Endogenous Information Choice

Topics in Macro Lecture 7

November 23, 2010

An optimizing trader will process those prices of most importance to his decision problem most frequently and carefully, those of less importance less so, and most prices not at all. Of the many sources of risk of importance to him, the business cycle and aggregate behavior generally is, for most agents, of no special importance, and there is no reason for traders to specialize their own information systems for diagnosing general movements correctly.

Robert E. Lucas (1977)

Endogenous Information Choice

Today we will talk mostly about "rational inattention"

- Agents can choose what to observe, i.e. choose D and Σ_{vv} in state space system

$$
\begin{aligned}
X_t &= AX_{t-1} + Cu_t \\
Z_t &= DX_t + v_t
\end{aligned}
$$

- A and C may in equilibrium depend on D and Σ_{vv}
- For this to be an interesting question, there has to be costs/constraints associated with acquiring more precise information (otherwise $D = I$ and $\Sigma_{vv} = \mathbf{0}$)

Rational Inattention

What is it?

- All information is in principle available, but an agent cannot pay attention to all available information
- Agents have information processing constraints and therefore chooses to observe the most important information

Some history of thought:

- Tools borrowed from information theory (a branch of applied mathematics)
- Sims (Carnegie-Rochester 1998) introduced it to economists
- Price setting model of Mackowiak and Wiederholt (AER 2009) probably the highest impact paper using the method

Entropy

- Entropy is a term from thermodynamics, roughly meaning the (inverse of) degree of order.
- Shannon (1948) used the term to describe information flows independently of the medium information is transmitted through
- In this context, entropy means how much information is required on average to describe the outcome of a random variable
- Entropy is measured in bits (per time unit)

Entropy

The entropy $h(x)$ of a continuous random variable x with density function $p(x)$ is given by

$$h(x) = \int p(x) \log_2 p(x) \, dx$$

The conditional entropy $h(x \mid y)$

$$h(x \mid y) = \int p(x, y) \log_2 p(x, y) \, dxdy$$

quantifies how much uncertainty about variable x remains after observing y.
If x and y are independent

$$h(x \mid y) = h(x)$$

Mutual Information

The mutual information $I(x;y)$ of x and y is a measure of how much we learn about x given y, and since mutual information is symmetric, i.e. since

$$I(x;y) = I(y;x)$$

it is also how much we learn about y from observing x. Formally, the mutual information of x and y are

$$\begin{aligned} I(x;y) &= h(x) - h(x \mid y) \\ &= h(y) - h(y \mid x) \\ &= I(y;x) \end{aligned}$$

Mutual information is independent of scale, i.e. is unaffected by units of measurement.

Bits and precision: Some intuition

Let's say an agent has a uniform prior about θ $U(-100, 100)$

- It takes one bit for to transmit whether θ is positive or negative
- It takes one more bit for to transmit whether θ is larger or smaller than 50 (-50)
- It takes one more bit for to transmit whether θ is larger or smaller than 25 (75,-25,-75)
- ...and so on.

With infinite capacity, we could partition the states of the world to an infinite precision

Gaussian entropy

The (log) entropy of a Gaussian random vector $X_t \sim N(0, \Sigma)$ is given by

$$\ln h(X_t) = \frac{n}{2}\left[\ln 2\pi + 1\right] + \frac{1}{2}\ln|\Sigma|$$

where n is the dimension of X so the interesting part is the determinant of the covariance matrix Σ

Gaussian entropy

For given variances entropy is maximized if X_t is a vector of uncorrelated variables:

$$\begin{aligned}
|\Sigma| &= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 1 - 0^2 \\
|\Sigma| &= \left| \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \right| = 1 - a^2 \\
-1 &< a < 1 \implies 1 - a^2 < 1
\end{aligned}$$

Intuition: If $a = -1$ or $a = 1$ we can perfectly transmit a two dimensional vector using a one dimensional signal

Gaussian signals and states

In economic (Gaussian) applications the constraint often takes the form

$$\begin{aligned} h(X) - h(X \mid Z) &\leq e^{\frac{1}{2}\kappa} \\ \left(\frac{n}{2}\left[\ln 2\pi + 1\right] + \frac{1}{2}\ln\left|\Sigma_{prior}\right|\right) - \left(\frac{n}{2}\left[\ln 2\pi + 1\right] + \frac{1}{2}\ln\left|\Sigma_{post}\right|\right) &\leq \frac{1}{2}\kappa \\ \ln\left|\Sigma_{prior}\right| - \ln\left|\Sigma_{post}\right| &\leq \kappa \end{aligned}$$

that is, the decrease in uncertainty is cannot be too large.
In a Kalman filter setting, the above constraint would be

$$\ln\left|P_{t|t-1}\right| - \ln\left|P_{t|t}\right| \leq \kappa$$

The "*No forgetting*" constraint

Posterior uncertainty cannot be larger than prior uncertainty

$$diag(P_{t|t-1}) >> diag(P_{t|t})$$

That is, an agent cannot intentionally increase the variance of the estimation error of one variable in order to get a more precise estimate of another variable.

Rational inattention and entropy in economics

Some nice properties:

1. When information processing capacity is large, behavior is close to full information
2. When a decision maker allocates a lot of attention to observing one variable, mistakes in responses to that variable becomes small
3. A decision maker need to allocate more attention to a variable (with given variance) to achieve a given precision if the variable has low serial correlation

A simple example of optimal information choice

Utility function

$$U = -E\left[(1-\lambda)(a-x_1)^2 + \lambda(a-x_2)^2\right] : 0 < \lambda < 1$$

FOC:

$$2(1-\lambda)(a-x_1) + 2\lambda(a-x_2) = 0$$

so that

$$a = (1-\lambda)E[x_1] + \lambda E[x_2]$$

Plugging in a into U gives the expected loss

$$EU = (1-\lambda)^2\sigma_1^2 + \lambda^2\sigma_2^2$$

where σ_1^2 and σ_2^2 are the posterior error variances of the estimates of x_1 and x_2.

A simple example of optimal information choice

Choose noise in signal Z to maximize expected utility (i.e. minimize expected loss)

$$Z = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v_t$$

subject to

$$\ln |\Sigma_{prior}| - \ln |\Sigma_{post}| \leq \kappa$$

A simple example of optimal information choice

For simplicity, we can restrict ourselves to diagonal covariance matrices

$$\begin{aligned}
\ln|\Sigma_{prior}| - \ln|\Sigma_{post}| &\leq \kappa \\
\frac{|\Sigma_{prior}|}{|\Sigma_{post}|} &\leq e^{\kappa} \\
\frac{|\Sigma_{prior}|}{e^{\kappa}} &\leq |\Sigma_{post}| = \sigma_1^2\sigma_2^2 \\
\frac{|\Sigma_{prior}|}{e^{\kappa}}\left(\sigma_1^2\right)^{-1} &\leq \sigma_2^2
\end{aligned}$$

(If Σ_{prior} is diagonal, the optimal Σ_{post} is also diagonal.)

A simple example of optimal information choice

Use that inequality will always be binding and plug into expected loss

$$(1-\lambda)^2 \sigma_1^2 + \lambda^2 \frac{|\Sigma_{prior}|}{e^{\kappa}} \left(\sigma_1^2\right)^{-1}$$

F.o.c.

$$\begin{aligned} (1-\lambda)^2 - \lambda^2 \frac{|\Sigma_{prior}|}{e^{\kappa}} \left(\sigma_1^2\right)^{-2} &= 0 \\ \frac{\lambda}{(1-\lambda)} |\Sigma_{prior}|^{\frac{1}{2}} e^{-\frac{1}{2}\kappa} &= \sigma_1^2 \end{aligned}$$

What happens when $\lambda \to 1$? And when $\lambda \to 0$?

Remember the *"No forgetting"* constraint!

Make sure that

$$diag(P_{t|t-1}) >> diag(P_{t|t})$$

by checking that

$$\frac{\lambda}{(1-\lambda)} \left|\Sigma_{prior}\right|^{\frac{1}{2}} e^{-\frac{1}{2}\kappa} < \sigma^2_{1,prior}$$

If inequality is violated at marginal condition, set $\sigma_1^2 = \sigma^2_{1,prior}$.

Application: A macro price setting model

Mackowiak and Wiederholt (AER 2009): *Optimal Sticky Prices under rational Inattention*

- Sets up a macro price setting model with rational inattention
- Explains large and persistent real effects of shocks to money supply
- Can match micro data on large magnitude of individual price changes, while aggregate price changes are small

Application: A macro price setting model

The optimal price of a good produced by firm i is given by

$$p_{it} = E\left[p_t + ay_t + bz_{it} \mid s_{ti}\right] \quad (1)$$

where p_t is the aggregate price level, y_t are aggregate shocks and z_{jt} are firm specific shocks.
The signal vector s_{ti} is chosen to minimize

$$L_i = cE\left[\left(X_t - X_{t|t}\right)\left(X_t - X_{t|t}\right)'\right]c' \quad (2)$$

subject to

$$I\left(X_t; s_{ti}\right) \leq K \quad (3)$$

where

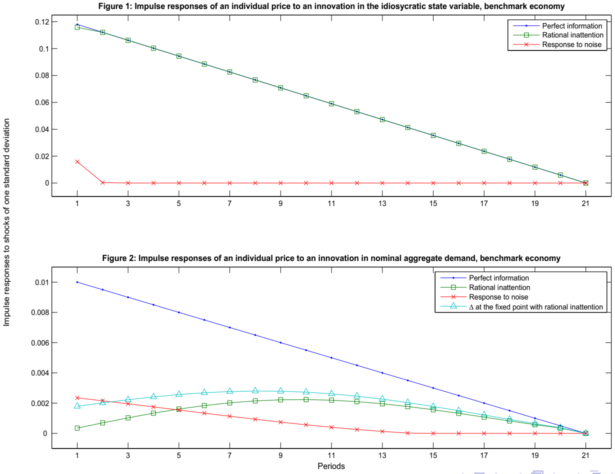
$$X_t = \begin{bmatrix} p_t & y_t & z_{it} \end{bmatrix}' \quad (4)$$

Mackowiak and Wiederholt 2009

Some important properties of the model:

- Optimal price in increasing in aggregate price level
- Idiosyncratic shocks are large relative to aggregate shocks
- Capacity K is set so that pricing errors are quite small
- Observing endogenous signals (like the price level) also uses up capacity.

Price responses to idiosyncratic and aggregate shocks



Pricing errors and the importance of feedback

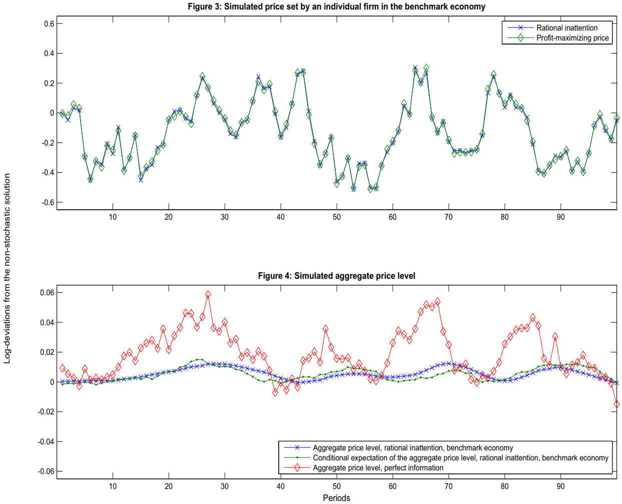


Figure 5: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand

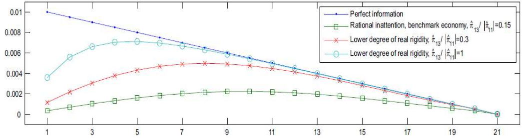


Figure 6: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand

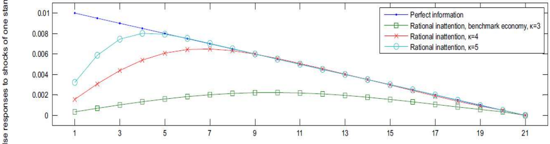
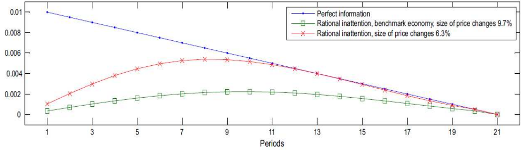


Figure 7: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand



Critique, limitations and further issues

The rational inattention framework yields nice results, and as a theory of information processing it has many appealing features. However, there are also some unresolved issues:

- How can the capacity constraint K be calibrated? In engineering, this is usually a physical constraint. What does it mean for humans? Are we Gaussian channels?
- Can processed information be traded?
- Is it realistic that entropy alone determines how hard a variable is to form an estimate about?

Rational inattention makes testable, but as of yet untested, predictions:

- If policy changes, agents should reallocate attention in a way that is optimal, given the new stochastic environment. This could perhaps be investigated using the monetary policy change in the early 1980's in the US, or the switch to inflation targeting in several other countries in the early 90's.

Other approaches

Rational inattentiveness: Mankiw and Reis (2002), Reis (2006a,2006b)

- Agents update information sets infrequently
- Random (Calvo-type) determination of exactly when updates occur, but frequency can be micro founded with fixed cost of gathering information
- When agents update, they observe state perfectly
  - Results in hierarchical information sets and keeps solution simple

Is it plausible?

- Actions changes even when information is not updated which implies smooth trajectories with intermittent jumps
- Since all agents that update in period t will choose the same action, you will have clustered actions which is not what we see in the data

But then again, maybe it is a tractable short cut that captures aggregate dynamics well