Bayesian Methods for Macroeconometrics

Motivation

Frank Schorfheide
Department of Economics
University of Pennsylvania

A Toy Model

- Phillips curve:

$$z_t = \theta \pi_t + \epsilon_{s,t}$$

- Demand:

$$\pi_t = m_t + \epsilon_{d,t}$$

- Policy

$$m_t = \begin{cases} 0 & \text{for } t = 1, \ldots, T \\ -\epsilon_{d,t} + \delta \epsilon_{s,t} & \end{cases}$$

- Policy maker's loss function:

$$\widetilde{L}_t = (\pi_t^2 + z_t^2)$$

- Note: all variables are in deviations from steady state, $\epsilon_{d,t}$ and $\epsilon_{s,t}$ are $iid\mathcal{N}(0,1)$ shocks.

Questions

- Forecast output given money:

$$z_t = \theta m_t + \underbrace{\theta \epsilon_{d,t} + \epsilon_{s,t}}_{u_t} \quad \Longrightarrow \quad I\!\!E[z_t|m_t] = \theta m_t$$

- Find optimal policy (under new rule):

$$z_t = \theta\delta\epsilon_{s,t} + \epsilon_{s,t}$$

Hence,

$$\widetilde{L}_t = \left[(\theta\delta + 1)^2 + \delta^2 \right] \epsilon_{s,t}^2$$

Expected loss (average over supply shocks):

$$L(\theta, \delta) = [(\theta\delta + 1)^2 + \delta^2]$$

Remarks

- Both $\mathbb{E}[z_t|m_t]$ and $L(\theta, \delta)$ depend on θ but we don't know θ.

- We also don't know $\epsilon_{d,t}$ and $\epsilon_{s,t}$ for future t.

  Solution: use probability calculus.

- Bayesian approach: treat ϵ and θ the same: place (prior) probability distribution on θ.

- Use Bayes theorem to find conditional distribution of θ given data.

Likelihood Function

- Suppose we observe z_t and π_t for $t = 1, \ldots, T$ (recall that $m_t = 0$). Let $y_t = [z_t, \pi_t]'$, $Y^T = [y_1, \ldots, y_T]$.

- Density of y_t:

$$p(y_t|\theta) = p(z_t, \pi_t|\theta) = \underbrace{p(z_t|\pi_t, \theta)}_{\mathcal{N}(\theta\pi_t, 1)} * \underbrace{p(\pi_t)}_{\mathcal{N}(0,1)}$$

- Thus,

$$\begin{aligned} p(y_t|\theta) &= (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(z_t - \theta\pi_t)^2\right\} (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}\pi_t^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}(z_t - \theta\pi_t)^2\right\} \end{aligned}$$

- Likelihood is joint density of data given parameters:

$$\mathcal{L}(\theta|Y) = p(Y|\theta) = \prod_{t=1}^{T} p(y_t|Y^{t-1}, \theta) = \prod_{t=1}^{T} p(y_t|\theta)$$

since observations are *iid*.

Bayesian Updating (I)

- Suppose 2 policy advisors:

  1. Peter: $\theta = 1/10$.

  2. Dave: $\theta = 1$.

- Should Chairman Al listen to Peter or Dave?

- Chairman Al's priors:

  – Probability that Peter is right: $\mathbb{P}\{\theta = 1/10\} = 1/4$.

  – Probability that Dave is right: $\mathbb{P}\{\theta = 1\} = 3/4$.

- One pair of observations: inflation $\pi = 2\%$, output $z = 0.25\%$.

Bayesian Updating (II)

- Use Bayes's theorem:

$$
\begin{aligned}
I\!P\{\theta = 1/10|Y\} &= \frac{I\!P\{\theta = 1/10\}\mathcal{L}(\theta = 1/10|Y)}{I\!P\{\theta = 1/10\}\mathcal{L}(\theta = 1/10|Y) + I\!P\{\theta = 1\}\mathcal{L}(\theta = 1|Y)} \\
&= \frac{\frac{1}{4}\exp\left\{-\frac{1}{2}\left(\frac{1}{4} - \frac{2}{10}\right)^2\right\}}{\frac{1}{4}\exp\left\{-\frac{1}{2}\left(\frac{1}{4} - \frac{2}{10}\right)^2\right\} + \frac{3}{4}\exp\left\{-\frac{1}{2}\left(\frac{1}{4} - \frac{2}{2}\right)^2\right\}} \\
&= 0.61
\end{aligned}
$$

Forecasting (I)

- Predict z_2 given $m_2 = 1$ and Y.

- Conditional on θ:

$$z_2|m_2, Y, \theta \sim \mathcal{N}(\theta m_2, 1 + \theta^2)$$

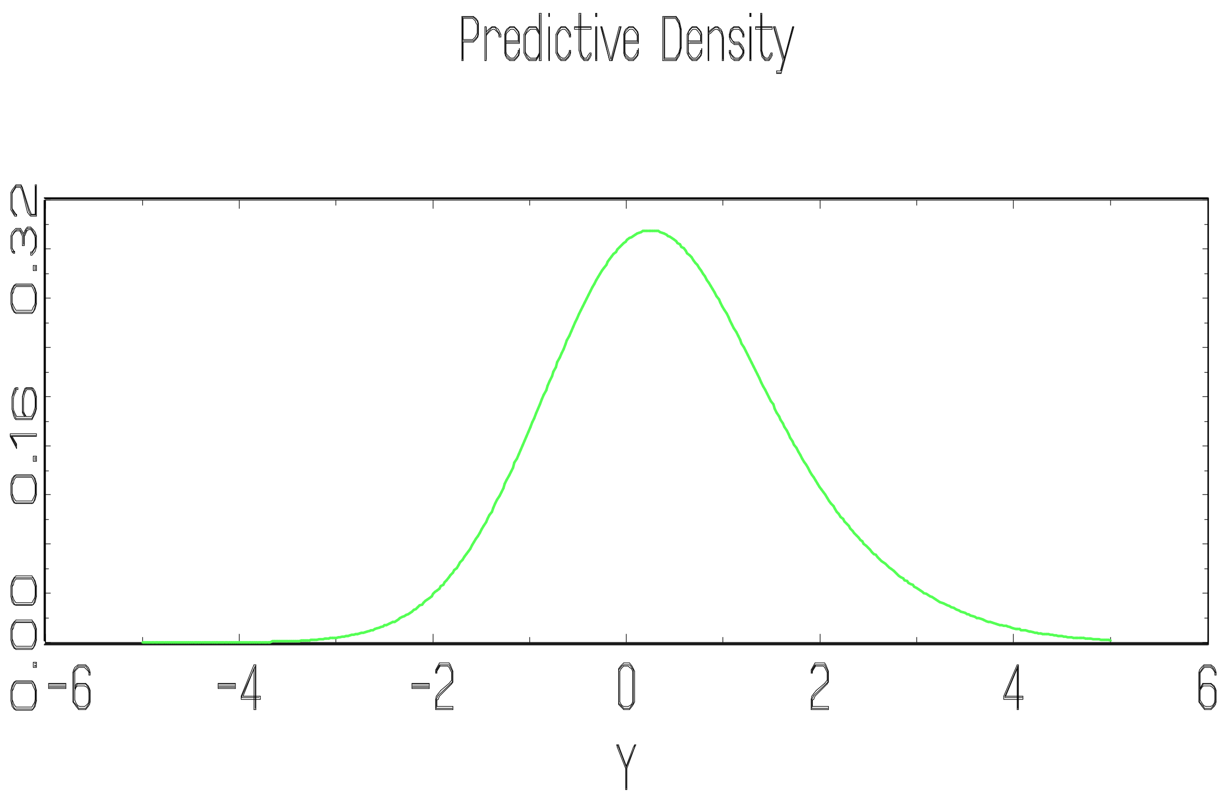
- Recall:

$$\begin{aligned} I\!P\{\theta = 1/10|Y\} &= 0.61 \\ I\!P\{\theta = 1|Y\} &= 0.39 \end{aligned}$$

- Thus:

$$p(z_2|m_2, Y) = 0.61 * \underbrace{p(z_2|m_2 = 1, Y, \theta = 1/10)}_{\mathcal{N}(0.1, 1.01)} + 0.39 * \underbrace{p(z_2|m_2 = 1, Y, \theta = 1)}_{\mathcal{N}(1, 2)}$$

(Figure)



Predictive Density
0.00
0.16
0.32
-6
-4
-2
0
2
4
6
Y

Forecasting (II)

- Summarize information in predictive density.
- Point forecast δ
- Define loss function $L(z,\delta)$:
  - quadratic error loss: $(z-\delta)^2$
  - absolute error loss: $|z-\delta|$
  - linex loss: $\exp\{a(z-\delta)\} - a(z-\delta) - 1$
- Find predictor that minimizes posterior expected loss:

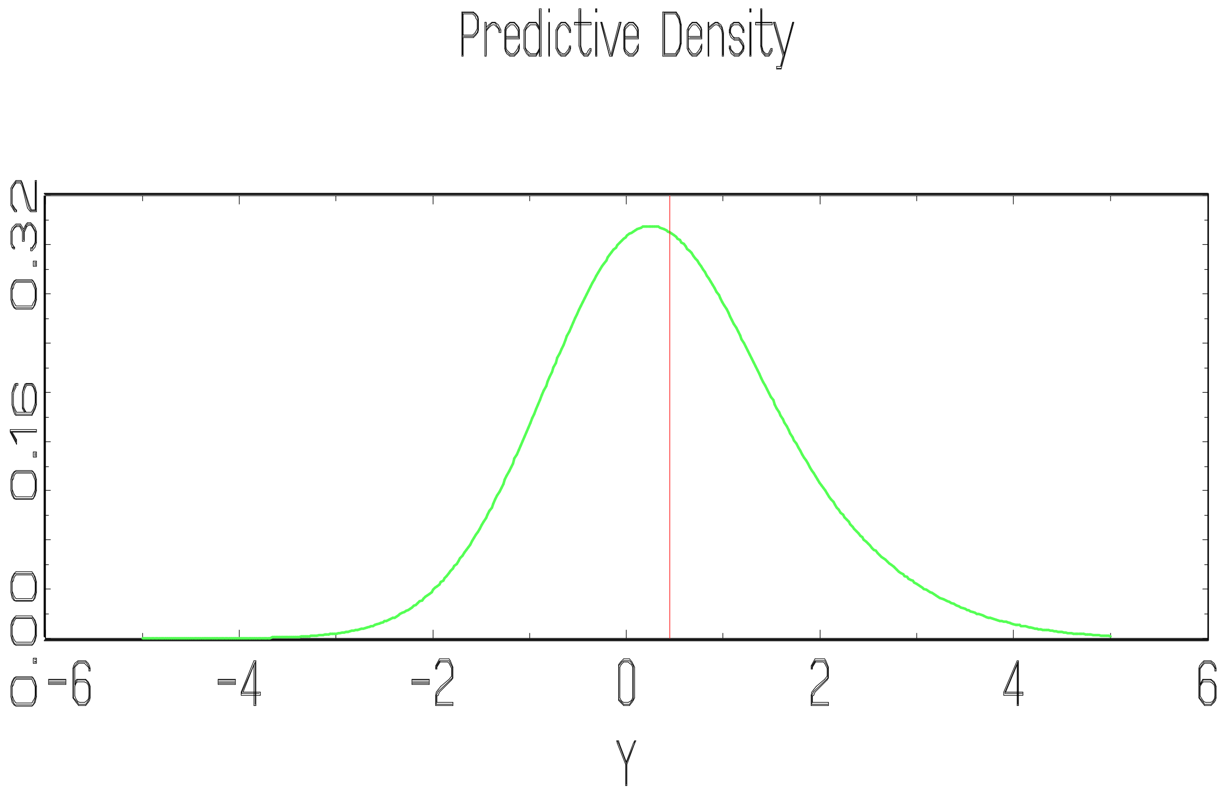
$$\delta^{Bayes} = \text{argmin} \int L(z,\delta)p(z|m,Y)dz$$

Forecasting (III)

- Example:
  - Loss function is quadratic.
  - Optimal predictor is posterior mean

$$\delta = I\!P\{\theta = 1/10 | Y\} * \frac{1}{10} + I\!P\{\theta = 1 | Y\} * 1 = 0.45$$

(Figure)


Predictive Density
0.00
0.16
0.32
-6
-4
-2
0
2
4
6
Y

Forecasting (IV)

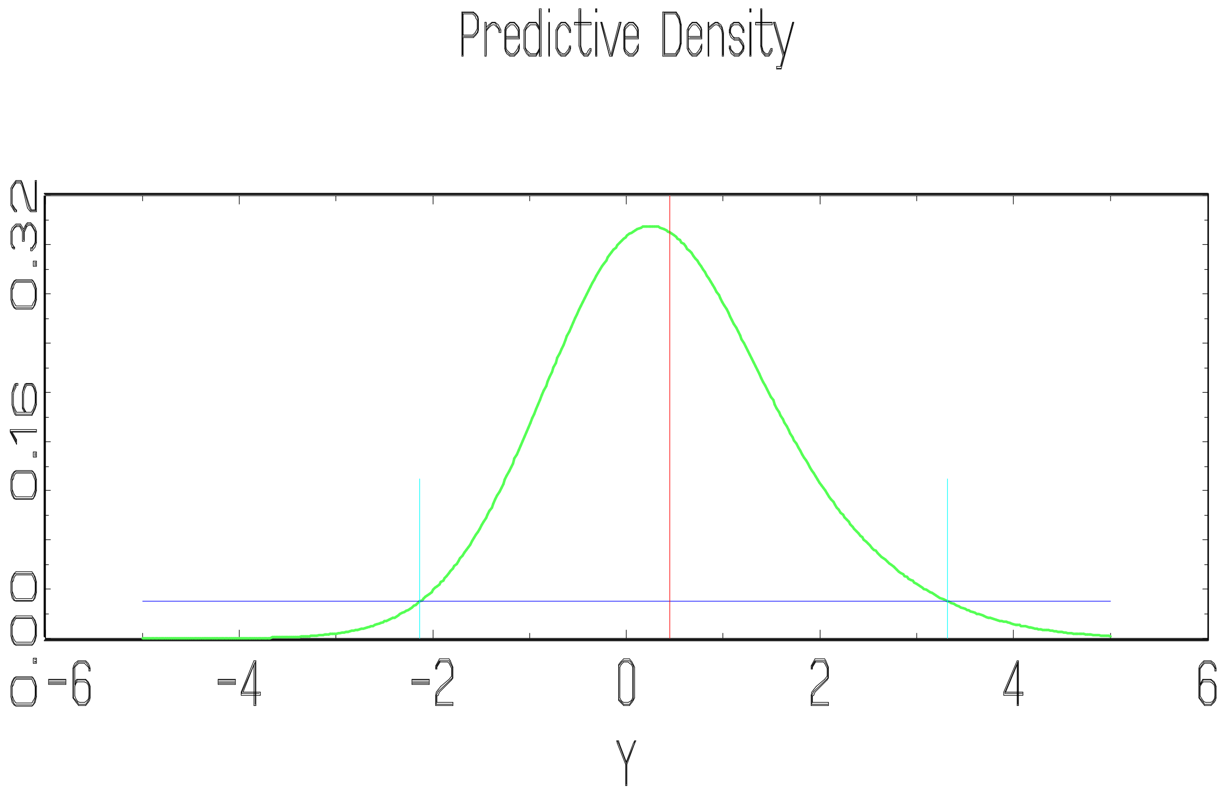
- Interval forecast: want the smallest set for z such that set is $100(1-\alpha)\%$ credible.

- Consider sets of the form:

$$\{z \in \mathbf{R} \mid p(z|m,Y) \geq k_\alpha\}$$

- Choose k_α such that set has desired posterior coverage probability.

(Figure)



Predictive Density
0.32
0.16
0.00
-6
-4
-2
0
2
4
6
Y

Optimal Policy (I)

- Suppose we knew θ...

- Then optimal policy is:

$$\begin{aligned}\delta^{opt}(\theta) &= \text{argmin}\left[\delta^2(1+\theta)^2 + 2\theta\delta + 1\right] \\ &= -\frac{\theta}{1+\theta^2}\end{aligned}$$

- Hence,

$$\begin{aligned}\text{Peter} &: \delta^{opt}(\theta = 1/10) = -0.10 \\ \text{Dave} &: \delta^{opt}(\theta = 1) = -0.50\end{aligned}$$

- Recall:

$$\begin{aligned}I\!P\{\theta = 1/10|Y\} &= 0.61 \\ I\!P\{\theta = 1|Y\} &= 0.39\end{aligned}$$

Optimal Policy – Attempt 1

- Chairman Al ranks advisors according posterior probabilities and listens to Peter, i.e.,

  $\delta = \delta^{opt}(\theta = 1/10) = -0.1.$

Optimal Policy – Attempt 2

- Chairman Al will either use $\delta^{opt}(\theta = 1/10)$ or $\delta^{opt}(\theta = 1)$.

- Expected losses (risks) are:

Decision	$\theta = 1/10$	$\theta = 1$	Posterior Risk
$\delta^{opt}(\theta = 1/10)$	0.99	0.82	0.92
$\delta^{opt}(\theta = 1)$	1.15	0.50	0.90

- Comparing the posterior risk of these two policies,

  Chairman Al chooses $\delta = \delta^{opt}(\theta = 1) = -0.5$.

Optimal Policy – Attempt 3

- Chairman Al minimizes posterior expected loss:

$$\mathbb{E}[L(\theta,\delta)|Y] = \delta^2(1+\mathbb{E}[\theta^2|Y]) + 2\delta\mathbb{E}[\theta|Y] + 1$$

- Solution:

$$\delta^{Bayes} = -\frac{\mathbb{E}[\theta|Y]}{1+\mathbb{E}[\theta^2|Y]} = -0.32$$

  Posterior Risk: 0.85

- Compare to posterior-mean plug-in solution:

$$\delta^{plug-in} = -\frac{\mathbb{E}[\theta|Y]}{1+(\mathbb{E}[\theta|Y])^2} = -0.38$$

Optimal Policy – Frequentist View

- Frequentist risk:

$$\mathbb{E}[L(\theta, \underbrace{\delta(Y)}_{r.v.})] \tag{1}$$

- We consider a class of decision rules $\delta(Y)$ and try to find the one with the best operating characteristics in repeated sampling.

- Choose $\delta(Y)$ among a class of functions to minimize

$$\mathbb{E}[L(\theta, \underbrace{\delta(Y)}_{r.v.})] = 1 + \mathbb{E}[\delta(Y)^2] + 2\theta\mathbb{E}[\delta(Y)] + \theta^2\mathbb{E}[\delta(Y)^2] \tag{2}$$

- We could estimate θ based on historical sample.

- Note: optimal decision is not necessarily of the MLE plug-in form:

$$\delta_{mle}(Y^T) = -\frac{\hat{\theta}_{mle}}{1 + \hat{\theta}^2_{mle}}. \tag{3}$$

Frequentist Decision Theory (I)

- What does a frequentist mean by "optimal"?

- The minimax risk associated with a loss function L is the value

$$\bar{R} = \inf_{\delta \in \mathcal{D}^*} \sup_{\theta \in \Theta} I\!\!E_\theta[L(\theta, \delta(Y))]$$

and a *minimax decision rule* is any (possibly randomized) decision rule δ_0 such that

$$\sup_{\theta \in \Theta} I\!\!E_\theta[L(\theta, \delta_0(Y))] = \bar{R}.$$

- A decision rule δ_0 is *inadmissible* if there exists a δ_1 which dominates δ_0, that is, for every θ

$$I\!\!E_\theta[L(\theta, \delta_0(Y))] \geq I\!\!E_\theta[L(\theta, \delta_1(Y))]$$

and, for at least one value θ_0

$$I\!\!E_\theta[L(\theta, \delta_0(Y))] > I\!\!E_\theta[L(\theta, \delta_1(Y))]$$

Frequentist Decision Theory (II)

- Suppose $Y \sim \mathcal{N}(\theta, \mathcal{I}_p)$ and the goal is to estimate θ. Estimator is evaluated according to

$$L(\delta, \theta) = \sum_{i=1}^{p} w_i(\delta_i - \theta_i)^2, \quad w_i > 0$$

- The OLS estimator $\delta_{ols} = Y$ is minimax.

- The OLS estimator $\delta_{ols} = Y$ is not admissible if $p \geq 3$ (Stein, 1955).

- If $w_i = 1$ for each i, the following (James and Stein, 1961) estimator uniformly dominates the OLS estimator:

$$\delta_{JS} = \left(1 - \frac{p-2}{\|Y\|}\right) Y.$$

Optimal Policy – continued

- Ranking of decision rules is not uniform over θ.
- We want good performance for values of θ that we regard as plausible.
- Frequentist risk is not very plausible for policy making: condition on θ (not observed) and average over trajectories that could have been observed but were not observed.