# Endogenous Information Choice

February 29, 2016

# Endogenous Information Choice

Today we will talk mostly about "rational inattention"

- Agents can choose what to observe, i.e. choose $D$ and $\Sigma_{vv}$ 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 $\Sigma_{vv}$
- 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
- 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

The entropy $H(x)$ of a random variable $x$ with probability mass function $p(x)$ is given by

$$H(x) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$$

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

$$H(x \mid y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 p(x \mid y)$$

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.

# Differential Entropy

Differential entropy $h(x)$ is the generalization of entropy to continuous random variables.

Differential entropy of the random variable $x$ with density function $p(x)$ is given by

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

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

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

quantifies how much uncertainty about variable $x$ remains after observing $y$.

If $x$ and $y$ are independent

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

# Coin flip

Coin flip (equal probability binomial)

$$H(x) = -\frac{1}{2}\ln\left(\frac{1}{2}\right) - \frac{1}{2}\ln\left(\frac{1}{2}\right) \\ = \ln(2)$$

where ln(2)=1 if the logarithm is taken w.r.t. base 2.

# Uniform

Entropy of uniform $x \sim U(0, a)$ is given by

$$\begin{aligned} h(x) &= -\int_0^a \frac{1}{a} \ln\left(\frac{1}{a}\right) dx \\ &= -\ln\left(\frac{1}{a}\right) = \ln a \end{aligned}$$

The larger the support, the larger the entropy.

# Gaussian entropy

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

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

where $n$ is the dimension of $x$ so the interesting part is the determinant of the covariance matrix $\Sigma$

# Gaussian entropy

For given variances entropy is maximized if $x$ 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

$$diag(\Sigma_{prior}) > diag(\Sigma_{post})$$

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:

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

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 $\sigma_1^2$ and $\sigma_2^2$ 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)

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

subject to

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

For simplicity, we can restrict ourselves to diagonal covariance matrices

# A simple example of optimal information choice

The capacity constraint

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

If $\Sigma_{prior}$ is diagonal, the optimal $\Sigma_{post}$ is also diagonal.

# A simple example of optimal information choice

Use that inequality will always be binding in an interior solution 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

$$diag(\Sigma_{prior}) > diag(\Sigma_{post})$$

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: Portfolio problem

Based on *Information Immobility and the Home Bias Puzzle* by van Nieuwerburgh and Veldkamp (JoF 2008)

Why do investor not diversify their portfolios internationally?

# 3 period asset pricing model

Time line:

1. Agents choose information
2. Agents purchase assets
3. Returns are realized

# Preferences

Mean-variance preferences

$$U = -E\left[-\rho q'\left(f - rp\right) + \frac{\rho^2}{2} q'\widehat{\Sigma} q\right]$$

- $r$ is risk free rate
- $f$ is vector of pay-offs
- $p$ is vector of prices
- $q$ is vector of quantities
- $\rho$ is risk aversion
- $\widehat{\Sigma}$ is conditional covariance of pay-offs

# Initial information

2 countries *Home* and *Foreign*

- Home priors beliefs are $\mu \sim N(f, \Sigma)$
- Foreign priors beliefs are $\mu^* \sim N(f, \Sigma^*)$

Home investor have lower variance of priors on home asset returns and foreign investors have lower variance on prior on foreign asset returns

# Information acquisition

Each investor chooses a vector $\eta$ to observe where $\eta \sim N(f, \Sigma_\eta)$ subject to the capacity constraint

$$\left|\widehat{\Sigma}\right| \geq \frac{1}{K} |\Sigma|$$

where

$$\widehat{\Sigma} = \left[\Sigma^{-1} + \Sigma_\eta^{-1}\right]^{-1}$$

and $K \geq 1$, $\Sigma_\eta$ is p.s.d.
With assets indexed by $i$ we can then write the constraints as

$$\begin{aligned} \prod \widehat{\sigma}_i^2 &\geq \frac{1}{K} \prod \sigma_i^2 \\ \sigma_{i\eta}^2 &\geq 0 \text{ (no forgetting)} \end{aligned}$$

# Posterior beliefs

The posterior mean is given by

$$
\begin{aligned}
\widehat{\mu}^j &\equiv \left[f \mid \mu^j, \eta^j, p\right] \\
&= \left(\left(\Sigma^j\right)^{-1} + \left(\Sigma_\eta^j\right)^{-1} + \Sigma_p^{-1}\right)^{-1} \\
&\quad \times \left(\left(\Sigma^j\right)^{-1} \mu^j + \left(\Sigma_\eta^j\right)^{-1} \eta^j + \Sigma_p^{-1} (rp - A)\right)
\end{aligned}
$$

and the posterior variance is

$$
\begin{aligned}
\widehat{\Sigma}^j &\equiv Var\left[f \mid \mu^j, \eta^j, p\right] \\
&= \left(\left(\Sigma^j\right)^{-1} + \left(\Sigma_\eta^j\right)^{-1} + \Sigma_p^{-1}\right)^{-1}
\end{aligned}
$$

# Equilibrium

Demand by investor $j$

$$q^j = \frac{1}{\rho}\left(\widehat{\Sigma}^j\right)^{-1}\left(\widehat{\mu}^j - pr\right)$$

Equate aggregate demand with supply

$$\int q^j \, dj = \overline{x} + x : x \sim N\left(0, \sigma_x^2 \times I\right)$$

# Optimal Information acquisition

Expected utility is given by

$$U = E\left[\frac{1}{2}\left(\widehat{\mu}^j - pr\right)'\left(\widehat{\Sigma}^j\right)^{-1}\left(\widehat{\mu}^j - pr\right)\right]$$

resulting in agent j's objective

$$\max_{\widehat{\sigma}^2_{i,j}} \sum_i \left(\sigma^2_{pi} + \left(\rho\overline{x}\sigma^2_{i,a}\right)^2\right)\left(\widehat{\sigma}^2_{i,j}\right)^{-1}$$

where $\sigma^2_{i,a} \equiv \left(\int_j \left(\widehat{\sigma}^2_{i,j}\right)^{-1}\right)^{-1}$

# Increasing returns to information

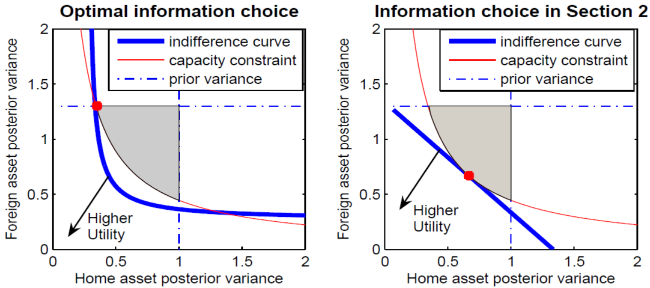



Figure 1: Objective and constraints in the optimal learning problem with 2 risk factors.

# 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

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

where $p_t$ is the aggregate price level, $y_t$ are aggregate shocks and $z_{jt}$ are firm specific shocks.

# Application: A macro price setting model

The signal vector $s_{ti}$ is chosen to minimize

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

subject to

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

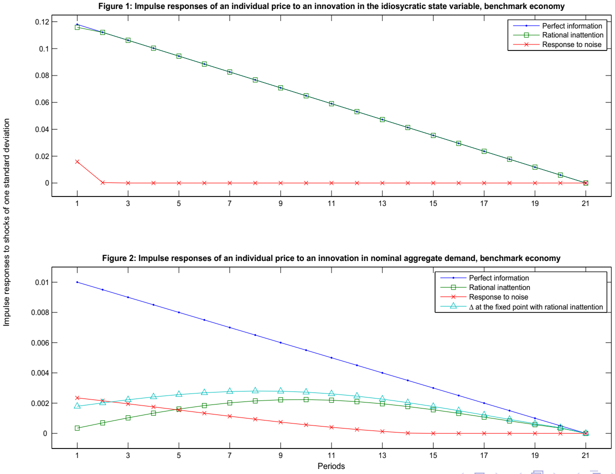
where

$$X_t = \begin{bmatrix} p_t & y_t & z_{it} \end{bmatrix}'$$

# 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

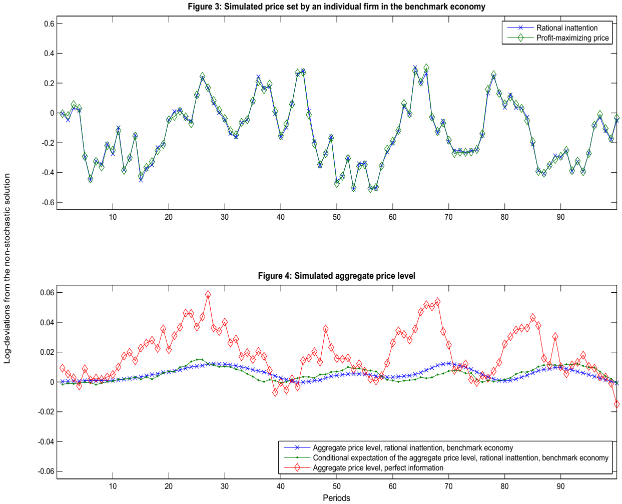


Figure 5: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand

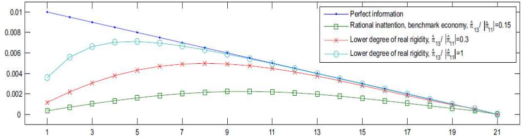



Figure 6: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand

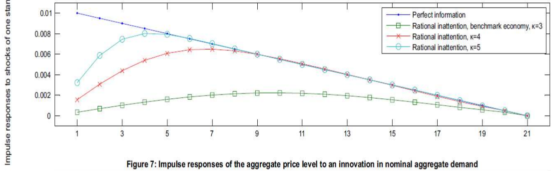
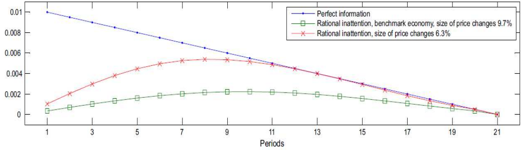



Figure 7: Impulse responses of the aggregate price level to an innovation in nominal aggregate demand



# 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?
- There appears to be markets for information. How do they fit in?
- Is it realistic that entropy alone determines how hard a variable is to form an estimate about?

# Critique, limitations and further issues

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