TOPICS IN MACROECONOMICS: MODELLING INFORMATION, LEARNING AND EXPECTATIONS

LECTURE NOTES 7

KRISTOFFER P. NIMARK

1. Entropy, information processing constraints and rational inattention

Chris Sims introduced information processing capacity constraints to economics, building on work by Shannon (1948) and dubbing it "rational inattention". For a recent text book treatment of information theory and its applications, see Cover and Thomas (2006). The approach borrows methods, concepts and technology from information theory, a branch of applied mathematics asking questions like *What is the most efficient compression to transmit messages without error?* and *How can information flow be quantified independently of the medium of transmission?* We start by introducing a few concepts and discuss how they are related to things more familiar to us.

1.1. **Entropy.** Entropy is a term taken from physics, and more specifically, thermodynamics. It was originally used to describe how far from equilibrium a physical system was, where equilibrium was characterized by equal pressure, temperature etc. Entropy is maximized in equilibrium. In statistics, the term entropy is used to describe uncertainty. A variable with high entropy requires more information to be described accurately. One way to think of entropy which works pretty well for applications in economics is simply as uncertainty, but the formal definition is how much information is needed *on average* to describe the outcome of a random variable. The more predictable a variable is, the lower is the entropy of its probability density.

Date: October 21, 2008.

To be more formal, the entropy $h(x)$ of a continuous random variable x with density function $p\left(x\right)$ is given by

$$h\left(x\right)=\int p\left(x\right)\log_{2}p\left(x\right)dx \tag{1.1}$$

where the log is with base 2, by tradition since this allows entropy to be quantified as bits.

Conditional entropy, $h\left(x\mid y\right)$ is defined as

$$h\left(x\mid y\right)=\int p\left(x,y\right)\log_{2}p\left(x,y\right)dxdy \tag{1.2}$$

The conditional entropy $h\left(x\mid y\right)$ quantifies how much uncertainty about variable x remains after observing y. Of course, if x and y are independent

$$h\left(x\mid y\right)=h\left(x\right) \tag{1.3}$$

This leads us to the concept of mutual information.

1.2. **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

$$I(x;y)=I(y;x) \tag{1.4}$$

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

$$\begin{aligned} I(x;y) &= h\left(x\right)-h\left(x\mid y\right) && (1.5)\\ &= h\left(y\right)-h\left(y\mid x\right) && (1.6)\\ &= I(y;x) && (1.7)\end{aligned}$$

The mutual information of two variables are naturally thought of as correlations: If the correlation is 1, the mutual information is the same as the entropy of each variable (which then also must be the same) and the conditional entropy must be zero, i.e. if x and y are perfectly correlated, there is no uncertainty about x once y has been observed.

One way to introduce information processing constraints is to model agents as having information processing constraints that takes the form of putting a an upper bound on the reduction in entropy that is possible by observing signals. That is, the posterior variance cannot be too small compared to the prior variance. In terms of mutual information, we then limit the mutual information between the signal and unobservable state. We can formalize this in the following way (see van Nieuwerburgh and Veldkamp (2008)). A channel with capacity K cannot transmit information more accurately than that

$$\frac{|\Sigma_{prior}|}{|\Sigma_{post}|} \leq K \tag{1.8}$$

holds. That is, the generalized posterior variance $|\Sigma_{post}|$ cannot be too small relative to the prior generalized uncertainty $|\Sigma_{prior}|$. The expression (1.8) should also make it clear that we cannot ”cheat” and try to increase capacity by rescaling the variables. That is, entropy is invarianbt to scaling, whichis a desirebly feature, as it the unit of measurement should not matter for uncertainty.

1.3. **A Kalman filter formulation.**

To make all this a bit more concrete, let us reformulate the constraint in a Kalman filter context. As usual, we will consider a system of the form

$$X_t = AX_{t-1} + C\mathbf{u}_t \tag{1.9}$$

$$Z_t = DX_t + \mathbf{v}_t \tag{1.10}$$

and define $P_{t|t-1}$ and $P_{t|t}$ as the prior and posterior error covariance matrix respectively

$$\begin{aligned} P_{t|t-1} &= E\left(X_t - E\left[X_t \mid Z^{t-1}\right]\right)\left(X_t - E\left[X_t \mid Z^{t-1}\right]\right)' \\ P_{t|t} &= E\left(X_t - E\left[X_t \mid Z^{t}\right]\right)\left(X_t - E\left[X_t \mid Z^{t}\right]\right)' \end{aligned}$$

We can then formulate the constraint (1.8) as

$$\frac{\left|P_{t|t-1}\right|}{\left|P_{t|t}\right|} \leq K \tag{1.11}$$

or in logs

$$\ln\left|P_{t|t-1}\right| - \ln\left|P_{t|t}\right| \leq \ln K \tag{1.12}$$

which sometime makes solving a model easier. The endogenous decision of the agents is then to choose D and Σ_{vv} in order to maximize utility, while respecting the constraint (1.12). Of course, to maximize utility we need to specify a model, and we now turn to two applications of this framework.

2. Information Acquisition and Under-Diversification (van Nieuwerburgh and Veldkamp 2008)

van Nieuwerburgh and Veldkamp (2008) present a static model where CARA utility investors can choose between multiple risky assets and also choose which assets to receive signals about as well as how precise these signals are. The cost of information is a utility cost of lost leisure denoted $c(K)$, where K is the reduction in entropy from priors to posteriors. Investor i's utility is given by

$$U_i = -E\left[\exp\left(-\rho W_i\right)\right] - c(K) \tag{2.1}$$

where end of period wealth W_i follows

$$W_i = W_{0i}r - q_i'\left(f - pr\right) \tag{2.2}$$

r is the risk free rate, q_i is vector of quantities of each asset held by investor i, p is a vector of prices and f is vector of asset returns. Denote the prior (or unconditional) return variance Σ and the posterior conditional return variance $\widehat{\Sigma}$. Investors then optimize by (sequentially) choosing a signal covariance matrix Σ_{vv} of the measurement errors of the signal vector Z

$$Z = f + \mathbf{v} \tag{2.3}$$

$$E\mathbf{v}\mathbf{v}' = \Sigma_{vv} \tag{2.4}$$

to maximize the expected utility (2.1) subject to the budget constraint (2.2) and the information flow constraint

$$\frac{|\Sigma|}{\left|\widehat{\Sigma}\right|} \leq K \tag{2.5}$$

In the case of independent asset returns, the return covariance matrices are diagonal and we can write out the constraint as

$$\ln \sigma_1^2 + \ln \sigma_2^2 + \ldots \ln \sigma_n^2 - \ln \widehat{\sigma}_1^2 + \ln \widehat{\sigma}_2^2 + \ldots \ln \widehat{\sigma}_n^2 \leq \ln K \tag{2.6}$$

where the $\widehat{\sigma}^2 s$ are choice variables with the additional constraint $\ln \sigma_n^2 \geq \ln \widehat{\sigma}_n^2$, that is, uncertainty cannot increase after observing a signal. In the second stage of optimization, investors receive the realizations of the signals that they have chosen, and make portfolio decisions. That is, after receiving the signal investors choose how much of their wealth to allocate to the safe asset and each of the risky assets.

2.1. **A simple example.** In the case of there being only one risky asset, the problem simplifies to choosing K, and allocating all processing capacity to the single risky asset. To solve the model, start from the end, i.e. from the portfolio decision's f.o.c.

$$q_i = \frac{E\left[f_1 \mid Z\right] - rp}{\rho} \widehat{\sigma}_1^{-2} \tag{2.7}$$

Plug into utility function

$$\begin{aligned}
EU_i &= -\exp\left(-\rho\left[RW_{0i}+q_i\left[E\left(f_1\right)-rp\right]-\left(\frac{\rho}{2}\right)^2 q_i^2\widehat{\sigma}_1^2\right]\right)-c(K) \quad (2.8)\\
&= -\exp\left(-\rho\left[RW_{0i}+q_i\left[E\left(f_1\right)-rp\right]-\left(\frac{\rho}{2}\right)^2\left(\frac{E\left[f_1\right]-rp}{\rho}\widehat{\sigma}_1^{-2}\right)^2\widehat{\sigma}_1^2\right]\right)-c(K) \quad (2.9)\\
&= -\exp\left(-\rho\left[RW_{0i}+q_i\left[E\left(f_1\right)-rp\right]-\frac{1}{4}\left(E\left[f_1\right]-rp\right)^2\widehat{\sigma}_1^2\right]\right)-c(K) \quad (2.10)
\end{aligned}$$

The f first order condition then is

$$\frac{\partial EU_i}{\partial K}=0 \quad (2.11)$$

or

$$\begin{aligned}
-\frac{1}{4}\left(E\left[f_1\right]-rp\right)^2\frac{\sigma_1^2}{K^2}+c'(K) &= 0 \quad (2.12)\\
&\Leftrightarrow \quad (2.13)\\
K &= \frac{1}{2}\sqrt{\frac{\left(E\left[f_1\right]-rp\right)^2}{c'(K)\sigma_1^2}} \quad (2.14)
\end{aligned}$$

Solving for K yields

$$K=\frac{1}{2}\sqrt{\frac{\left(E\left[f_1\right]-rp\right)^2}{c'(K)\sigma_1^2}} \quad (2.15)$$

Which can be plugged into the second round of optimization, i.e. in the portfolio choice f.o.c.

$$q_i=\frac{E\left[f_1\mid Z\right]-rp}{\rho\sigma_1^2/K} \quad (2.16)$$

2.2. **Multiple risky assets.** van Nieuwerburgh and Veldkamp show that solving for the optimal allocation of information processing capacity across signals and the portfolio problem when there are multiple risky assets can be done in a similar way. They derive the following substantive results:

- Independent asset returns: Investor choose to allocate all their capacity to processing information about a single asset which investors also choose to hold a larger fraction of in his portfolio. I.e. information choice leads to specialization and asset concentration. The intuition is the following. When agents have a more precise signal of the return of an asset, they expect to hold more of it so choosing to purchase information about a particular asset makes it more likely that you will hold a larger fraction of your wealth in that asset. But you are also more likely to buy more information about an asset that you are more likely to have a large fraction of your wealth invested in, so this creates a feedback effect resulting in specialization in information acquisition.

- Correlated asset returns: Correlated asset return generate a similar effect, except that investors then choose to specialize in information about a single risk factor, where the risk factors are the orthogonal components of the asset return. Asset allocations are on average concentrated in assets with returns that are to a large part explained by the risk factor the investor has chosen to process more information about.

3. Optimal Sticky Prices under Rational Inattention (Mackowiak and Wiederholt 2008)

Mackowiak and Wiederholt (2008) presents a macro economic price setting model where the optimal price of a good produced by firm i is given by

$$p_{it} = E\left[p_t + ay_t + bz_{it} \mid s_{ti}\right] \tag{3.1}$$

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 by firms in order to minimize the loss function

$$L_i = cE\left(X_t - X_{t|t}\right)\left(X_t - X_{t|t}\right)c' \tag{3.2}$$

subject to

$$I\left(X_t; s_{ti}\right) \leq K \tag{3.3}$$

where

$$X_t = \begin{bmatrix} p_t & y_t & z_{it} \end{bmatrix}' \tag{3.4}$$

That is, firms minimize a quadratic loss function that is decreasing in the posterior state estimate error variance, subject to a constraint on the maximum (K) mutual information $I(X_t; s_{ti})$ of their signals and the true state. Mackowiak and Wiederholt uses the model to show that it can be rational for agent to allocate more attention to firm specific shocks and less to aggregate conditions, and that this can explain the slow response of the aggregate price level to monetary policy shocks, while still accounting for the large changes in individual goods prices observed in the data.

4. 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?
- Can processed information be traded?
- Is it realistic that the only thing determining how hard a variable is to form an estimate about is its entropy? What about the CPI versus the output gap? If you ask any central banker, he will tell you that getting an accurate estimate of the output gap is harder than getting an accurate estimate of the CPI, and for reason unrealted to these variables entropy.

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 invstigated 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.

5. Other approaches

In this note we focused on a particular way of modelling information choice, that is, modelling agents as having finite information processing constraints. This choice should not be taken to imply that other approaches are not worth studying. Another approach, with a deceptively similar name is the "inattentiveness literature", most prominently advocated by Ricardo Reis (2007, 2006a, 2006b). In the inattentiveness literature, agents choose when to receive information, but not how precise it is. (It is assumed that when agents get a signal, they get a perfect signal of the state.) With a fixed cost of acquiring information it turns out to be optimal to sample information only infrequently. This approach provides a tractable way of modelling information choice in a dynamic setting, but is subject to some of the same critique as the rational inattention literature.

References

[1] Cover and Thomas, 2006, *Elements of Information Theory*, Wiley-Interscience.

[2] Mackowiak and Wiederholt, forthcoming, Optimal Sticky Prices under rational inattention, *AER*.

[3] Reis, Ricardo and N. Gregory Mankiw., 2007, Sticky Information in General Equilibrium, *Journal of the European Economic Association*, 5 (2-3), 603-613.

[4] Reis, Ricardo, 2006a, Inattentive Consumers. *Journal of Monetary Economics*, 53 (8), 1761-1800.

[5] Reis, Ricardo, 2006b, Inattentive Producers. *Review of Economic Studies*, 73 (3), 793-821.

[6] Sims, Chris, 2003, Implications of rational inattention, *Journal of Monetary Economics*, p665-690.

[7] Shannon, Claude, A mathematical theory of communications, 1948.

[8] van Nieuwerburgh, S and L. Veldkamp, 2008, Information Acquisition and Under-Diversification, *mimeo*, Stern/NYU.