Bayesian Analysis of a Linear Regression Model

Frank Schorfheide
Department of Economics
University of Pennsylvania

Introduction to Bayesian Statistics

- Ingredients of Bayesian Analysis:
  - Likelihood function $\mathcal{L}(\theta|Y^T) = p(Y^T|\theta)$
  - Prior density $p(\theta)$
  - Marginal data density $p(Y^T) = \int p(Y^T|\theta)p(\theta)d\theta$
- Bayes Theorem:

$$p(\theta|Y^T) = \frac{\mathcal{L}(\theta|Y^T)p(\theta)}{p(Y^T)} \tag{1}$$

Linear Regression

- Consider linear regression model:

$$y_t = x_t'\theta + u_t, \quad u_t \sim iid\mathcal{N}(0,1), \tag{2}$$

  or

$$Y = X\theta + U.$$

  Assume θ is $k \times 1$.

- Notice: we treat the variance of the errors as know. The generalization to unknown variance is straightforward but tedious.

- Likelihood function:

$$\mathcal{L}(\theta|Y,X) = (2\pi)^{-T/2} \exp\left\{-\frac{1}{2}(Y - X\theta)'(Y - X\theta)\right\}. \tag{3}$$

A Convenient Prior

- Prior:

$$\theta \sim \mathcal{N}\left(0_{k\times 1}, \tau^2 \mathcal{I}_{k\times k}\right), \quad p(\theta) = (2\pi\tau^2)^{-k/2} \exp\left\{-\frac{1}{2\tau^2}\theta'\theta\right\} \tag{4}$$

- Large τ means diffuse prior.

- Small τ means tight prior.

Deriving the Posterior (I)

- Bayes Theorem:

$$p(\theta|Y,X) \propto p(\theta)\mathcal{L}(\theta|Y,X). \tag{5}$$

- Right-hand-side is given by

$$\begin{aligned} & p(\theta)\mathcal{L}(\theta|Y,X) \\ \propto \;& (2\pi)^{-\frac{T+k}{2}}\tau^{-k}\exp\left\{-\frac{1}{2}[Y'Y - \theta'X'Y - Y'X\theta + \theta'X'X\theta + \tau^{-2}\theta'\theta]\right\}. \end{aligned} \tag{6}$$

- Rewrite exponential term

$$\begin{aligned} & Y'Y - \theta'X'Y - Y'X\theta + \theta'X'X\theta + \tau^{-2}\theta'\theta \\ =\;& Y'Y - \theta'X'Y - Y'X\theta + \theta'(X'X + \tau^{-2}\mathcal{I})\theta \qquad (7) \\ =\;& \Big(\theta - (X'X + \tau^{-2}\mathcal{I})^{-1}X'Y\Big)'\Big(X'X + \tau^{-2}\mathcal{I}\Big)\Big(\theta - (X'X + \tau^{-2}\mathcal{I})^{-1}X'Y\Big) \\ & + Y'Y - Y'X(X'X + \tau^{-2}\mathcal{I})^{-1}X'Y. \end{aligned}$$

Deriving the Posterior (II)

- Exponential term is a quadratic function of θ.

- Deduce: posterior distribution of θ must be a multivariate normal distribution

$$\theta|Y, X \sim \mathcal{N}(\tilde{\theta}_T, \tilde{V}_T) \tag{8}$$

with

$$\begin{aligned} \tilde{\theta}_T &= (X'X + \tau^{-2}\mathcal{I})^{-1}X'Y \\ \tilde{V}_T &= (X'X + \tau^{-2}\mathcal{I})^{-1}. \end{aligned}$$

- $\tau \longrightarrow \infty$:

$$\theta|Y, X \overset{approx}{\sim} \mathcal{N}\left(\hat{\theta}_{mle}, (X'X)^{-1}\right). \tag{9}$$

- $\tau \longrightarrow 0$:

$$\theta|Y, X \overset{approx}{\sim} \text{Pointmass at } 0 \tag{10}$$

Point Estimation (I)

- Interpret point estimation as decision problem.

- Consider quadratic loss:

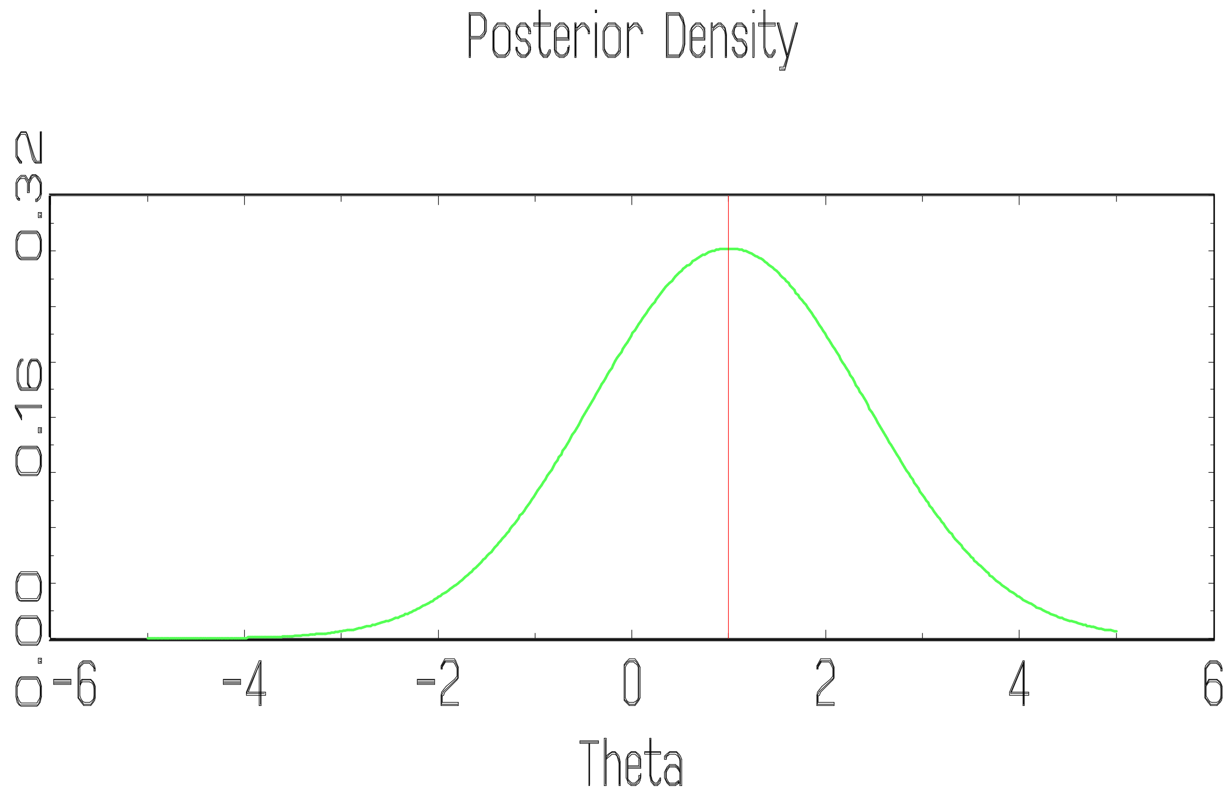
$$L(\theta, \delta) = (\theta - \delta)^2 \tag{11}$$

- Optimal decision rule is obtained by minimizing

$$\min_{\delta \in \mathcal{D}} \; \mathbb{E}[(\theta - \delta)^2 | Y, X] \tag{12}$$

- Solution: $\delta = \mathbb{E}[\theta | Y, X]$, i.e., posterior mean.

(Figure)


Posterior Density
0.32
0.16
0.00
-6
-4
-2
0
2
4
6
Theta

Point Estimation (II)

- Consistency: Suppose data are generated from the model $y_t = x_t'\theta_0 + u_t$. Asymptotically the Bayes estimator converges to the "true" parameter θ_0.

- Consider

$$\begin{aligned} \tilde{\theta}_T &= (X'X + \tau^{-2}\mathcal{I})^{-1}X'Y \\ &= \theta_0 + \left(\frac{1}{T}X'X + \frac{1}{\tau^2 T}\mathcal{I}\right)^{-1}\left(\frac{1}{T}X'U\right) \\ &\xrightarrow{p} \theta_0 \end{aligned} \tag{13}$$

- Disagreement between two Bayesians who have different priors will asymptotically vanish.

Confidence Sets

- Bayesian: $C_Y \subseteq \Theta$ is $1-\alpha$ credible if

$$\mathbb{P}_Y\{\underbrace{\theta}_{r.v.} \in C_Y\} \geq 1-\alpha \tag{14}$$

- Bayesian: a highest posterior density region (HPD) is of the form

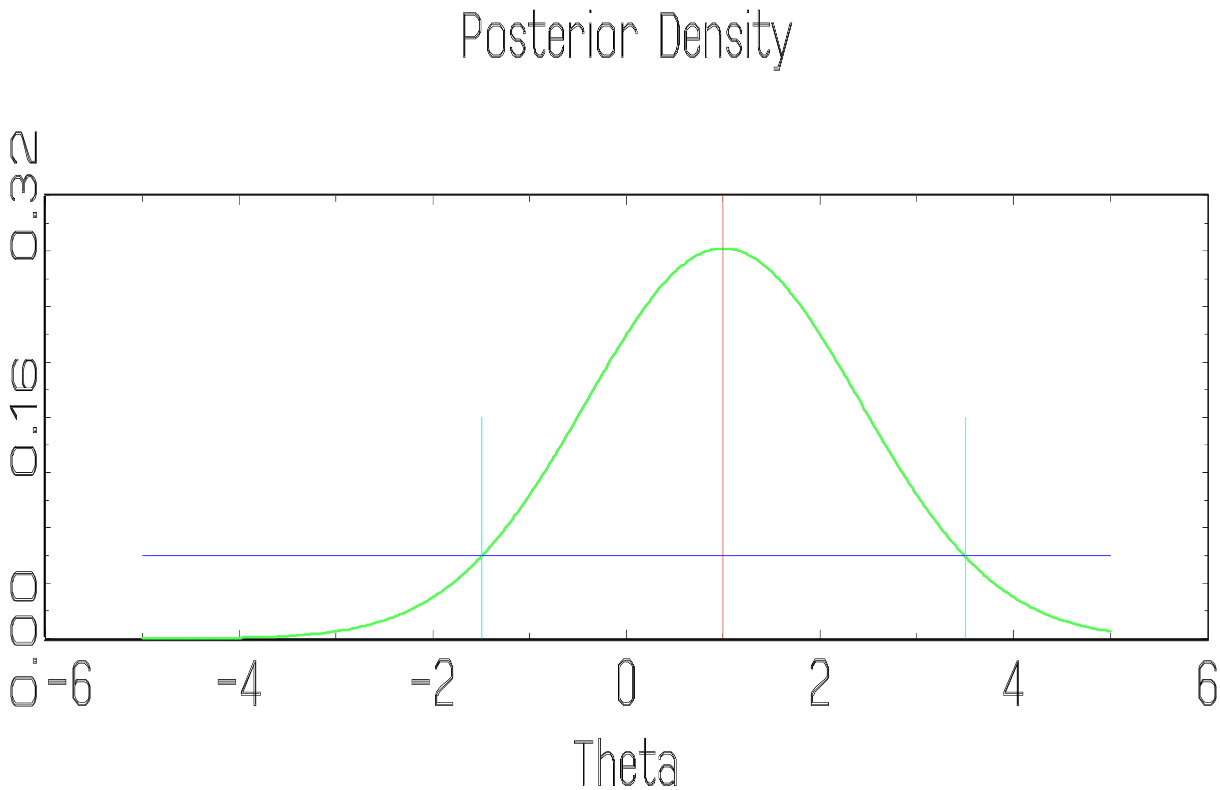
$$C_Y = \{\theta : p(\theta|Y) \geq k_\alpha\} \tag{15}$$

where k_α is the largest bound such that

$$\mathbb{P}_Y\{\theta \in C_Y\} \geq 1-\alpha$$

HPD regions have the smallest volume among all α credible regions of the parameter space Θ.

(Figure)


Posterior Density
0.32
0.16
0.00
-6
-4
-2
0
2
4
6
Theta

Testing (I)

- $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$.

- Decision space is 0 ("reject") and 1 ("accept").

- Loss function

$$L(\theta, \delta) = \begin{cases} 0 & \delta = \{\theta \in \Theta_0\} \quad \text{correct decision} \\ a_0 & \delta = 0,\ \theta \in \Theta_0 \quad \text{Type 1 error} \\ a_1 & \delta = 1,\ \theta \in \Theta_1 \quad \text{Type 2 error} \end{cases} \tag{16}$$

  Note that the parameters a_1 and a_2 are part of the econometricians preferences.

- Optimal decision:

$$\delta(Y^T) = \begin{cases} 1 & I\!P_Y\{\theta \in \Theta_0\} \geq \frac{a_1}{a_0+a_1} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Testing (II)

- Posterior odds:

$$\frac{I\!\!P_Y\{\theta \in \Theta_0\}}{I\!\!P_Y\{\theta \in \Theta_1\}}$$

- Often, hypotheses are evaluated according to Bayes factors:

$$B(Y^T) = \frac{\text{Posterior Odds}}{\text{Prior Odds}} = \frac{I\!\!P_Y\{\theta \in \Theta_0\}/I\!\!P_Y\{\theta \in \Theta_1\}}{I\!\!P\{\theta \in \Theta_0\}/I\!\!P\{\theta \in \Theta_1\}} \tag{18}$$

- How do we test $\theta = 0$? We'll come back to that later...