[p_t + \alpha(y_t - y_t^n)]di \tag{4.10}$$

Now differentiate with respect to time to obtain

$$\dot{p}_t = \alpha\rho(y_t - y_t^n) + \rho \int_{-\infty}^{t} e^{-\rho(t-i)} E_i[\dot{p}_t + \alpha(\dot{y}_t - \dot{y}_t^n)]di \tag{4.11}$$

which corresponds to the sticky information Phillips Curve we have derived earlier.

5 Microfoundation II: Constraints on Information Processing

5.1 Preliminaries

Individuals need to process and use information. They choose how much information they use, not its content. Before we had that once you get information, you can perfectly use it. Now you cannot. Hence, one's behaviour can react continuously, but imprecisely to the available information.

We borrow a lot from *information theory.*[5] An input X is transformed into an output Y through a "communication channel" which reduces the uncertainty about X by providing additional information $p(Y|X)$ (see diagram).

[5] A great reference for this topic is the book by Clover and Thomas (2006).

We view rational inattention as the "capacity" of the channel to be limited, so that $p(Y|X)$ cannot perfectly relate the information.

5.2 Basic Concepts

Shannon's entropy measures the uncertainty of a random variable

$$H(X) = -E\left[\log(p(x))\right] \tag{5.1}$$

For example, use $\log_2$ and interpret the entropy as the number of bits required to describe the distribution of the random variable.

Let $X = \{0, 1\}$ with equal probability. Then, we have that

$$H(X) = -\int \frac{1}{2}\log_2\left(\frac{1}{2}\right)dx = -2\log_2\left(\frac{1}{2}\right) = 1; \tag{5.2}$$

that is we need 1 bit to describe this random variable.

The mutual information describes how much Y tells us about X

$$I(X,Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y). \tag{5.3}$$

The capacity of an information channel is given by

$$K = \max_{p(x)} I(X,Y) \tag{5.4}$$

where $p(x)$ is describing the input distribution for the channel that is described by $p(Y|X)$.

5.3 Deriving the Rational Inattention Constraint

Let X be a multivariate normal variable with $\mathcal{N}(\mu, \Sigma)$ and Y be a signal about X which errors distributed according to $\mathcal{N}(0, \hat{\Sigma})$.

We have in nats

$$
\begin{aligned}
I(X,Y) &= H(X) - H(X|Y) \\
&= \frac{1}{2}\ln\left((2\pi e)^n|\Sigma|\right) - \frac{1}{2}\ln\left((2\pi e)^n|\hat{\Sigma}|\right) \\
&= \frac{1}{2}\ln\left(\frac{|\Sigma|}{|\hat{\Sigma}|}\right)
\end{aligned} \tag{5.5}
$$

With X and Y normally distributed, they maximize mutual information and, hence, define capacity K of any channel so that

$$K = \frac{1}{2}\ln\left(\frac{|\Sigma|}{|\hat{\Sigma}|}\right) \tag{5.6}$$

This yields then the constraint for the posterior covariance matrix

$$|\hat{\Sigma}| = e^{-2K}|\Sigma|. \tag{5.7}$$

The idea is here to choose the covariance matrix of the posterior belief – or, equivalently, the covariance matrix of the signals – so that to satisfy this constraint.

Assume, for example that all the signals are specific to each risk, i.e. the errors are uncorrelated across risks. Then, we have that

$$\Pi_{i=1}^{N}\hat{\sigma}_{ii}^{-2} \leq \frac{e^{2K}}{|\Sigma|} \tag{5.8}$$

which is a constraint on the *joint* precision of the signals.

5.4 Example – Mackowiak and Wiederholt

Consider a firm that has an information set s_{it} and sets its price equal to

$$p_{it} = E[p_t - az_{it}|s_{it}] \tag{5.9}$$

It obtains signals

$$s_p = p_t + \epsilon_t \text{ where } \epsilon_t \sim \mathcal{N}(0,\sigma_\epsilon^2) \tag{5.10}$$

$$s_z = z_{it} + \psi_t \text{ where } \psi_t \sim \mathcal{N}(0,\sigma_\psi^2) \tag{5.11}$$

The firm wants to minimize its expected losses relative to its full information price p_t^*, but is constrained through its information capacity K.

$$\min_{\hat{\sigma}_p,\hat{\sigma}_z} E[(p_{it} - p_t^*)^2] \tag{5.12}$$

subject to

$$\left(\frac{\hat{\sigma}_p\hat{\sigma}_z}{\sigma_p\sigma_z}\right)^{-2} \leq e^{2K} \tag{5.13}$$

$$\hat{\sigma}_p^2 \leq \sigma_p^2 \tag{5.14}$$

$$\hat{\sigma}_z^2 \leq \sigma_z^2 \tag{5.15}$$

Recall that Bayesian updating just adds the precisions of the two signals so that

$$\hat{\sigma}_p^{-2} = \sigma_p^{-2} + \sigma_\epsilon^{-2} \text{ and } \hat{\sigma}_z^{-2} = \sigma_z^{-2} + \sigma_\psi^{-2} \tag{5.16}$$

Rewriting the problem and using the constraint to get rid of $\hat{\sigma}_z$, we obtain for the objective function

$$\min_{\hat{\sigma}_p} \frac{1}{2}\left(\hat{\sigma}_p^2 + a^2\frac{\sigma_p^2\sigma_z^2}{e^{2K}\hat{\sigma}_p^2}\right) \tag{5.17}$$

Hence, for an interior solution we have

$$\left(\frac{\hat{\sigma}_p}{\sigma_p}\right)^{-2} = \frac{1}{a}e^K\frac{\sigma_p}{\sigma_z}. \tag{5.18}$$

More attention is paid to the aggregate shock p_t, the larger the capacity K, the less precise the prior information on p_t relative to z_{it} and the smaller the weight on the idiosyncratic shock in the objective function.

5.5 Remarks

- General equilibrium applications are a problem. Prices for example cannot be market clearing as decision makers are not reacting precisely and immediately. Also, individual decision makers cannot optimize precisely. For example, one chooses consumption, but savings need to be a residual over which one cannot necessarily optimize.

- Institutions can be designed to precisely deal with information processing and, hence, are an artefact of rational inattention.
- What about aggregation of correlated information processing?
- Small shocks might lead to inertia or small reaction. Large shocks will change this. Hence, structural parameters are unlikely to be invariant to the size of shocks. This makes models calibrated or estimated in normal times irrelevant (quantitatively) for periods of big shocks.

6 Literature

Veldkamp, Chapter 6 (2011)

Cover and Thomas (2006)

Mackowiak and Wiederholt, AER (2009)

Mankiw and Reis, HB of Monetary Economics (2011)

Reis, REStud (2006)

Sims, HB of Monetary Economics (2011)