# Bayesian Methods for Macroeconometrics

**Frank Schorfheide**
Department of Economics, University of Pennsylvania

# Posterior Odds and Marginal Data Densities

- Posterior model probabilities can be computed as follows:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_j \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad j = 1, \ldots, 4, \tag{1}$$

- where

$$p(Y) = \int \mathcal{L}(\theta|Y) p(\theta) d\theta \tag{2}$$

- Posterior odds and Bayes Factor

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \underbrace{\frac{\pi_{1,T}}{\pi_{2,T}}}_{Prior\,Odds} \times \underbrace{\frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_2)}}_{Bayes\,Factor} \tag{3}$$

# Example: Linear Regression

- Simple example: compare

$$\mathcal{M}_0 \;:\; y_t = x_t^{(0)'}\theta^{(0)} + u_t^{(0)} \tag{4}$$

$$\mathcal{M}_1 \;:\; y_t = x_t^{(1)'}\theta^{(1)} + u_t^{(1)} \tag{5}$$

- Prior probabilities: $\pi_{i,0}$.

- Posterior probabilities:

$$\pi_{i,T} = \frac{\pi_{i,0}p(Y^T|\mathcal{M}_i)}{\sum_{i=0,1}\pi_{i,0}p(Y^T|\mathcal{M}_i)}. \tag{6}$$

  where marginal data density is

$$p(Y^T|\mathcal{M}_i) = \int \mathcal{L}(\theta^{(i)}|Y^T,\mathcal{M}_i)p(\theta^{(i)}|\mathcal{M}_i)d\theta^{(i)} \tag{7}$$

# Example: Linear Regression

- Here calculation is relatively simple:

$$p(Y|X) = \frac{\mathcal{L}(\theta|Y,X)p(\theta)}{p(\theta|Y,X)}. \tag{8}$$

- Since, we previously showed that the posterior $p(\theta|Y,X)$ is multivariate normal all the terms on the right-hand-side are known:

$$\begin{aligned} p(Y|X) &= \frac{(2\pi)^{-T/2}(2\pi)^{-k/2}\tau^{-k}\exp\left\{-\frac{1}{2}[(Y-X\theta)'(Y-X\theta)+\theta'\theta/\tau^2]\right\}}{(2\pi)^{-k/2}|\tilde{V}|^{-1/2}\exp\left\{-\frac{1}{2}[(\theta-\tilde{\theta})'\tilde{V}^{-1}(\theta-\tilde{\theta})]\right\}} \\ &= (2\pi)^{-T/2}\tau^{-k}|X'X+\tau^{-2}|^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}[Y'Y - Y'X(X'X+\tau^{-2}\mathcal{I})^{-1}X'Y]\right\}. \end{aligned} \tag{9}$$

using the definition of $\tilde{\theta}$ and $\tilde{V}$:

$$\tilde{\theta}_T = (X'X+\tau^{-2}\mathcal{I})^{-1}X'Y, \quad \tilde{V}_T = (X'X+\tau^{-2}\mathcal{I})^{-1}.$$

# Example: Linear Regression

- Schwarz Criterion: The terms of the marginal data density that asymptotically dominate are

$$\begin{aligned}\ln p(Y|X) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}(Y'Y - Y'X(X'X)^{-1}X'Y) - \frac{k}{2}\ln T + small \\ &= \underbrace{\ln p(Y|X,\hat{\theta}_{mle})}_{\text{max likelihood}} - \underbrace{\frac{k}{2}\ln T}_{\text{penalty}} + small \qquad (10)\end{aligned}$$

- Notice that

$$\begin{aligned}\ln|X'X+\tau^{-2}|^{-1/2} &= -\frac{1}{2}\ln\left|T\left(\frac{1}{T}X'X+\frac{1}{T\tau^2}\right)\right| \qquad (11) \\ &= \underbrace{-\frac{k}{2}\ln T}_{O(\ln(T))} \underbrace{-\frac{1}{2}\ln\left|\frac{1}{T}X'X+\frac{1}{T\tau^2}\right|}_{O_p(1)}\end{aligned}$$

# Consistency

- If data are generated from model $\mathcal{M}_i$ then as $T \longrightarrow \infty$ the posterior probability of model $\mathcal{M}_i$ converges to one for almost all data sets.

  If the models are nested, then the posterior prob of the smaller model will converge to one (because the penalty is smaller).

- If data are not generated from any of the models under consideration, then, roughly speaking, the posterior probability of the model that is closest in the Kullback-Leibler sense to the "truth" converges to one.

  What happens if there are ties? Is this a very useful result?

# Example: Linear Regression

- Suppose we compare: $\mathcal{M}_0$ $y_t = u_t$ versus $\mathcal{M}_1$ $y_t = x_t'\theta_0 + u_t$.

- Under $\mathcal{M}_0$:

$$\ln p(Y|X) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}Y'Y$$

whereas under $\mathcal{M}_1$:

$$\ln p(Y|X) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}(Y'Y - Y'X(X'X + \tau^{-2}\mathcal{I})^{-1}X'Y) - \frac{k}{2}\ln T + small$$

# Example: Linear Regression

- Assume that data were generated from the model $y_t = x_t'\theta_0 + u_t$.

$$\begin{aligned}
& Y'X(X'X+\tau^{-2})^{-1}X'Y \\
&= \theta_0'X'X(X'X+\tau^{-2})^{-1}X'X\theta_0 + U'X(X'X+\tau^{-2})^{-1}X'U \qquad (12)\\
&\quad +U'X(X'X+\tau^{-2})^{-1}X'X\theta_0 + \theta_0'X(X'X+\tau^{-2})^{-1}X'U \\
&= T\theta_0'\left(\frac{1}{T}\sum x_t x_t'\right)^{-1}\theta_0 + \sqrt{T}2\left(\frac{1}{\sqrt{T}}\sum x_t u_t\right)'\theta_0 \\
&\quad +\left(\frac{1}{\sqrt{T}}\sum x_t u_t\right)'\left(\frac{1}{T}\sum x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum x_t u_t\right) + O_p(1).
\end{aligned}$$

- If $\theta_0 = 0$ then

$$\ln\left[\frac{I\!P_{Y^T}\{\theta=0\}}{I\!P_{Y^T}\{\theta\neq 0\}}\right] = \frac{k}{2}\ln T + small \longrightarrow +\infty. \qquad (13)$$

- If $\theta_0 \neq 0$ then

$$\ln\left[\frac{I\!P_{Y^T}\{\theta=0\}}{I\!P_{Y^T}\{\theta\neq 0\}}\right] = -\frac{T}{2}\theta_0'\left(\frac{1}{T}\sum x_t x_t'\right)^{-1}\theta_0 + small \longrightarrow -\infty. \qquad (14)$$

# Finite-Sample Challenges: Lindley's Paradox

- Test $H_0 : \theta = 0$:

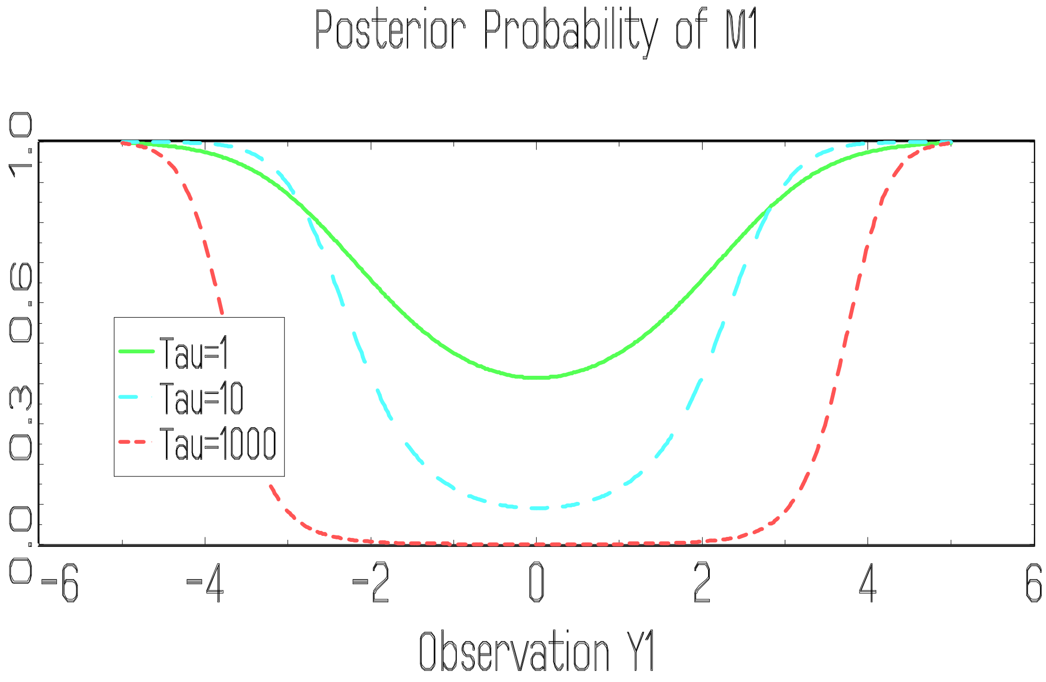
$$\frac{I\!\!P_Y\{\theta = 0\}}{I\!\!P_Y\{\theta \neq 0\}} = \tau^k |X'X + \tau^{-2}|^{1/2} \exp\left\{ -\frac{1}{2}[Y'X(X'X + \tau^{-2}\mathcal{I})^{-1}X'Y] \right\} \tag{15}$$

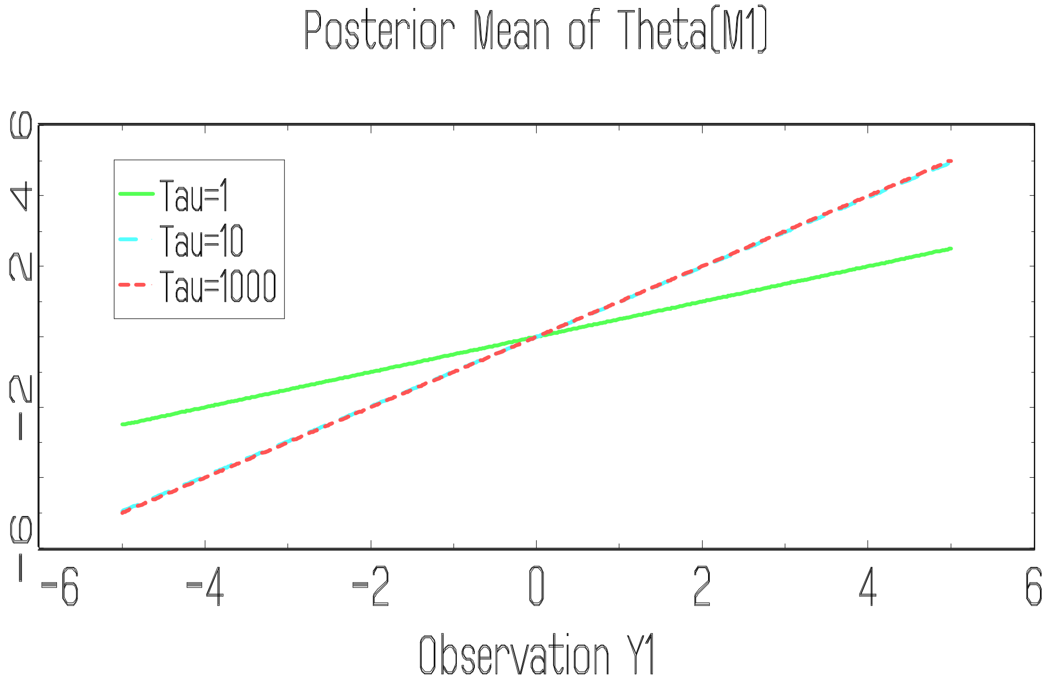
- Lindley's Paradox: suppose $Y$ is fixed, as prior on alternative becomes more diffuse ($\tau \longrightarrow \infty$)

$$\frac{I\!\!P_Y\{\theta = 0\}}{I\!\!P_Y\{\theta \neq 0\}} \longrightarrow \infty \tag{16}$$

  regardless of $Y$.

- Important for non-nested model comparisons. Changing prior variance can have small effect on posterior distribution of parameters yet large effect on posterior model probabilities.


Posterior Probability of M1


Posterior Mean of Theta(M1)

# Example: Model Selection vs Likelihood Ratio Test

- Comparison of Bayesian test to LR test.

- In the regression example: $(H_0 : \theta = 0)$

$$\begin{aligned} LR &= 2 \ln \left[ \frac{p(Y|X, \hat{\theta}_{mle})}{p(Y|X, \theta = 0)} \right] && (17) \\ &= Y'X(X'X)^{-1}X'Y. && (18) \end{aligned}$$

- If $H_0$ is true, then

$$LR = \left( \frac{1}{\sqrt{T}} \sum x_t u_t \right)' \left( \frac{1}{T} \sum x_t x_t' \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum x_t u_t \right) \Longrightarrow \chi^2_k \qquad (19)$$

# Example: Model Selection vs. Likelihood Ratio Tests

- Frequentist decision rule: accept / don't reject $\theta = 0$ if

$$Y'X(X'X)^{-1}X'Y < \chi^2_{k,crit} \tag{20}$$

- Bayesian decision rule: accept $\theta = 0$ if

$$Y'X(X'X)^{-1}X'Y < k \ln T + small \tag{21}$$

- Note: the implied Bayesian critical value tends to infinity at logarithmic rate. Consequently, the size of the test converges to zero asymptotically and the Type 1 error vanishes.

# In General...

- Laplace approximation ($\tilde{\theta}$ is mode, $\tilde{\Sigma}$ is inv Hessian)

$$\hat{p}(Y|\mathcal{M}) = (2\pi)^{d/2} \underbrace{p(Y|\tilde{\theta},\mathcal{M})p(\tilde{\theta}|\mathcal{M})}_{In-sample\,Fit} \times \underbrace{|\tilde{\Sigma}|^{1/2}}_{Dimensionality\,Penalty} \tag{22}$$

- In "regular" models $T \cdot \tilde{\Sigma} = O_p(1)$. Thus,

$$\frac{1}{2}\ln|\tilde{\Sigma}| = -\frac{d}{2}\ln T \tag{23}$$

  The larger the dimensionality $d$, the larger the penalty.

- $\ln p(Y|\mathcal{M})$ can be interpreted as predictive score (Good, 1952)

$$\sum_{t=1}^{T} \ln p(y_t|Y^{t-1},\mathcal{M}) = \sum_{t=1}^{T} \ln\left[\int p(y_t|Y^{t-1},\theta,\mathcal{M})p(\theta|Y^{t-1},\mathcal{M})d\theta\right], \tag{24}$$

  Model comparison based on posterior odds captures the relative one-step-ahead predictive performance.

# Numerical Approximations

- Naive approach: draw $\theta^{(s)}$ from prior $p(\theta)$ and use

$$\ln p(Y) \approx \frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} p(Y|\theta^{(s)})$$

- Why does this not work?

- Refinement:

$$\ln p(Y) = \prod_{t=1}^{T} p(y_t|Y^{t-1}) \approx \prod_{t=1}^{T} \frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} p(y_t|Y^{t-1}, \theta_{t-1}^{(s)}),$$

  where $\theta_{t-1}^{(s)})$ is drawn from $p(\theta|Y^{t-1})$.

- We will now discuss Geweke's modified harmonic mean estimator and the Chib and Jeliazkov method.

# Numerical Approximations: Harmonic Mean

- Harmonic mean estimators are based on the following identity

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{\mathcal{L}(\theta|Y)p(\theta)} p(\theta|Y)d\theta, \tag{25}$$

where $\int f(\theta)d\theta = 1$.

- Conditional on the choice of $f(\theta)$ an obvious estimator is

$$\hat{p}_G(Y) = \left[ \frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f(\theta^{(s)})}{\mathcal{L}(\theta^{(s)}|Y)p(\theta^{(s)})} \right]^{-1}, \tag{26}$$

where $\theta^{(s)}$ is drawn from the posterior $p(\theta|Y)$.

- Geweke (1999):

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta - \bar{\theta})'V_\theta^{-1}(\theta - \bar{\theta})\right] \\ &\quad \times \left\{ (\theta - \bar{\theta})'V_\theta^{-1}(\theta - \bar{\theta}) \le F_{\chi_d^2}^{-1}(\tau) \right\}. \end{aligned} \tag{27}$$

# Numerical Approximations: Chib and Jeliazkov (I)

- Rewrite Bayes Theorem:

$$p(Y) = \frac{\mathcal{L}(\theta|Y)p(\theta)}{p(\theta|Y)}. \tag{28}$$

- Thus,

$$\hat{p}_{CS}(Y) = \frac{\mathcal{L}(\tilde{\theta}|Y)p(\tilde{\theta})}{\hat{p}(\tilde{\theta}|Y)}, \tag{29}$$

  where we replaced the generic $\theta$ in (28) by the posterior mode $\tilde{\theta}$.

- Within the RWM Algorithm denote the probability of moving from $\theta$ to $\vartheta$ by

$$\alpha(\theta, \vartheta|Y) = \min \; \{1, r(\theta, \vartheta|Y)\}, \tag{30}$$

  where $r(\theta, \vartheta|Y)$ was in the description of the algorithm. Moreover, let $q(\theta, \tilde{\theta}|Y)$ be the proposal density for the transition from $\theta$ to $\tilde{\theta}$.

# Numerical Approximations: Chib and Jeliazkov (II)

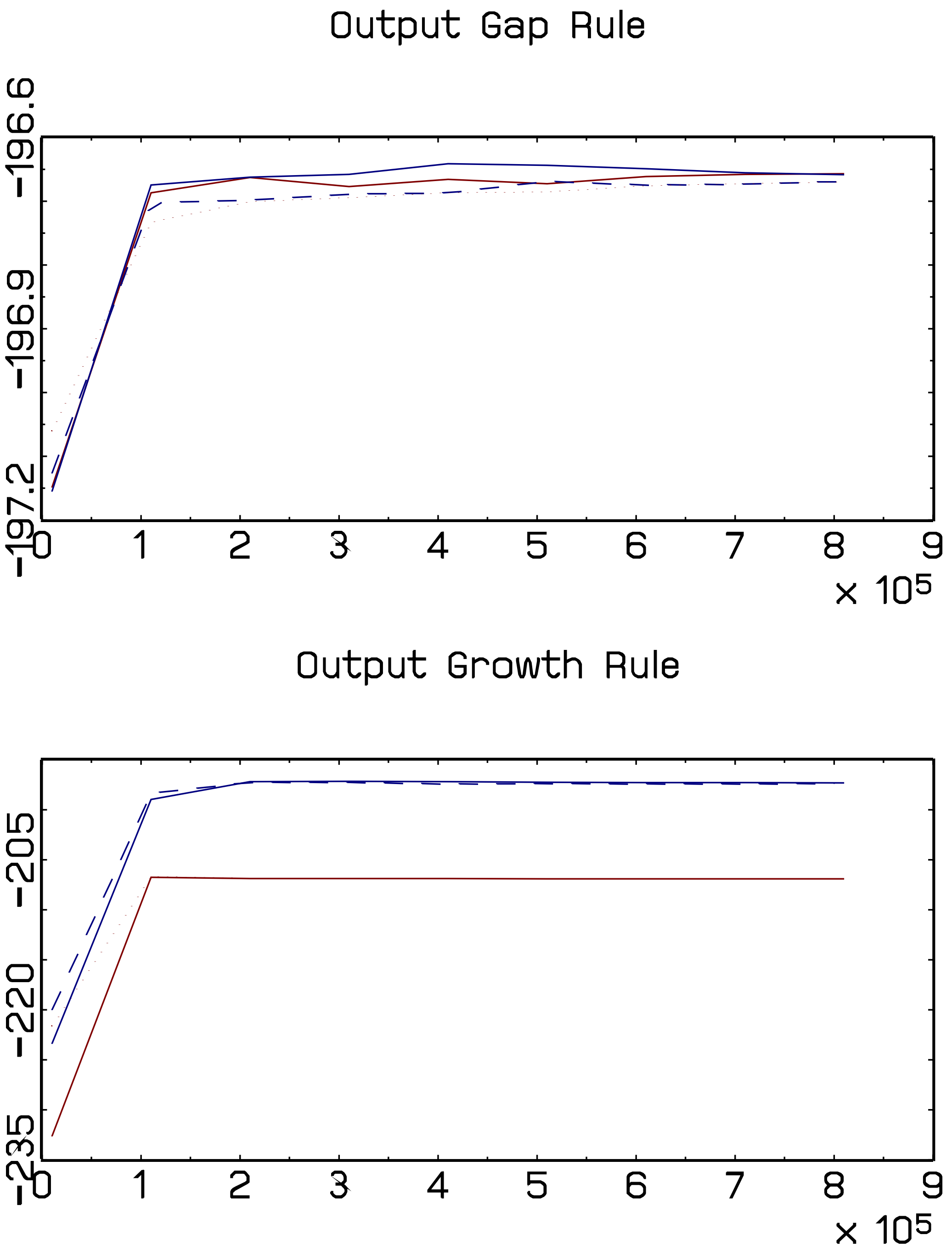
- Then the posterior density at the mode can be approximated as follows

$$\hat{p}(\tilde{\theta}|Y) = \frac{\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \alpha(\theta^{(s)}, \tilde{\theta}|Y) q(\theta^{(s)}, \tilde{\theta}|Y)}{J^{-1} \sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{(j)}|Y)}, \tag{31}$$

  where $\{\theta^{(s)}\}$ are sampled draws from the posterior distribution with the RWM Algorithm and $\{\theta^{(j)}\}$ are draws from $q(\tilde{\theta}, \theta|Y)$ given the fixed posterior mode value $\tilde{\theta}$.

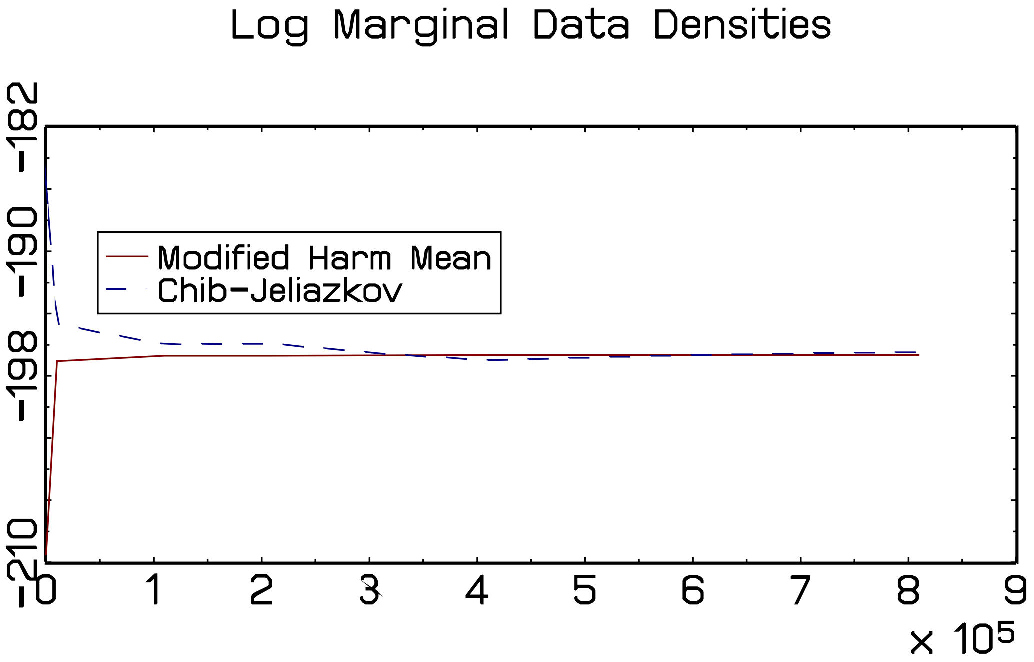
(insert figures here)

Figure 9: Log Marginal Data Densities from Multiple Chains



*Notes:* Output gap rule specification (top panel) and output growth rule specification (bottom panel), Data Sets 1-$\mathcal{M}_1$ and 1-$\mathcal{M}_2$, respectively. For each Markov chain, log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator and plotted as a function of the number of draws.

Figure 10: Log Marginal Data Densities – Geweke vs. Chib-Jeliazkov



*Notes:* Output gap rule specification $\mathcal{M}_1$, Data Set 1-$\mathcal{M}_1$. Log marginal data densities are computed recursively with Geweke's modified harmonic mean estimator as well as the Chib-Jeliazkov estimator and plotted as a function of the number of draws.

# Bayes Factors / Posterior Odds: Example

- Two alternative specifications: $\mathcal{M}_3$ prices are nearly flexible ($\kappa = 5$); $\mathcal{M}_4$ central bank does not respond to output ($\psi_2 = 0$).

- Marginal data densities are -196.7 for $\mathcal{M}_1$, -245.6 for $\mathcal{M}_3$, and -201.9 for $\mathcal{M}_4$.

- Bayes factors:

  - $\mathcal{M}_1$ versus $\mathcal{M}_3$ is approximately $e^{49}$;

  - $\mathcal{M}_1$ versus $\mathcal{M}_4$ is approximately $e^{4}$.