DSGE Model Evaluation

Frank Schorfheide
University of Pennsylvania
Gerzensee Ph.D. Course on Bayesian Macroeconometrics

May 28, 2019
Two Questions

• Question 1: Does model $M_1$ fit better than model $M_2$?
  $\Rightarrow$ Posterior model odds.

• Question 2: Are there patterns in the data that are inconsistent with model $M_1$?
  $\Rightarrow$ Posterior predictive checks.
Based on Chang, Doh, and Schorfheide (JMCB 2007): “Non-stationary Hours in a DSGE Model”

Many researchers doubt that hours worked are stationary as we have observed apparent changes in labor-supply patterns over recent decades.

We present a modified stochastic growth model in which hours worked have a stochastic trend, generated by a non-stationary labor supply shock.

Based on output and hours data we evaluate the stochastic growth models.
We consider four versions of the stochastic growth model:

- In $M_0$ and $M_1$ firms can choose the employment level at the given wage rate without any adjustment cost.
- In $A_0$ and $A_1$, on the other hand, it is costly for firms to adjust the employment level.
- In $A_0$ and $M_0$ the labor supply shock is a stationary AR(1) process, whereas it is modeled as random walk in $A_1$ and $M_1$. 
Households

- The representative household maximizes the expected discounted lifetime utility from consumption $C_t$ and hours worked $H_t$:

$$E_t \left[ \sum_{s=0}^{\infty} \beta^{t+s} \left( \ln C_{t+s} - \frac{(H_{t+s}/B_{t+s})^{1+1/\nu}}{1 + 1/\nu} \right) \right].$$

- The log utility in consumption implies a constant long-run labor supply in response to a permanent change in technology. The short-run (Frisch) labor supply elasticity is $\nu$. The labor supply shock is denoted by $B_t$.

- Per-period budget constraint faced by the household is

$$C_t + K_{t+1} - (1 - \delta)K_t = W_t H_t + R_t K_t.$$
• Firms rent capital, hire labor services, and produce final goods according to the following Cobb-Douglas technology:

\[ Y_t = (A_t H_t)^\alpha K_t^{1-\alpha} \left(1 - \varphi \cdot \left(\frac{H_t}{H_{t-1}} - 1\right)^2\right). \]  

(3)

• The stochastic process \( A_t \) represents the exogenous labor augmenting technical progress. The last term captures the cost of adjusting labor inputs: \( \varphi \geq 0 \).

• The firms maximize expected discounted future profits

\[ \mathbb{E}_t \left[ \sum_{s=0}^{\infty} \beta^{t+s} \lambda_{t+s} (Y_t - W_t H_t^d - R_t K_t^d) \right], \]  

(4)

where \( \lambda_t \) is the marginal value of a unit consumption to a household, which is treated as exogenous to the firm.
In equilibrium \( \lambda_t = 1/C_t \) and the goods, labor, and capital markets clear:

\[
Y_t = C_t + K_{t+1} - (1 - \delta)K_t, \quad H^d_t = H_t, \quad \text{and} \quad K^d_t = K_t.
\]

We assume that the log production technology evolves according to a random walk with drift:

\[
\ln A_t = \gamma + \ln A_{t-1} + \epsilon_{a,t}, \quad \epsilon_{a,t} \sim iid \mathcal{N}(0, \sigma_a^2).
\]

The level of technology in period 0 is denoted by \( A_0 \).
• In models $\mathcal{M}_0$ and $\mathcal{A}_0$, the labor supply shock follows a stationary AR(1) process:

$$\mathcal{M}_0: \quad \ln B_t = \rho_b \ln B_{t-1} + (1 - \rho_b) \ln B_0 + \epsilon_{b,t}, \quad \epsilon_{b,t} \sim iid \mathcal{N}(0, \sigma_{\epsilon_b}^2), \quad (6)$$

where $0 \leq \rho_b < 1$ and $\ln B_0$ is the unconditional mean of $\ln B_t$.

• In model $\mathcal{M}_1$ and $\mathcal{A}_1$ the innovation $\epsilon_{b,t}$ only has a transitory effect. Alternatively, in models $\mathcal{M}_1$ and $\mathcal{A}_1$ the labor supply shock evolves according to a random walk:

$$\mathcal{M}_1: \quad \ln B_t = \ln B_{t-1} + \epsilon_{b,t}, \quad \epsilon_{b,t} \sim iid \mathcal{N}(0, \sigma_{\epsilon_b}^2) \quad (7)$$

and we use $\ln B_0$ to denote the initial level of $\ln B_t$. 
• It is well known that in models $M_0$ and $A_0$ hours are stationary and that output, consumption, and capital grow according to the technology process $A_t$. Hence, one can induce stationarity with the following transformation:

$$M_0: \quad \tilde{Y}_t = \frac{Y_t}{A_t}, \quad \tilde{C}_t = \frac{C_t}{A_t}, \quad \tilde{K}_{t+1} = \frac{K_{t+1}}{A_t}.$$

• In models $M_1$ and $A_1$, on the other hand, the labor supply shock $B_t$ induces a stochastic trend into hours as well as output, consumption, and capital. To obtain a stationary equilibrium these variables have to be detrended according to:

$$M_1: \quad \tilde{H}_t = \frac{H_t}{B_t}, \quad \tilde{Y}_t = \frac{Y_t}{A_t B_t}, \quad \tilde{C}_t = \frac{C_t}{A_t B_t}, \quad \tilde{K}_{t+1} = \frac{K_{t+1}}{A_t B_t}.$$
• We fit the DSGE models to observations on the log level of real per capita output and hours worked, denoted by the $2 \times 1$ vector $y_t$.

• Let $\epsilon_t = [\epsilon_{a,t}, \epsilon_{b,t}]'$ and define the vector of structural model parameters as $\theta = [\alpha, \beta, \gamma, \delta, \nu, \ln A_0, \ln B_0, \rho_b, \sigma_a, \sigma_b]'$.

• It is well known that log-linearized DSGE models have a state space representation:

$$y_t = \Gamma_0 + \Gamma_1 s_{1,t} + \Gamma_2 s_{2,t} + \Gamma_3 t$$

(8)

$$s_{1,t} = \Phi_1 s_{1,t-1} + \Psi_1 \epsilon_t$$

(9)

$$s_{2,t} = s_{2,t-1} + \Psi_2 \epsilon_t.$$  

(10)
• The trend in (8) captures the effect of the drift in the random walk technology process $A_t$.
• Equation (9) represents the law of motion for the state variables of the detrended model, and (10) describes the evolution of trends: $s_{2,t} = \ln A_t - \gamma t$ in models $\mathcal{M}_0$ and $\mathcal{A}_0$ and $s_{2,t} = [\ln A_t - \gamma t, \ln B_t]'$ in $\mathcal{M}_1$ and $\mathcal{A}_1$.
• The Kalman filter can be used to compute the likelihood function $p(Y|\theta)$ for the state space system (8) - (10).
• To initialize the Kalman filter a distribution for the state vector in period $t = 0$ has to be specified.

• We factorize the initial distribution as $p(s_{1,0})p(s_{2,0})$ and set the first component equal to the unconditional distribution of $s_{1,t}$, whereas the second component, composed of the distribution of $\ln A_0$ (for $M_0, A_0$) and $[\ln A_0, \ln B_0]'$ (for $M_1, A_1$), respectively, is absorbed into the specification of our prior $p(\theta)$. 
• Paper uses three different data sets comprised of quarterly U.S. real per capita GDP and hours worked from 1954:Q2 to 2001:Q4.

• For Data Set 1 we use real GDP from the DRI-Global Insight database (GDPQ) and divide it by population of age 20 or older (PM20+PF20). Hours worked is measured as average weekly hours of all people in the non-farm business sector compiled by the Bureau of Labor Statistics (EEU005000005). We multiply the hours series by the employment ratio, which is the number of people employed (LHEM, DRI-Global Insight) divided by population (PM20+PF20).

• The observations from 1954:Q2 to 1958:Q4 are treated as pre-sample to quantify prior distributions.
• We assume all parameters to be *a priori* independent.
• By and large, the prior means are chosen based on a pre-sample of observations from 1954:Q2 to 1958:Q4.
• The prior mean of the labor share $\alpha$ is 0.66 and that for the quarter-to-quarter growth rate of productivity, $\gamma$, is 0.5%.
• The prior for $\beta$ is centered at 0.995. Combined with the prior mean of $\gamma$, this corresponds to an annualized real return of about 4%.
• The depreciation rate $\delta$ lies between 1.8% and 3.3% per quarter.
• The 90% probability interval for the Frisch labor supply elasticity $\nu$ ranges from 0.3 to 1.8.
• We specify a prior for the adjustment cost parameter $\varphi$ as follows.
  • In order to recruit labor $\Delta H$, firms can either search for workers, incurring adjustment costs $\varphi \left( \frac{\Delta H}{H} \right)^2 Y$, or pay head hunters for finding workers.
  • In the latter case the head hunters service fee is $\zeta W \Delta H$ where $\zeta$ is the fraction of the salary of the job to fill.
  • It is known that the head hunters tend to charge about 1/3 to 2/3 of quarterly earnings of a worker (i.e., $\zeta = 1/3$ to 2/3).
  • At the margin, the recruiting costs should be the same: $\varphi \left( \frac{\Delta H}{H} \right)^2 Y = \zeta W \Delta H$.
  • With the labor share of 1/3 ($= \frac{W}{Y}$) for a size of one percent increase of employment, $\frac{\Delta H}{H} = 1\%$, we obtain a range of 22 to 44 for $\varphi$.
  • We use a fairly diffuse prior distribution that is centered at 33 and has a standard deviation of 15.
Prior

- The presence of adjustment costs dampens the effect of technology and labor supply shocks on output and hours worked.
- In order to guarantee that the adjustment cost specifications have a priori similar implications for the volatility of the endogenous variable as $\mathcal{M}_0$ and $\mathcal{M}_1$ we use slightly different priors for the standard deviations of the structural shocks.
- Under $\mathcal{M}_0$ and $\mathcal{M}_1$ the priors for $\sigma_a$ and $\sigma_b$ are centered at 0.010, whereas under $\mathcal{A}_0$ and $\mathcal{A}_1$ they are centered at 0.015.
• For $M_0$ and $A_0$ the prior mean of $\ln B_0$ is constructed by matching average hours worked over the pre-sample period with the steady state level of hours worked $H^*$, evaluated at the prior mean values of the remaining structural parameters.

• For $M_1$ and $A_1$ the prior mean of $\ln B_0$ is obtained by equating hours worked in 1958:Q4 with the steady state level $B_0 H^*$. Similarly, we select the prior mean of $\ln A_0$ by matching $A_0 Y^*$ and $A_0 B_0 Y^*$, respectively, with the level of output in 1958:Q4.

• The prior standard deviations for $\ln A_0$ and $\ln B_0$ are 0.2. Finally, for $M_0$ and $A_0$ the 90% probability interval for the autoregressive parameter $\rho_b$ ranges from 0.825 to 0.977, implying a fairly persistent labor supply process.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Density</th>
<th>Data Set</th>
<th>Model</th>
<th>Para (1)</th>
<th>Para (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Beta</td>
<td>all</td>
<td>all</td>
<td>0.660</td>
<td>0.020</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Beta</td>
<td>all</td>
<td>all</td>
<td>0.995</td>
<td>0.002</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Normal</td>
<td>all</td>
<td>all</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Beta</td>
<td>all</td>
<td>all</td>
<td>0.025</td>
<td>0.005</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Gamma</td>
<td>all</td>
<td>all</td>
<td>1.000</td>
<td>0.500</td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>Beta</td>
<td>all</td>
<td>$M_0, A_0$</td>
<td>0.900</td>
<td>0.050</td>
</tr>
<tr>
<td>$\sigma_a$</td>
<td>InvGamma</td>
<td>all</td>
<td>$M_0, M_1$</td>
<td>0.010</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>$M_1, A_1$</td>
<td>0.015</td>
<td>1.000</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>InvGamma</td>
<td>all</td>
<td>$M_0, M_1$</td>
<td>0.010</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>$M_1, A_1$</td>
<td>0.015</td>
<td>1.000</td>
</tr>
<tr>
<td>$\ln A_0$</td>
<td>Normal</td>
<td>1</td>
<td>$M_0, A_0$</td>
<td>5.647</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>$M_1, A_1$</td>
<td>5.674</td>
<td>0.200</td>
</tr>
<tr>
<td>$\ln B_0$</td>
<td>Normal</td>
<td>1</td>
<td>$M_0, A_0$</td>
<td>3.236</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>$M_1, A_1$</td>
<td>3.209</td>
<td>0.200</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>Gamma</td>
<td>all</td>
<td>$A_0, A_1$</td>
<td>33.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Parameter</td>
<td>Domain</td>
<td>Density</td>
<td>Data Set</td>
<td>Model</td>
<td>Para (1)</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In $B_0$</td>
<td>$\mathbb{R}$</td>
<td>Normal</td>
<td>1</td>
<td>$M_1$</td>
<td>3.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>$M_1$</td>
<td>6.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>$M_1$</td>
<td>6.309</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In $B_0$</td>
<td>$\mathbb{R}$</td>
<td>Normal</td>
<td>1</td>
<td>$M_1$</td>
<td>3.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>$M_1$</td>
<td>6.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>$M_1$</td>
<td>6.309</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>[0, 1)</td>
<td>Beta</td>
<td>all</td>
<td>$M_0$</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_b$</td>
<td>[0, 1)</td>
<td>Beta</td>
<td>all</td>
<td>$M_0$</td>
<td>0.800</td>
</tr>
</tbody>
</table>
Recall: Posterior Odds and Marginal Data Densities

• Posterior model probabilities can be computed as follows:

\[
\pi_{i,T} = \frac{\pi_{i,0} p(Y|M_i)}{\sum_j \pi_{j,0} p(Y|M_j)}, \quad j = 1, \ldots, 2,
\] (11)

• where

\[
p(Y|M) = \int p(Y|\theta, M)p(\theta|M)d\theta
\] (12)

• Note:

\[
\ln p(Y_{1:T}|M) = \sum_{t=1}^{T} \ln \int p(y_t|\theta, Y_{1:t-1}, M)p(\theta|Y_{1:t-1}, M)d\theta
\]

• Posterior odds and Bayes Factor

\[
\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}} \times \frac{p(Y|M_1)}{p(Y|M_2)}
\] (13)
Computation of Marginal Data Densities

- Importance sampling
- Reciprocal importance sampling: Geweke’s modified harmonic mean estimator.
- Chib and Jeliazkov’s estimator

For a survey, see Ardia, Hoogerheide, and van Dijk (2009).
Recall: Importance Sampling

- Let \( \pi(\theta) = \frac{f(\theta)}{Z} \), where \( f(\theta) = p(Y|\theta)p(\theta) \) and \( Z = \int f(\theta)d\theta = p(Y) \).

- Monte Carlo approximation of

\[
\mathbb{E}_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{1}{Z} \int h(\theta)w(\theta)g(\theta)d\theta,
\]

where \( w(\theta) = \frac{f(\theta)}{g(\theta)} \).

- We defined

\[
\bar{h}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{h(\theta^i)w(\theta^i)}{\sum_{i=1}^{N} w(\theta^i)},
\]

where the “particles” \( \theta^i \)'s are drawn from the distribution with density \( g(\cdot) \).

- Note that \( \frac{1}{N} \sum_{i=1}^{N} w(\theta^i) \xrightarrow{a.s.} p(Y) \).
Reciprocal importance samplers are based on the following identity:

\[
\frac{1}{p(Y)} = \int \frac{f(\theta)}{p(Y|\theta)p(\theta)} p(\theta|Y) d\theta, \quad (14)
\]

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

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

\[
\hat{p}_G(Y) = \left[ \frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f(\theta^{(s)})}{p(Y|\theta^{(s)})p(\theta^{(s)})} \right]^{-1}, \quad (15)
\]

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

Geweke (1999):

\[
f(\theta) = \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp \left[ -0.5(\theta - \bar{\theta})'V_\theta^{-1}(\theta - \bar{\theta}) \right] \times \left\{ (\theta - \bar{\theta})'V_\theta^{-1}(\theta - \bar{\theta}) \leq F_{\chi^2_d}^{-1}(\tau) \right\}. \quad (16)
\]
Rewrite Bayes Theorem:

\[ p(Y) = \frac{p(Y|\theta)p(\theta)}{p(\theta|Y)} \]  \hspace{2cm} (17)

Thus,

\[ \hat{p}_{CS}(Y) = \frac{p(Y|\tilde{\theta})p(\tilde{\theta})}{\hat{p}(\tilde{\theta}|Y)} \]  \hspace{2cm} (18)

where we replaced the generic \( \theta \) in (17) by the posterior mode \( \tilde{\theta} \).
• Use output of Metropolis-Hastings Algorithm.
• Proposal density for transition $\theta \mapsto \tilde{\theta}$: $q(\theta, \tilde{\theta}|Y)$.
• Probability of accepting proposed draw:

$$\alpha(\theta, \tilde{\theta}|Y) = \min \left\{ 1, \frac{p(\tilde{\theta}|Y)/q(\theta, \tilde{\theta}|Y)}{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)} \right\}.$$ 

• Note that

$$\int \alpha(\theta, \tilde{\theta}|Y)q(\theta, \tilde{\theta}|Y)p(\theta|Y)d\theta$$

$$= \int \min \left\{ 1, \frac{p(\tilde{\theta}|Y)/q(\theta, \tilde{\theta}|Y)}{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)} \right\} q(\theta, \tilde{\theta}|Y)p(\theta|Y)d\theta$$

$$= p(\tilde{\theta}|Y) \int \min \left\{ \frac{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)}{p(\tilde{\theta}|Y)/q(\tilde{\theta}, \theta|Y)}, 1 \right\} q(\tilde{\theta}, \theta|Y)d\theta$$

$$= p(\tilde{\theta}|Y) \int \alpha(\tilde{\theta}, \theta|Y)q(\tilde{\theta}, \theta|Y)d\theta$$
Posterior density at the mode can be approximated as follows

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

\[ (19) \]

• \{\theta^{(s)}\} are posterior draws obtained with the M-H Algorithm;
• \{\theta^{(j)}\} are additional draws from \( q(\tilde{\theta}, \theta|Y) \) given the fixed value \( \tilde{\theta} \).
### Log Marginal Data Densities

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Prior</th>
<th>$\mathcal{M}_0$</th>
<th>$\mathcal{M}_1$</th>
<th>$\mathcal{A}_0$</th>
<th>$\mathcal{A}_1$</th>
<th>VAR(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>1176.33</td>
<td>1178.45</td>
<td>1182.10</td>
<td>1180.21</td>
<td>1180.49</td>
</tr>
<tr>
<td>1</td>
<td>P1</td>
<td></td>
<td>1176.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>P2</td>
<td></td>
<td>1178.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>P3</td>
<td>1177.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>P4</td>
<td>1174.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• In our application the marginal likelihood values were pretty close to each other.

• However, in many DSGE applications the differences on a log scale could be 50, 100, 200, ...

• In many instances, such odds are “implausible” and suggest that the wrong models are being compared.
Now we are switching from the assessment of relative fit to the assessment of absolute fit.


Prior predictive check: does the model have a chance explaining salient features of the data?

Posterior predictive check: tries to assess the “absolute” fit of the model – similar to classical specification test.

We will also return to the DSGE model application.
Prior Predictive Checks

- Let $Y^{rep}$ be a sample of observations of length $T$ that we could have observed in the past or that we might observe in the future.
- Let’s construct a predictive distribution based on our prior knowledge for $Y^{rep}$:
  \[
  p(Y^{rep}) = \int p(Y^{rep} | \theta) p(\theta) \, d\theta
  \]
  
  - Let $S(Y)$ be a sample statistic of interest. From $p(Y^{rep})$ we can derive the predictive distribution of $p(S)$.
  - Compute the observed value of $S$ based on the actual data and assess how far it lies in the tails of its predictive distribution.
Posterior Predictive Checks

- Let $Y^{rep}$ be a sample of observations of length $T$ that we could have observed in the past or that we might observe in the future.
- Let’s construct a predictive distribution based on our posterior knowledge for $Y^{rep}$:
  \[
  p(Y^{rep}) = \int p(Y^{rep} | \theta) \, p(\theta | Y) \, d\theta
  \]
  Posterior
- Let $S(Y)$ be a sample statistic of interest. From $p(Y^{rep})$ we can derive the predictive distribution of $p(S)$.
- Compute the observed value of $S$ based on the actual data and assess how far it lies in the tails of its predictive distribution.
For $s = 1$ to $n_{sim}$:

1. Generate a draw $\theta^{(s)}$ from prior (posterior).
2. Simulate data $Y^{(s)}$ from model conditional on $\theta^{(s)}$.
3. Compute $S(Y^{(s)})$. 

Frank Schorfheide

DSGE Model Evaluation
Sample Moments of Hours – Posterior Predictive Distribution

For Model $M_0$ and Model $M_1$

- $\text{Autocorrelation} \times \text{Standard Deviation}$
- $\text{Autocorrelation} \times $
Joint Distribution of $\varphi$ and $\rho_b$

Prior Draws

<table>
<thead>
<tr>
<th>PHI</th>
<th>RHO_b</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

Posterior Draws

<table>
<thead>
<tr>
<th>PHI</th>
<th>RHO_b</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>
Comparison of DSGE and VAR

- ... has a long history.

- Goal: document in which dimensions the DSGE model dynamics are (in)consistent with the data.


- Important issues:
  - Can the state-space representation of the DSGE model be approximated by a VAR?
  - Precise estimation of the VAR coefficients (there are many!)
  - Identification of structural shocks in VAR
A Simple Example

- Consider the following models with $0 \leq \theta < 1$ and $\epsilon_t \sim iid(0, 1)$:

  $M_1: \quad y_t = \epsilon_t + \theta \epsilon_{t-1} = (1 + \theta L)\epsilon_t$

  $M_2: \quad y_t = \theta \epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t$

- $M_1$ and $M_2$ are observationally equivalent.

- Roots of MA polynomial

  $M_1: \quad z_* = 1/\theta > 1, \quad (1 + \theta L)^{-1} = \sum_{j=0}^{\infty} (-\theta)^j L^j$

  $M_2: \quad z_* = \theta < 1, \quad (\theta + L)^{-1} = \sum_{j=0}^{\infty} (-1/\theta)^j L^j$  (not convergent)

- AR($\infty$) representation in terms of $\epsilon_t$:
  - $M_1$: yes.
  - $M_2$: no.
A Simple Example

• What happens if we estimate AR(2) approximating model

\[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t. \]

• In population

\[
\begin{bmatrix}
\phi_1 \\
\phi_2
\end{bmatrix} = 
\begin{bmatrix}
\gamma_0 & \gamma_1 \\
\gamma_1 & \gamma_0
\end{bmatrix}^{-1} 
\begin{bmatrix}
\gamma_1 \\
\gamma_2
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
1 + \theta^2 & \theta \\
\theta & 1 + \theta^2
\end{bmatrix}^{-1} 
\begin{bmatrix}
\theta \\
0
\end{bmatrix} = 
\frac{1}{1 + \theta^2 + \theta^4} 
\begin{bmatrix}
1 + \theta^2 \\
-\theta
\end{bmatrix}
\]

• This is an AR(2) approximation of the AR(\(\infty\)) representation of model \(M_1\).

• If we increase the number of lags, the impulse responses will more and more look like the monotone IRFs of \(M_1\) rather than the hump-shaped response of \(M_2\).

• Lesson: if DSGE model generates non-invertible moving average terms its impulse responses cannot be approximated by a VAR(\(\infty\)) and a direct comparison of VAR and DSGE IRFs will be misleading.

2. Evaluation under a minimal econometric interpretation: Geweke (2010)

   - “Improve” DSGE model by relaxing its restrictions
   - Compare impulse responses of DSGE model and the “improved” DSGE model
• Idea: construct a prior distribution for parameters of a VAR that is centered at DSGE model restrictions.

• This relaxes potentially misspecified DSGE model restrictions.
• VAR(p):

\[ y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_0 + u_t, \quad \mathbb{E}[u_t u'_t] = \Sigma. \]

• Identification: \( u_t = \Sigma_{tr} \Omega \)

• Write VAR as \( Y = X\Phi + U \), \( Y \) is \( T \times n \), \( X \) is \( T \times k \).

• Create hierarchical model:

\[
p(Y, \Phi, \Sigma, \Omega, \theta) = p(Y|\Phi, \Sigma)p_{\lambda}(\Phi, \Sigma|\theta)p(\Omega|\theta)p(\theta),
\]

where \( p_{\lambda}(\Phi, \Sigma|\theta) \) is a prior distribution of the VAR parameters given the DSGE model parameters \( \theta \).
Specifying a Prior $p(\Phi, \Sigma|\theta)$

- Prior contours for misspecification parameters $\Phi^\Lambda$
- $\Phi^*(\theta)$: Cross-equation restriction for given value of $\theta$
- Subspace generated by the DSGE model restrictions
- $\Phi^*(\theta) + \Phi^\Lambda$
Specifying a Prior $p(\Phi, \Sigma|\theta)$

- Quasi-likelihood function for artificial observations (sample size $T^* = \lambda T$) generated from DSGE model:
  \[
p(Y^*(\theta)|\Phi, \Sigma) \propto |\Sigma_u|^{-\lambda T/2} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (-2\Phi'X^* Y^* + \Phi'X^*X^*\Phi)] \right\}.
  \]

- Let $\mathbb{E}_\theta^D[\cdot]$ be the expectation under DSGE model and define the autocovariance matrices
  \[
  \Gamma_{XX}(\theta) = \mathbb{E}_\theta^D [x_t x_t'], \quad \Gamma_{XY}(\theta) = \mathbb{E}_\theta^D [x_t y_t'].
  \]

- Replace sample moments $Y^* Y^*$ by $\mathbb{E}_\theta^D [Y^* Y^*] = \lambda T \Gamma_{YY}(\theta)$, etc.
Specifying a Prior $p(\Phi, \Sigma|\theta)$

- Define
  \[
  \Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta) \\
  \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta)
  \]

- Prior distribution:
  \[
  \Sigma|\theta \sim IW\left(\lambda T\Sigma^*(\theta), \lambda T - k\right) \\
  \Phi|\Sigma, \theta \sim N\left(\Phi^*(\theta), \frac{1}{\lambda T}\left[\Sigma^{-1} \otimes \Gamma_{XX}(\theta)\right]^{-1}\right),
  \]
• According to the DSGE model, the one-step-ahead forecast errors $u_t$ are functions of the structural shocks $\epsilon_t$: $u_t = \Sigma_{tr}\Omega\epsilon_t$.

• Let $A_0(\theta)$ be the contemporaneous impact of $\epsilon_t$ on $y_t$ according to the DSGE model and use the factorization

$$
\left( \begin{array}{c}
\frac{\partial y_t}{\partial \epsilon'_t} \\
\end{array} \right)_{DSGE} = A_0(\theta) = \Sigma^*_t(\theta)\Omega^*(\theta),
$$

where $\Sigma^*_t(\theta)$ is lower triangular and $\Omega^*(\theta)$ is an orthogonal matrix.

• Initial impact of $\epsilon_t$ on $y_t$ in the VAR:

$$
\left( \begin{array}{c}
\frac{\partial y_t}{\partial \epsilon'_t} \\
\end{array} \right)_{VAR} = \Sigma_{tr}\Omega.
$$

• Replace the rotation $\Omega$ in (21) with the function $\Omega^*(\theta)$ in (20).
• See DSGE model estimation.
The following factorization is useful for MCMC
\[
p_\lambda(\theta, \Phi, \Sigma, \Omega|Y) = p_\lambda(\theta|Y) \\
\times p_\lambda(\Phi, \Sigma|Y, \theta) \\
\times p(\Omega|Y, \Phi, \Sigma, \theta)
\]
The marginal posterior density of $\theta$ can be obtained by evaluating the marginal likelihood $p_\lambda(Y|\theta)$

$$p_\lambda(Y|\theta) = (2\pi)^{-nT/2} \frac{\lambda^T \Gamma XX(\theta) + X'X|^{-\frac{n}{2}} |(1 + \lambda) T \hat{\Sigma}_b(\theta)|^{-\frac{(1+\lambda)T-k}{2}}}{|\lambda^T \Gamma XX(\theta)|^{-\frac{n}{2}} |\lambda^T \Sigma^*(\theta)|^{-\frac{\lambda T-k}{2}}} \times 2^{\frac{n((1+\lambda)T-k)}{2}} \prod_{i=1}^n \Gamma[((1 + \lambda) T - k + 1 - i)/2] \times 2^{\frac{n(\lambda T-k)}{2}} \prod_{i=1}^n \Gamma[(\lambda T - k + 1 - i)/2].$$

and the prior density $p(\theta)$.

Draws from this posterior can be obtained in the same manner as draws in the regular Bayesian estimation of a DSGE model, e.g. with RW Metropolis algorithm.

Based on the MCMC output the marginal data density

$$p_\lambda(Y) = \int p_\lambda(\theta|Y)p(\theta)d\theta$$

can be approximated.
• The posterior distribution of $\Phi$ and $\Sigma$ is of the Inverted Wishart – Normal form:

$$
\Sigma|Y, \theta \sim IW\left((1 + \lambda) T \hat{\Sigma}_b(\theta), (1 + \lambda) T - k\right)
$$

$$
\Phi|Y, \Sigma, \theta \sim N\left(\hat{\Phi}_b(\theta), \Sigma \otimes (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}\right),
$$

• where $\hat{\Phi}_b(\theta)$ and $\hat{\Sigma}_b(\theta)$ are the given by

$$
\hat{\Phi}_b(\theta) = (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T \Gamma_{XY} + X'Y)
$$

$$
\hat{\Sigma}_b(\theta) = \frac{1}{(1 + \lambda) T} \left[(\lambda T \Gamma_{YY}(\theta) + Y'Y) - (\lambda T \Gamma_{YX}(\theta) + Y'X) \times (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T \Gamma_{XY}(\theta) + X'Y)\right].
$$
Posterior $p(\Omega | Y, \Phi, \Sigma, \theta)$

• Recall joint distribution:

$$p(Y, \Phi, \Sigma, \Omega, \theta) = p(Y | \Phi, \Sigma)p(\Phi, \Sigma | \theta)p(\Omega | \theta)p(\theta).$$

• Deduce:

$$p(\Omega | Y, \Phi, \Sigma, \theta) \propto p(\Omega | \theta)$$

• Here the conditional prior of $\Omega$ does not get updated because $\Omega$ does not enter the likelihood function.

• Also note that

$$p(\theta | Y, \Phi, \Sigma) \propto p(\Phi, \Sigma | \theta)p(\theta)$$

• But: marginal posterior $p(\theta | Y)$ gets updated because we learn about $(\Phi, \Sigma)$ from the data.
Use RWM Algorithm to generate a sequence of draws \( \theta^{(s)} \), \( s = 1, \ldots, n_{\text{sim}} \), from the posterior distribution of \( \theta \), given by \( p_\lambda(\theta|Y) \propto p_\lambda(Y|\theta)p(\theta) \). Moreover, compute \( \Omega^{(s)} = \Omega^*(\theta^{(s)}) \).

For \( s = 1, \ldots, n_{\text{sim}} \): draw a pair \((\Phi^{(s)}, \Sigma^{(s)})\) from its conditional MNIW posterior distribution given \( \theta^{(s)} \). \( \square \)
DSGE-VARs: Posterior of $\lambda$

- Numerical Illustration in An and Schorfheide (2005):
- Data Set 1: generated from the state-space representation of DSGE model that is used to construct prior for DSGE-VAR
- Data Set 2: generated from a different DSGE model

<table>
<thead>
<tr>
<th>Specification</th>
<th>Data Set 1</th>
<th>Data Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSGE Model</td>
<td>-196.66</td>
<td>-279.38</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = \infty$</td>
<td>-196.88</td>
<td>-277.49</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = 5.00$</td>
<td>-198.87</td>
<td>-270.46</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = 1.00$</td>
<td>-206.57</td>
<td>-258.25</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = 0.75$</td>
<td>-209.53</td>
<td>-257.53</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = 0.50$</td>
<td>-215.06</td>
<td>-258.73</td>
</tr>
<tr>
<td>DSGE-VAR $\lambda = 0.25$</td>
<td>-231.20</td>
<td>-269.66</td>
</tr>
</tbody>
</table>
• How well is the state-space representation of the linearized DSGE model approximated by the finite-order VAR? Compare DSGE-VAR(\(\hat{\lambda}\)) and DSGE IRFs.

• For each \(\theta\) draw compare responses of the state-space version of the DSGE to the DSGE-VAR(\(\lambda = \infty\)) version.
• How different are the IRFs of the VAR that is estimated subject to the DSGE model restrictions from the IRFs of the VAR in which restrictions are relaxed?

• For each \((\Phi, \Sigma, \theta)\) draw compare responses of the state-space version of the DSGE to the DSGE-VAR\((\lambda = \infty)\) version.

• We plot posterior mean responses of DSGE-VAR\((\lambda = \infty)\).

• Moreover, for each draw we compute the difference between DSGE-VAR\((\lambda)\) and DSGE-VAR\((\lambda = \infty)\). We use these differences to compute a posterior mean and 90% probability bands.
Suppose we rewrite the structural equations in a New Keynesian DSGE model as follows:

\[
\begin{align*}
\hat{y}_t - \hat{y}_{t+1} + \frac{1}{\tau} \left[ \hat{R}_t - \hat{E}_t \pi_{t+1} \right] &= (1 - \rho_g) \hat{g}_t + E_t \hat{z}_{t+1} \\
\hat{\pi}_t - \beta \hat{E}_t \left[ \hat{\pi}_{t+1} \right] - \kappa \hat{y}_t &= -\kappa \hat{g}_t \\
\hat{R}_t - \rho_R \hat{R}_{t-1} - (1 - \rho_R) \psi_1 \hat{\pi}_t &= -(1 - \rho_R) \psi_2 \hat{g}_t + \epsilon_{R,t} \\
-(1 - \rho_R) \psi_2 \hat{y}_t &=
\end{align*}
\]

For instance, in response to a monetary policy shock, the right-hand-side of the Euler equation and the Phillips curve equation has to be zero.

We can check these conditions for the DSGE-VAR(\(\lambda\)) response.

We overlay the right-hand-side for DSGE and DSGE-VAR.
Recall the estimated monetary DSGE model...

Monetary Policy Rule:

\[ R_t = R_{\*,t}^{1-\rho_R} R_{t-1}^{\rho_R} \exp\{\sigma_R \epsilon_{R,t}\} \]

\[ R_{\*,t} = (r_{\*,\pi_{\*,t}}) \left( \frac{\pi_t}{\pi_{\*,t}} \right)^{\psi_1} \left( \frac{Y_t}{\gamma Y_{t-1}} \right)^{\psi_2} \]

Agents forecast target inflation according to:

\[ \pi_{\*,t} = \pi_{\*,t-1} + \epsilon_{\pi,t} \]
Suppose we compute the posterior odds of the DSGE model versus a VAR of the form

\[
y_{1,t} = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Psi \Delta y_{2,t} + u_{1,t}
\]
\[
y_{2,t} = y_{2,t-1} + \sigma_{\pi_*} \epsilon_{\pi_*},t,
\]

- \(y_{1,t}\): output, inflation, interest rates, inverse velocity
- \(y_{2,t}\): target inflation rate constructed from inflation expectations and low-freq band-pass filtered inflation.

\[u_{1,t} \sim \mathcal{N}(0, \Sigma_{11})\] and is independent of \(\epsilon_{\pi_*},t\). This identifies \(\epsilon_{\pi_*},t\).

For now the VAR is equipped with Minnesota prior (see Del Negro and Schorfheide, 2010).

Posterior odds of VAR versus DSGE model

\[
\frac{\pi_{V,T}}{\pi_{D,T}} = \frac{\pi_{V,0}}{\pi_{D,0}} \frac{p(Y|M_V)}{p(Y|M_D)} \left[ e^{25} \right]
\]
Response to Target Inflation Shock

Real GDP

Inflation

Nom. Interest

Real Money

Frank Schorfheide
DSGE Model Evaluation
Response to Target Inflation Shock – DSGE (Blue) versus VAR (Red) IRFs

Real GDP

Inflation

Nom. Interest

Real Money
Now let’s construct a prior from the DSGE model...

Two modifications for the benchmark setup:

1. We combine the DSGE prior with a Minnesota prior such that the prior remains proper as $\lambda \to 0$

2. We make adjustments to account for the unit root in the DSGE model
The VAR can be written as $Y = X\Phi + U$

Recall that priors can be represented through dummy observations $Y^*$ and $X^*$.

Combine dummy observations that represent Minnesota prior with dummies that represent DSGE prior, e.g.:

$$X^*(\theta)'X^*(\theta) = (\lambda_D T)\Gamma_{XX}^D(\theta) + X^M'X^M$$

Similarly for $Y^*(\theta)'X^*(\theta)$ and $Y^*(\theta)'Y^*(\theta)$. 
Allowing for Unit Roots

• DSGE model has state-space representation

\[ y_t = \Psi_0 + \Psi_s s_t, \quad s_t = \Phi_1 s_{t-1} + \Phi_\epsilon \epsilon_t. \]

• If DSGE model has unit roots then autocovariances \( \Gamma_{XX}^D(\theta) \) are not time-invariant.

• Assume \( s_{-\tau} = 0 \) and \( \epsilon_t \sim iidN(0, \Sigma_\epsilon) \). Iterate state-transition equation forward to obtain joint distribution of \( y_0, \ldots, y_p \).

• Define matrices \( \Gamma_{XX}^D, \Gamma_{XY}^D, \) and \( \Gamma_{YY}^D \) based on covariance matrix of \( y_0, \ldots, y_p \).

• If some of the elements of \( s_t \) are non-stationary and others are stationary, the stationary ones can be initialized in period \( -\tau \) through their ergodic distribution, and the non-stationary ones with a pointmass at zero.

• In our application, \( s_t \) contains one non-stationary element, namely the target inflation rate, and we set \( \tau = 40 \).
Effect of a Change in Target Inflation (as Function of $\lambda$)