Truth and Probability

Thomas J. Sargent

May 2008

Murphy's law, statistical version

The probability of anything happening is in inverse ratio to its desirability.

Bernanke, 2007: great speech because . . .

- Probabilities
- Utility functions
- Bayesian model averaging
- Model selection
- Robust control and multiple priors

Bernanke, October 2007

- He mentions Bayesian subjective expected utility analyses and highlights a decision maker's *unique probability density* over states of the world, which for the monetary policy maker is a dynamic model of the economy.
- A Bayesian analysis also involves a *utility function* over states, which for a monetary policy maker is a welfare function to be used as an ingredient of an optimal policy problem. A timing protocol captures whether the policy maker has a commitment technology.
- He mentions multiple-priors models cast in terms of robust control theory. Here a decision maker is ambiguous about probability distributions and cannot elude that ambiguity by putting a prior over multiple models that would reduce those multiple models to a single model.

Bernanke, October 2007

- He describes how a multiple-priors (a.k.a. robust) decision maker engages in a worst-case analysis over probability distributions as a device for designing a policy rule that works well over the range of probability models that express his ignorance.
- I read passages in his remarks as describing policy procedures that first do *model selection*, then make decisions that maximize expected utility using the winning model.
- I read other passages as alluding to *Bayesian model averaging* that effectively always keeps all models in play because Bayes's Law is too forgiving ever completely to destroy an ill-fitting model.

Bernanke, October 2007

- He discusses how Bayesian and multiple-priors robust control models often recommend decisions that are more aggressive than ones that would be made if a policy maker having more confidence in a reduced set of models.
- Under concerns about robustness, highly serially correlated worst-case shocks rationalize the policy maker's more aggressive policy stance.

Objects in play

There is a finite space of states $\mathcal{I} = \{i = 1, \ldots, I\}$. A (consumption) plan is a function $c : \mathcal{I} \to \mathbb{R}$. Let π be an $I \times 1$ vector of nonnegative probabilities over states and $u : \mathbb{R} \to \mathbb{R}$ be a utility function. The relative entropy of a probability vector $\hat{\pi}$ with respect to probability vector π is the expected value of the likelihood ratio $m_i = \left(\frac{\hat{\pi}_i}{\pi_i}\right)$ under the $\hat{\pi}$ distribution:

$$ent(\pi, \hat{\pi}) = \sum_{i=1}^{I} \hat{\pi}_i \log\left(\frac{\hat{\pi}_i}{\pi_i}\right) = \sum_{i=1}^{I} \pi_i \left(\frac{\hat{\pi}_i}{\pi_i}\right) \log\left(\frac{\hat{\pi}_i}{\pi_i}\right)$$

or

$$ent(\pi, \hat{\pi}) = \sum_{i=1}^{I} \pi_i m_i \log m_i.$$

Expected utility

A decision maker is said have *expected utility* preferences when he rank plans c by their expected utility

$$\sum_{i=1}^{I} u(c_i)\pi_i$$

where u is a unique utility function and π is a unique probability measure.

Constraint preferences

A decision maker is said to have *constraint preferences* when he ranks plans c according to

$$\min_{\{m_i \geq 0\}_{i=1}^{I}} \sum_{i=1}^{I} m_i \pi_i u(c_i)$$

where the minimization is subject to

$$\sum_{i=1}^{I} \pi_i m_i \log m_i \leq \eta$$

and

$$\sum_{i=1}^{I} \pi_i m_i = 1.$$

Here $\eta \geq 0$ specifies the size of an entropy-ball of probability distributions $\hat{\pi}$ surrounding a baseline distribution π.

Multiplier preferences

A decision maker is said to have *multiplier preferences* when he ranks consumption plans c according to

$$\min_{\{m_i \geq 0\}_{i=1}^{I}} \sum_{i=1}^{I} \pi_i m_i [u(c_i) + \theta \log m_i]$$

where the minimization is subject to

$$\sum_{i=1}^{I} \pi_i m_i = 1.$$

Here $\theta \in (\underline{\theta}, +\infty)$ is a parameter that, by penalizing choices of m_i that enlarge entropy, expresses the decision maker's concern about possible misspecification of the baseline model π.

Slopes of indifference curves

- Expected utility:
$$\frac{dc_2}{dc_1} = -\frac{\pi_1}{\pi_2}\frac{u'(c_1)}{u'(c_2)}$$
- Constraint preferences:
$$\frac{dc_2}{dc_1} = -\frac{\hat{\pi}_1}{\hat{\pi}_2}\frac{u'(c_1)}{u'(c_2)}$$
where $\hat{\pi}_1, \hat{\pi}_2$ are the minimizing probabilities.
- Multiplier preferences:
$$\frac{dc_2}{dc_1} = -\frac{\pi_1}{\pi_2}\frac{\exp(-u(c_1)/\theta)}{\exp(-u(c_2)/\theta)}\frac{u'(c_1)}{u'(c_2)}$$

Entropy

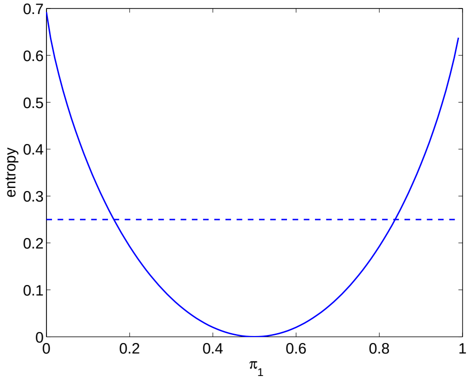

Figure: Entropy as a function of $\hat{\pi}_1$ when $\pi_1 = .5$.

Level curves

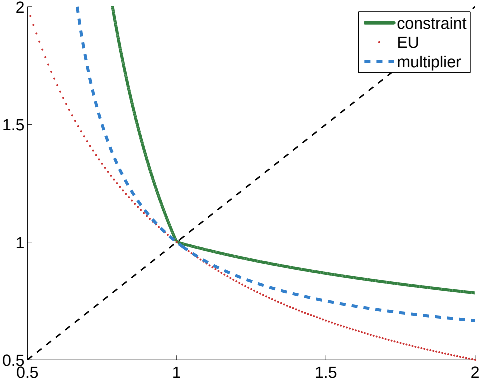

Figure: Indifference curves for expected utility, multiplier (smooth), and constraint (kinked at 45 degree line) preferences. The worst case probability $\hat{\pi}_1 < .5$ when $c_1 > c_2$ and $\hat{\pi}_1 > .5$ when $c_1 < c_2$.

Level curves

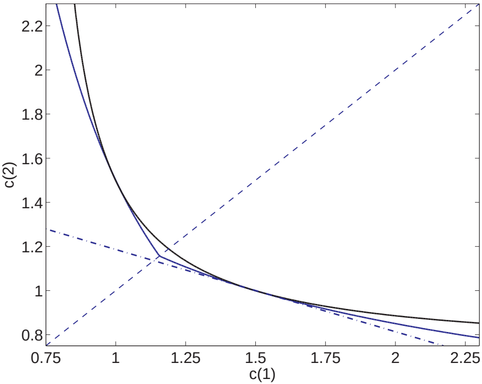

Figure: Indifference curves for multiplier (smooth) and constraint (kinked at 45 degree line) preferences. The worst case probability $\hat{\pi}_1 < .5$ when $c_1 > c_2$ and $\hat{\pi}_1 > .5$ when $c_1 < c_2$.

Isoentropy curves

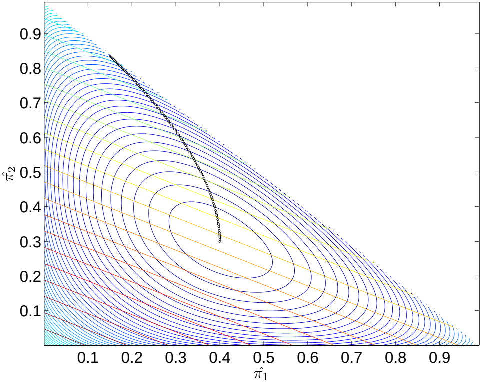

Figure: Iso-entropy and iso-expected utility, $\alpha = 0$. The 'expansion path', or locus of tangencies, shows the worst case probabilities associated with values of θ over the interval $\theta^{-1} \in [0, 2]$. Entropy increases and expected utility decreases as we move northwest along an expansion path.

Isoentropy curves

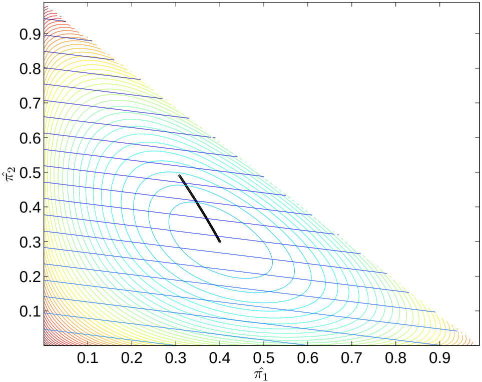

Figure: Iso-entropy and iso-expected utility, $\alpha = 3$. The 'expansion path', or locus of tangencies, shows the worst case probabilities associated with values of θ over the interval $\theta^{-1} \in [0, 2]$. Entropy increases and expected utility decreases as we move northwest along an expansion path.

Murphy's Law

Multiplier preferences can be represented with the indirect utility function

$$\mathcal{R}u(c) = -\theta \log \sum_{i=1}^{I} \pi_i \exp\bigl(-u(c_i)/\theta\bigr).$$

Evidently,

$$\mathcal{R}u(c) = \sum_{i=1}^{I} \pi_i \hat{m}_i [u(c_i) + \theta \log \hat{m}_i]$$

where

$$\hat{m}_i = \exp(-u(c_i)/\theta) / \Bigl(\sum_j \pi_j \exp(-u(c_j)/\theta)\Bigr)$$

attains the $\min_{\{m_i \geq 0\}_{i=1}^I} \sum_{i=1}^I \pi_i m_i [u(c_i) + \theta \log m_i]$, subject to $\sum_{i=1}^I \pi_i m_i = 1$. $\mathcal{R}u$ is called the *risk-sensitivity* operator.

Murphy's Law, 2

It follows from the definition of $\mathcal{R}$ that

$$\sum_{i=1}^{I} m_i \pi_i u(c_i) \geq \mathcal{R} u(c) - \theta \sum_{i=1}^{I} \pi_i m_i \log m_i.$$

Dynamics is a special case of statics

There is a set of states Ω with elements $\omega \in \Omega$. Let $c^t = c_t, c_{t-1}, \ldots, c_0$. A plan is now an infinite dimensional random sequence $^\infty$, with a time t component that is a measurable function of ω. A utility function $W : c^\infty \to \mathbb{R}$. The probability of states is described by a density $\pi(\omega)$ and relative entropy is

$$ent(\pi, \hat{\pi}) = \int \log\big(\hat{\pi}(\omega)/\pi(\omega)\big)\big(\hat{\pi}(\omega)/\pi(\omega)\big)\pi(\omega)d\omega.$$

Dynamics

Multiplier preferences are characterized by an indirect utility function like the one above constructed by applying the risk-sensitivity operator to W:

$$\mathcal{R}(W) = -\theta \log \int \exp\Big(-W(c^\infty(\omega))/\theta\Big)\pi(\omega)d\omega$$

and the associated worst-case probability density is

$$\hat{\pi}(\omega) = \pi(\omega)\exp\Big(-W(c^\infty(\omega))/\theta\Big) \Big/ \int \exp\Big(-W(c^\infty(\tilde{\omega}))/\theta\Big)d\tilde{\omega}$$

(Murphy's law)

Romer and Romer, 2008

Should monetary policymakers take the staff forecast of the effects of policy actions as given, or should they attempt to include additional information? This paper seeks to shed light on this question by testing the usefulness of the FOMCs own forecasts. Twice a year, the FOMC makes forecasts of major macroeconomic variables. FOMC members have access to the staff forecasts when they prepare their forecasts. We find that the optimal combination of the FOMC and staff forecasts in predicting inflation and unemployment puts a weight of essentially zero on the FOMC forecast and essentially one on the staff forecast: the FOMC appears to have no value added in forecasting. The results for predicting real growth are less clear-cut. We also find statistical and narrative evidence that differences between the FOMC and staff forecasts help predict monetary policy shocks, suggesting that policymakers act in part on the basis of their apparently misguided information. *[emphasis added by Sargent]*

Adapted Primiceri (2006) model

$$u_{t+1}^{*} = u_{t}^{*} + c_{u^{*}} w_{t+1}$$

where $w_{t+1} \sim \mathcal{N}(0, I), u_0^{*} \sim \mathcal{N}(\mu_0^{*}, \sigma_0^{*2})$. Inflation π_t and unemployment U_t are related to a policy variable v_t by

$$\begin{aligned} \pi_{t+1} &= \pi_t + \gamma_0 (U_t - u_t^{*}) + \gamma_1 (U_{t-1} - u_{t-1}^{*}) + c_\pi w_{t+1} \\ (U_{t+1} - u_{t+1}^{*}) &= \rho_1 (U_t - u_t^{*}) + \rho_2 (U_{t-1} - u_{t-1}^{*}) + v_t + c_U w_{t+1} \end{aligned}$$

The government's objective is the expected value of

$$-.5 \sum_{t=0}^{\infty} \beta^t \left((\pi_t - \pi^{*})^2 + \lambda (U_t - k u_t^{*})^2 + \phi (v_t - v_{t-1})^2 \right)$$

Bayes' Law (Kalman filter)

$$
\begin{aligned}
K_2(\Delta) &= (A_{22}\Delta D_2' + C_2 G')(D_2 \Delta D_2' + GG')^{-1} \\
\mathcal{C}(\Delta) &\equiv A_{22}\Delta A_{22}' + C_2 C_2' - K_2(A_{22}\Delta D_2' + C_2 G')'.
\end{aligned}
$$

Law of motion

In terms of the state variables $(y, \check{z}, \Delta)$, the law of motion for $(y, z, \check{z}, \Delta)$ can be written as

$$\begin{aligned}
y^* &= A_{11}y + A_{12}\check{z} + B_1 a + C_1 w^* + A_{12}(z - \check{z}) \\
z^* &= A_{21}y + A_{22}\check{z} + B_2 a + C_2 w^* + A_{22}(z - \check{z}) \\
\check{z}^* &= A_{21}y + A_{22}\check{z} + B_2 a + K_2(\Delta)Gw^* + K_2(\Delta)D_2(z - \check{z}) \\
\Delta^* &= \mathcal{C}(\Delta) \qquad (1)
\end{aligned}$$

where $w^* \sim \mathcal{N}(0, I)$ and $z - \check{z} \sim \mathcal{N}(0, \Delta)$.

Perturbations

- Distribution of w^*
- Distribution of $(z - \check{z})$ emerging from Bayes' Law.

Inflation forecasts (FOMC and staff)

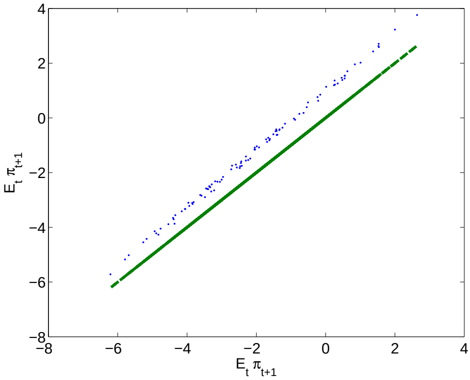

Figure: Worst case forecast $\tilde{E}_t\pi_{t+1}$ versus $E_t\pi_{t+1}$.

Unemployment forecasts (FOMC and staff)

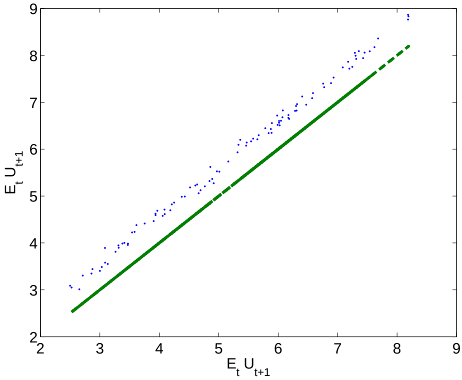

Figure: Worst case forecast $\tilde{E}_t U_{t+1}$ versus $E_t U_{t+1}$.

NAIRU forecasts (FOMC and staff)

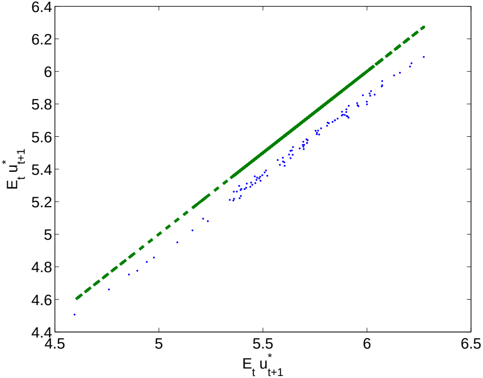

Figure: Worst case forecast $\tilde{E}_t u^*_{t+1}$ versus $E_t u^*_{t+1}$.

Unlikely worst-case?

The density for the approximating model is

$$\log c_{t+1} - \log c_t = \mu + \sigma_c \epsilon_{t+1}$$

where $\epsilon_{t+1} \sim \mathcal{N}(0,1)$ and μ and σ_c are estimated by maximum likelihood from the data in the histogram, data for 1948I-2006IV. The worst-case density has mean shift $-\sigma_c w$ where w is calculated by setting a detection error probability to .05. The worst case model appears to fit the histogram nearly as well as the approximating model.

Consumption growth

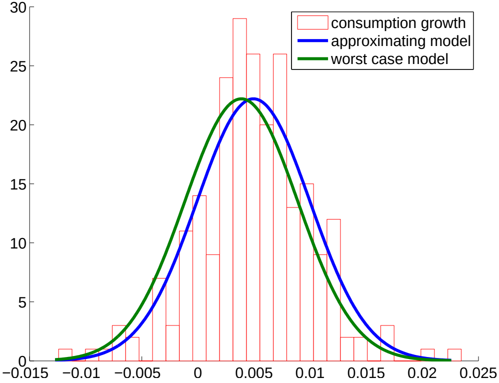

Figure: Histogram and maximum likelihood and worst-case densities for U.S. quarterly consumption growth, 1948I-2006IV.

Wedgeology

Consumption Euler equation:

$$E\beta \frac{u'(c_{t+1})}{c_t} R_{t+1} = 1.$$

Wedgeology

$$\sum_{t}\sum_{s^t} u(c(s^t))\frac{\check{\Pi}(s^t)}{\hat{\Pi}(s^t)}\frac{\hat{\Pi}(s^t)}{\Pi(s^t)}\Pi(s^t)$$

where $\Pi(s^t), \hat{\Pi}(s^t), \check{\Pi}(s^t)$ are joint densities over histories of states, and Π is a true or physical measure, $\hat{\Pi}$ is a measure induced by Bayes' Law, and $\check{\Pi}$ is a twisted measure induced by robust adjustments to either $\hat{\Pi}$ or just to Π (in settings without learning).

Wedgeology

What is called the New Keynesian IS curve simply rearranges

$$1 = \beta \sum_{s_{t+1}} \left(\frac{u'(c_{t+1})}{u'(c_t)} \right) R_{t+1}(s^{t+1}) \frac{\check{\pi}(s_{t+1}|s_t)}{\hat{\pi}(s_{t+1}|s_t)} \frac{\hat{\pi}(s_{t+1}|s_t)}{\pi(s_{t+1}|s_t)} \pi(s_{t+1}|s_t)$$

Wedge:

$$\frac{\check{\pi}(s_{t+1}|s_t)}{\hat{\pi}(s_{t+1}|s_t)} \frac{\hat{\pi}(s_{t+1}|s_t)}{\pi(s_{t+1}|s_t)}$$

IS curve wedgeology

- Shocks to the stochastic discount factor *are* shocks to the NK IS curve.
- Bayes' law leads to a backwards looking model of shocks.
- Robustness leads to a forward-looking model of shocks, due to the 'Murphy's law' feature of exponential twisting that involves *value functions*.
- Alternatives to habit persistence.
- Kreps-Porteus interpretation (e.g., see Tallarini (2000)).

Paraphrase of Bernanke, October 2007

When the day arrives that monetary policy makers come to know the probabilities that govern the risky outcomes they face, the Bayesian approach will tell us how to make policy. Then our policy prescriptions will reflect both the probability weights we attach to outcomes and the continuation values we associate with them. Of course, until Congress tells the FOMC to target inflation, continuation values will reflect the 'dual mandate' the Humphrey-Hawkins Act assigns.

Paraphrase of Bernanke, October 2007

In the mean time, it is worth thinking about other approaches to model uncertainty that, like robust control, require less confidence in the probabilities that our favorite model assigns. To acquire robustness across a set *of probability models, robust control adopts an instrumental pessimism that allows a decision maker to put* bounds *on performance by using a min-max expected utility approach. To construct these bounds, the decision maker contemplates worst-scenarios among a set of probability models restricted as the policy maker wishes. Sometimes a thoughtful policy maker who has serious model specification doubts will appear to act more aggressively than one who is sure that, probabilistically speaking, he knows the truth.*

Great speech because . . .

- Probabilities
- Utility functions
- Bayesian model averaging
- Model selection
- Robust control and multiple priors