Model Evaluation

Frank Schorfheide
University of Pennsylvania
Econ 722 – Part 1

February 13, 2019

Two Questions

- Question 1: Does model $\mathcal{M}_1$ fit better than model $\mathcal{M}_2$?
  $\Longrightarrow$ Posterior model odds.
- Question 2: Are there patterns in the data that are inconsistent with model $\mathcal{M}_1$?
  $\Longrightarrow$ Posterior predictive checks.

DSGE Application

- 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.

Model Specifications

- We consider four versions of the stochastic growth model:
- In $\mathcal{M}_0$ and $\mathcal{M}_1$ firms can choose the employment level at the given wage rate without any adjustment cost.
- In $\mathcal{A}_0$ and $\mathcal{A}_1$, on the other hand, it is costly for firms to adjust the employment level.
- In $\mathcal{A}_0$ and $\mathcal{M}_0$ the labor supply shock is a stationary AR(1) process, whereas it is modeled as random walk in $\mathcal{A}_1$ and $\mathcal{M}_1$.

Households

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

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

- 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 ν. The labor supply shock is denoted by B_t.
- Per-period budget constraint faced by the household is

$$C_t + K_{t+1} - (1-\delta)K_t = W_t H_t + R_t K_t. \quad (2)$$

Firms

- Firms rent capital, hire labor services, and produce final goods according to the following Cobb-Douglas technology:

$$Y_t = (A_t H_t)^\alpha K_t^{1-\alpha} \left(1 - \varphi \cdot \left(\frac{H_t}{H_{t-1}} - 1 \right)^2 \right). \tag{3}$$

- 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

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^{t+s} \lambda_{t+s} (Y_t - W_t H_t^d - R_t K_t^d) \right], \tag{4}$$

where λ_t is the marginal value of a unit consumption to a household, which is treated as exogenous to the firm.

Equilibrium Conditions

- In equilibrium $\lambda_t = 1/C_t$ and the goods, labor, and capital markets clear:

$$Y_t = C_t + K_{t+1} - (1-\delta)K_t, \qquad H_t^d = H_t, \quad \text{and} \quad K_t^d = K_t.$$

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

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_{a,t}, \quad \epsilon_{a,t} \sim iid\mathcal{N}(0, \sigma_a^2). \tag{5}$$

The level of technology in period 0 is denoted by A_0.

Preference Shock

- In models $\mathcal{M}_0$ and $\mathcal{A}_0$, the labor supply shock follows a stationary AR(1) process:

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

  where $0 \leq \rho_b < 1$ and $\ln B_0$ is the unconditional mean of $\ln B_t$.
- In model $\mathcal{M}_0$ and $\mathcal{A}_0$ 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:

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

  and we use $\ln B_0$ to denote the initial level of $\ln B_t$.

Detrending

- It is well known that in models $\mathcal{M}_0$ and $\mathcal{A}_0$ 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:

$$\mathcal{M}_0: \quad \widetilde{Y}_t = \frac{Y_t}{A_t}, \quad \widetilde{C}_t = \frac{C_t}{A_t}, \quad \widetilde{K}_{t+1} = \frac{K_{t+1}}{A_t}.$$

- In models $\mathcal{M}_1$ and $\mathcal{A}_1$, 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:

$$\mathcal{M}_1: \quad \widetilde{H}_t = \frac{H_t}{B_t}, \quad \widetilde{Y}_t = \frac{Y_t}{A_t B_t}, \quad \widetilde{C}_t = \frac{C_t}{A_t B_t}, \quad \widetilde{K}_{t+1} = \frac{K_{t+1}}{A_t B_t}.$$

State-Space Representation

- We fit the DSGE models to observations on the log level of real per capita output and hours worked, denoted by the 2×1 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:

$$
\begin{aligned}
y_t &= \Gamma_0 + \Gamma_1 s_{1,t} + \Gamma_2 s_{2,t} + \Gamma_3 t & (8)\\
s_{1,t} &= \Phi_1 s_{1,t-1} + \Psi_1 \epsilon_t & (9)\\
s_{2,t} &= s_{2,t-1} + \Psi_2 \epsilon_t. & (10)
\end{aligned}
$$

State-Space Representation

- 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).

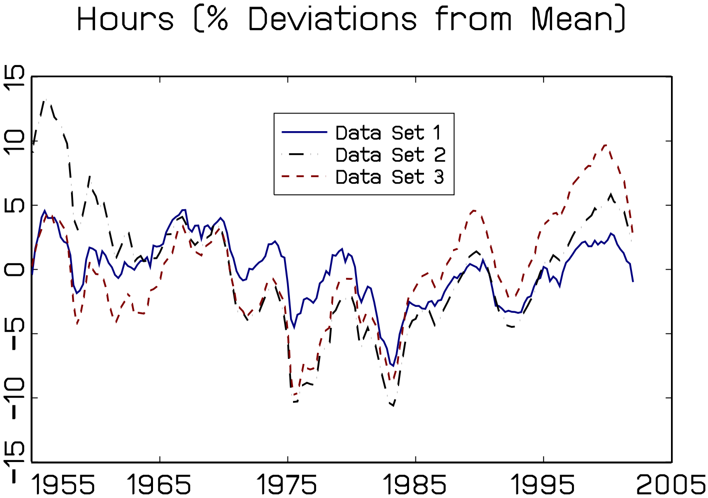
Initialization of Kalman Filter

- 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 $\mathcal{M}_0, \mathcal{A}_0$) and $[\ln A_0, \ln B_0]'$ (for $\mathcal{M}_1, \mathcal{A}_1$), respectively, is absorbed into the specification of our prior $p(\theta)$.

Data

- 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 (EEU00500005). 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.

Hours Worked



Prior

- 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 α is 0.66 and that for the quarter-to-quarter growth rate of productivity, γ, is 0.5%.
- The prior for β is centered at 0.995. Combined with the prior mean of γ, this corresponds to an annualized real return of about 4%.
- The depreciation rate δ lies between 1.8% and 3.3% per quarter.
- The 90% probability interval for the Frisch labor supply elasticity ν ranges from 0.3 to 1.8.

Prior

- We specify a prior for the adjustment cost parameter φ as follows.
  - In order to recruit labor ΔH, firms can either search for workers, incurring adjustment costs $\varphi(\frac{\Delta H}{H})^2 Y$, or pay head hunters for finding workers.
  - In the latter case the head hunters service fee is $\zeta W \Delta H$ where ζ 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: $\varphi(\frac{\Delta H}{H})^2 Y = \zeta W \Delta H$.
  - With the labor share of $1/3$ $(= \frac{WH}{Y})$ for a size of one percent increase of employment, $\frac{\Delta H}{H} = 1\%$, we obtain a range of 22 to 44 for φ.
  - 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 σ_a and σ_b are centered at 0.010, whereas under $\mathcal{A}_0$ and $\mathcal{A}_1$ they are centered at 0.015.

Prior

- For $\mathcal{M}_0$ and $\mathcal{A}_0$ 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 $\widetilde{H}^*$, evaluated at the prior mean values of the remaining structural parameters.
- For $\mathcal{M}_1$ and $\mathcal{A}_1$ the prior mean of $\ln B_0$ is obtained by equating hours worked in 1958:Q4 with the steady state level $B_0\widetilde{H}^*$. Similarly, we select the prior mean of $\ln A_0$ by matching $A_0\widetilde{Y}^*$ and $A_0 B_0\widetilde{Y}^*$, 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 $\mathcal{M}_0$ and $\mathcal{A}_0$ the 90% probability interval for the autoregressive parameter ρ_b ranges from 0.825 to 0.977, implying a fairly persistent labor supply process.

Prior

Parameter	Density	Data Set	Model	Para (1)	Para (2)
α	Beta	all	all	0.660	0.020
β	Beta	all	all	0.995	0.002
γ	Normal	all	all	0.005	0.005
δ	Beta	all	all	0.025	0.005
ν	Gamma	all	all	1.000	0.500
ρ_b	Beta	all	$\mathcal{M}_0$, $\mathcal{A}_0$	0.900	0.050
σ_a	InvGamma	all	$\mathcal{M}_0$, $\mathcal{M}_1$	0.010	1.000
		all	$\mathcal{M}_1$, $\mathcal{A}_1$	0.015	1.000
σ_b	InvGamma	all	$\mathcal{M}_0$, $\mathcal{M}_1$	0.010	1.000
		all	$\mathcal{M}_1$, $\mathcal{A}_1$	0.015	1.000
$\ln A_0$	Normal	1	$\mathcal{M}_0$, $\mathcal{A}_0$	5.647	0.200
		1	$\mathcal{M}_1$, $\mathcal{A}_1$	5.674	0.200
$\ln B_0$	Normal	1	$\mathcal{M}_0$, $\mathcal{A}_0$	3.236	0.200
		1	$\mathcal{M}_1$, $\mathcal{A}_1$	3.209	0.200
φ	Gamma	all	$\mathcal{A}_0$, $\mathcal{A}_1$	33.00	15.00

Prior

Parameter	Domain	Density	Data Set	Model	Para (1)	Para (2)
			Alternative Prior P1			
$\ln B_0$	$\mathbb{R}$	Normal	1	$\mathcal{M}_1$	3.209	2.000
			2	$\mathcal{M}_1$	6.405	2.000
			3	$\mathcal{M}_1$	6.309	2.000
			Alternative Prior P2			
$\ln B_0$	$\mathbb{R}$	Normal	1	$\mathcal{M}_1$	3.209	0.020
			2	$\mathcal{M}_1$	6.405	0.020
			3	$\mathcal{M}_1$	6.309	0.020
			Alternative Prior P3			
ρ_b	$[0, 1)$	Beta	all	$\mathcal{M}_0$	0.980	0.005
			Alternative Prior P4			
ρ_b	$[0, 1)$	Beta	all	$\mathcal{M}_0$	0.800	0.100

Recall: Posterior Odds and Marginal Data Densities

- Posterior model probabilities can be computed as follows:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_j \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad j = 1, \dots, 2, \tag{11}$$

- where

$$p(Y|\mathcal{M}) = \int p(Y|\theta, \mathcal{M}) p(\theta|\mathcal{M}) d\theta \tag{12}$$

- Note:

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

- Posterior odds and Bayes Factor

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \underbrace{\frac{\pi_{1,0}}{\pi_{2,0}}}_{\textit{Prior Odds}} \times \underbrace{\frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_2)}}_{\textit{Bayes Factor}} \tag{13}$$

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 $\pi(\theta) = f(\theta)/Z$, 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

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

  where the "particles" θ^i's are drawn from the distribution with density $g(\cdot)$.
- Note that $\frac{1}{N}\sum_{i=1}^N w(\theta^i) \xrightarrow{a.s.} p(Y)$.

Modified Harmonic Mean

- Reciprocal importance samplers are based on the following identity:

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{p(Y|\theta)p(\theta)} p(\theta|Y)d\theta, \tag{14}$$

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

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

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

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

- Geweke (1999):

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta})\right] \\ &\quad \times \left\{ (\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta}) \le F_{\chi^2_d}^{-1}(\tau) \right\}. \end{aligned} \tag{16}$$

Chib and Jeliazkov

- Rewrite Bayes Theorem:

$$p(Y) = \frac{p(Y|\theta)p(\theta)}{p(\theta|Y)}. \tag{17}$$

- Thus,

$$\hat{p}_{CS}(Y) = \frac{p(Y|\tilde{\theta})p(\tilde{\theta})}{\hat{p}(\tilde{\theta}|Y)}, \tag{18}$$

where we replaced the generic θ in (17) by the posterior mode $\tilde{\theta}$.

Chib and Jeliazkov

- 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

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

- Posterior density at the mode can be approximated as follows

$$\hat{p}(\tilde{\theta}|Y) = \frac{\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \alpha(\theta^{(s)}, \tilde{\theta}|Y) q(\theta^{(s)}, \tilde{\theta}|Y)}{J^{-1} \sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{(j)}|Y)}, \tag{19}$$

- $\{\theta^{(s)}\}$ are posterior draws obtained with the 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

Data Set	Prior	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{A}_0$	$\mathcal{A}_1$	VAR(4)
1	B	1176.33	1178.45	1182.10	1180.21	1180.49
	P1		1176.81			
	P2		1178.61			
	P3	1177.64				
	P4	1174.85				

Some Difficulties

- 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.

Predictive Checks

- Now we are switching from the assessment of relative fit to the assessment of absolute fit.
- See, for instance, Gelman, Carlin, Stern, and Rubin (1995), Lancaster (2003), Geweke (2005).
- 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}:

$$p(Y^{rep}) = \int p(Y^{rep}|\theta) \underbrace{p(\theta)}_{\text{Prior}} d\theta$$

- Let $\mathcal{S}(Y)$ be a sample statistic of interest. From $p(Y^{rep})$ we can derive the predictive distribution of $p(\mathcal{S})$.
- Compute the observed value of $\mathcal{S}$ 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}:

$$p(Y^{rep}) = \int p(Y^{rep}|\theta) \underbrace{p(\theta|Y)}_{\text{Posterior}} d\theta$$

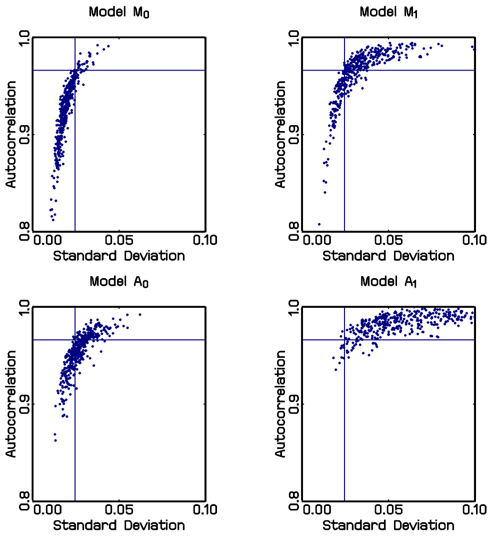
- Let $\mathcal{S}(Y)$ be a sample statistic of interest. From $p(Y^{rep})$ we can derive the predictive distribution of $p(\mathcal{S})$.
- Compute the observed value of $\mathcal{S}$ based on the actual data and assess how far it lies in the tails of its predictive distribution.

Implementation

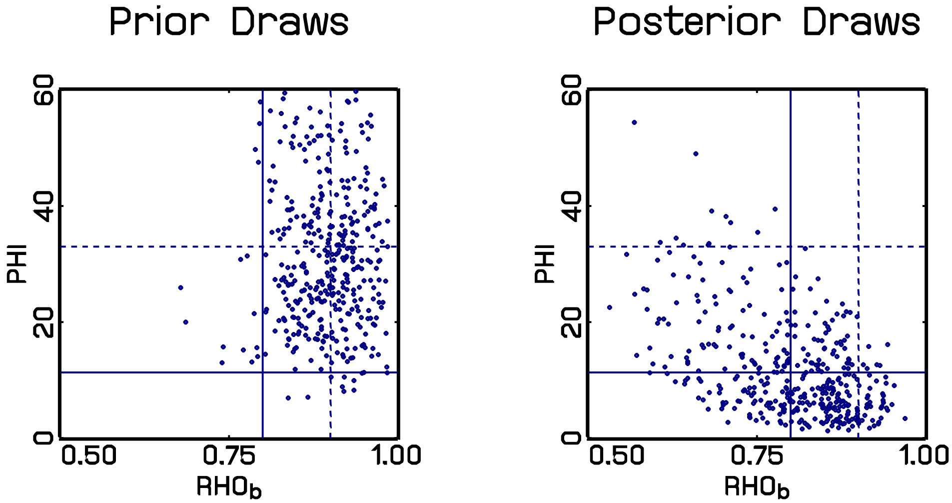
For $s = 1$ to n_{sim}:

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

Sample Moments of Hours – Posterior Predictive Distribution



Joint Distribution of φ and ρ_b



Comparison of DSGE and VAR

- ... has a long history.
- Goal: document in which dimensions the DSGE model dynamics are (in)consistent with the data.
- Examples: Cogley and Nason (1994), Rotemberg and Woodford (1997), Schorfheide (2000), Boivin and Giannoni (2003), and Christiano, Eichenbaum, and Evans (2004), to name a few.
- 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 $\epsilon_t \sim iid(0,1)$:

$$\begin{aligned} M_1 &: \quad y_t = \epsilon_t + \theta \epsilon_{t-1} = (1 + \theta L)\epsilon_t \\ M_2 &: \quad y_t = \theta \epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t \end{aligned}$$

- M_1 and M_2 are observationally equivalent.
- Roots of MA polynomial

$$\begin{aligned} M_1 &: \quad z_* = 1/\theta > 1, \quad (1 + \theta L)^{-1} = \sum_{j=0}^{\infty} (-\theta)^j L^j \\ M_2 &: \quad z_* = \theta < 1, \quad (\theta + L)^{-1} = \sum_{j=0}^{\infty} (-1/\theta)^j L^j \quad \text{(not convergent)} \end{aligned}$$

- AR(∞) representation in terms of ϵ_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

$$\begin{aligned} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} &= \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \\ &= \begin{bmatrix} 1+\theta^2 & \theta \\ \theta & 1+\theta^2 \end{bmatrix}^{-1} \begin{bmatrix} \theta \\ 0 \end{bmatrix} = \frac{1}{1+\theta^2+\theta^4} \begin{bmatrix} 1+\theta^2 \\ -\theta \end{bmatrix} \end{aligned}$$

- This is an AR(2) approximation of the AR(∞) 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(∞) and a direct comparison of VAR and DSGE IRFs will be misleading.

Bayesian Approaches for Comparison of DSGE Models and VARs

1. Loss function based evaluation: Schorfheide (2000)
2. Evaluation under a minimal econometric interpretation: Geweke (2010)
3. DSGE-VARs: Del Negro and Schorfheide (2004)
   - "Improve" DSGE model by relaxing its restrictions
   - Compare impulse responses of DSGE model and the "improved" DSGE model

Loss Function-Based Evaluation

- Question: does DSGE model $\mathcal{M}_1$ or model $\mathcal{M}_2$ deliver more precise predictions of a vector of population characteristics φ under a loss function $L(\varphi, \hat{\varphi})$?
- Use VAR as reference model $\mathcal{M}_0$.
- Create predictive density for φ:

$$p(\varphi|Y) = \sum_{i=0,1,2} \pi_{i,T} p(\varphi|Y, \mathcal{M}_i).$$

- Create DSGE model-specific predictions:

$$\hat{\varphi}_{(i)} = \text{argmin}_{\tilde{\varphi}} \int L(\tilde{\varphi}, \varphi) p(\varphi|Y, \mathcal{M}_i) d\varphi, \quad i = 1, 2.$$

- Rank models according to the overall posterior loss:

$$\int L(\hat{\varphi}_{(i)}, \varphi) p(\varphi|Y) d\varphi.$$

Loss Function-Based Evaluation

Some Applications

- Comparison of monetary DSGE models: Schorfheide (2000)
- Comparison of a learning-by-doing model with a regular stochastic growth model: Chang, Gomes, Schorfheide (2002).

Evaluation under Minimal Econometric Interpretation

- DSGE models only generate $p(\varphi|\mathcal{M}_i)$ instead of $p(Y|\mathcal{M}_i)$.
- VAR provides

$$p(\theta_{(0)}, Y|\varphi, \mathcal{M}_0) = p(\theta_{(0)}, Y|\varphi, \mathcal{M}_0)p(Y|\theta_{(0)}, \varphi, \mathcal{M}_0)$$

- Conditional on DSGE models we obtain

$$\begin{aligned} & p(\varphi, \theta_{(0)}, Y|\mathcal{M}_i, \mathcal{M}_0) \\ &= p(\varphi|\mathcal{M}_i)p(\theta_{(0)}, Y|\varphi, \mathcal{M}_0)p(Y|\theta_{(0)}, \varphi, \mathcal{M}_0) \end{aligned}$$

- Now notice that

$$\begin{aligned} & p(Y|\varphi, \mathcal{M}_i, \mathcal{M}_0) \\ &= \frac{\int_{\theta_{(0)}} p(\varphi|\mathcal{M}_i)p(\theta_{(0)}, Y|\varphi, \mathcal{M}_0)p(Y|\theta_{(0)}, \varphi, \mathcal{M}_0)d\theta_{(0)}}{p(\varphi|\mathcal{M}_i)} \\ &= \int_{\theta_{(0)}} p(\theta_{(0)}, Y|\varphi, \mathcal{M}_0)p(Y|\theta_{(0)}, \varphi, \mathcal{M}_0)d\theta_{(0)} \\ &= p(Y|\varphi, \mathcal{M}_0) \end{aligned}$$

is independent of $p(\varphi|\mathcal{M}_i)$.

Evaluation under Minimal Econometric Interpretation

- Finally (using the previous result),

$$\frac{p(\mathcal{M}_1|Y,\mathcal{M}_0)}{p(\mathcal{M}_2|Y,\mathcal{M}_0)} = \frac{p(\mathcal{M}_1|\mathcal{M}_0)}{p(\mathcal{M}_2|\mathcal{M}_0)}\frac{p(Y|\mathcal{M}_1,\mathcal{M}_0)}{p(Y|\mathcal{M}_2,\mathcal{M}_0)} = \frac{p(\mathcal{M}_1|\mathcal{M}_0)}{p(\mathcal{M}_2|\mathcal{M}_0)}\frac{\int p(\varphi|\mathcal{M}_1)p(Y|\varphi,\mathcal{M}_0)d\varphi}{\int p(\varphi|\mathcal{M}_2)p(Y|\varphi,\mathcal{M}_0)d\varphi}$$

- To construct $p(Y|\varphi,\mathcal{M}_0)$ introduce auxiliary model $\mathcal{M}_0^*$ such that:

$$p(Y|\varphi,\mathcal{M}_0) = \int_{\theta_{(0)}} p(Y,\theta_{(0)}|\varphi,\mathcal{M}_0)d\theta_{(0)}$$

$$\propto \int_{\theta_{(0)}} p(\varphi|\theta_{(0)},\mathcal{M}_0^*)p(Y|\theta_{(0)},\mathcal{M}_0^*)p(\theta_{(0)},\mathcal{M}_0^*)d\theta_{(0)}.$$

- Thus, to obtain $\int p(\varphi|\mathcal{M}_i)p(Y|\varphi,\mathcal{M}_0)d\varphi$ we are averaging

$$\int p(\varphi|\mathcal{M}_i^*)p(\varphi|\theta_{(0)},\mathcal{M}_0^*)d\varphi$$

with respect to the posterior $p(\theta_{(0)}|Y,\mathcal{M}_0^*)$.

Application

- Assess asset market implications of DSGE models: Geweke (2010)

DSGE-VARs

- Idea: construct a prior distribution for parameters of a VAR that is centered at DSGE model restrictions.
- This relaxes potentially misspecified DSGE model restrictions.

DSGE-VARs

- VAR(p):

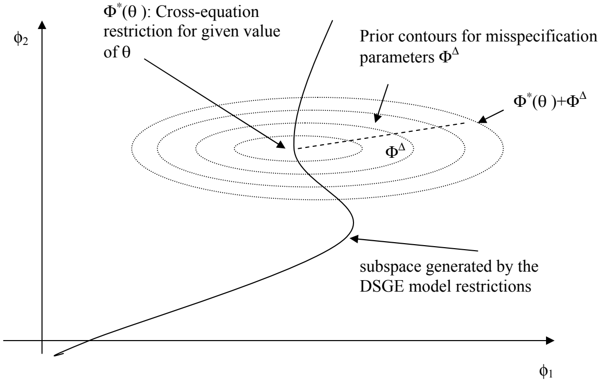
$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_0 + u_t, \quad \mathbb{E}[u_t u_t'] = \Sigma.$$

- 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 θ.

Specifying a Prior $p(\Phi, \Sigma|\theta)$



Specifying a Prior $p(\Phi, \Sigma|\theta)$

- Quasi-likelihood function for artificial observations (sample size $T^* = \lambda T$) generated from DSGE model:

$$p(Y^*(\theta)|\Phi, \Sigma) \propto$$
$$|\Sigma_u|^{-\lambda T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(-2\Phi' X^{*'} Y^* + \Phi' X^{*'} X^* \Phi)]\right\}.$$

- Let $\mathbb{E}^D_\theta[\cdot]$ be the expectation under DSGE model and define the autocovariance matrices

$$\Gamma_{XX}(\theta) = \mathbb{E}^D_\theta[x_t x_t'], \quad \Gamma_{XY}(\theta) = \mathbb{E}^D_\theta[x_t y_t'].$$

- Replace sample moments $Y^{*'} Y^*$ by $\mathbb{E}^D_\theta[Y^{*'} Y^*] = \lambda T \Gamma_{YY}(\theta)$, etc.

Specifying a Prior $p(\Phi, \Sigma|\theta)$

- Define

$$\begin{aligned} \Phi^*(\theta) &= \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta) \\ \Sigma^*(\theta) &= \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta) \end{aligned}$$

- Prior distribution:

$$\begin{aligned} \Sigma|\theta &\sim IW\left(\lambda T \Sigma^*(\theta), \lambda T - k\right) \\ \Phi|\Sigma,\theta &\sim N\left(\Phi^*(\theta), \frac{1}{\lambda T}\left[\Sigma^{-1} \otimes \Gamma_{XX}(\theta)\right]^{-1}\right), \end{aligned}$$

Prior for Ω

- According to the DSGE model, the one-step-ahead forecast errors u_t are functions of the structural shocks ϵ_t: $u_t = \Sigma_{tr}\Omega\epsilon_t$.
- Let $A_0(\theta)$ be the contemporaneous impact of ϵ_t on y_t according to the DSGE model and use the factorization

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{DSGE} = A_0(\theta) = \Sigma^*_{tr}(\theta)\Omega^*(\theta), \qquad (20)$$

  where $\Sigma^*_{tr}(\theta)$ is lower triangular and $\Omega^*(\theta)$ is an orthogonal matrix.
- Initial impact of ϵ_t on y_t in the VAR:

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{VAR} = \Sigma_{tr}\Omega. \qquad (21)$$

- Replace the rotation Ω in (21) with the function $\Omega^*(\theta)$ in (20).

Prior for θ

- See DSGE model estimation.

Posterior Distribution

- The following factorization is useful for MCMC

$$
\begin{aligned}
p_\lambda(\theta, \Phi, \Sigma, \Omega | Y) &= p_\lambda(\theta | Y) \\
&\times p_\lambda(\Phi, \Sigma | Y, \theta) \\
&\times p(\Omega | Y, \Phi, \Sigma, \theta)
\end{aligned}
$$

Posterior $p(\theta|Y)$

- The marginal posterior density of θ can be obtained by evaluating the marginal likelihood

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

  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.

Posterior $p_\lambda(\Phi, \Sigma | Y, \theta)$

- The posterior distribution of Φ and Σ is of the Inverted Wishart – Normal form:

$$\begin{aligned} \Sigma | Y, \theta &\sim IW\Big((1+\lambda) T \hat{\Sigma}_b(\theta), (1+\lambda)T - k \Big) \\ \Phi | Y, \Sigma, \theta &\sim N\Big(\hat{\Phi}_b(\theta), \Sigma \otimes (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1} \Big), \end{aligned}$$

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

$$\begin{aligned} \hat{\Phi}_b(\theta) &= (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1} (\lambda T \Gamma_{XY} + X'Y) \\ \hat{\Sigma}_b(\theta) &= \frac{1}{(1+\lambda)T} \Big[(\lambda T \Gamma_{YY}(\theta) + Y'Y) - (\lambda T \Gamma_{YX}(\theta) + Y'X) \\ &\quad \times (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1} (\lambda T \Gamma_{XY}(\theta) + X'Y) \Big]. \end{aligned}$$

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

- Recall joint distribution:

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

- Deduce:

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

- Here the conditional prior of Ω does not get updated because Ω does not enter the likelihood function.

- Also note that

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

- But: marginal posterior $p_\lambda(\theta|Y)$ gets updated because we learn about (Φ,Σ) from the data.

Posterior Draws for DSGE-VAR

1. Use RWM Algorithm to generate a sequence of draws $\theta^{(s)}$, $s = 1, \dots, n_{sim}$, from the posterior distribution of θ, given by $p_\lambda(\theta|Y) \propto p_\lambda(Y|\theta)p(\theta)$. Moreover, compute $\Omega^{(s)} = \Omega^*(\theta^{(s)})$.

2. For $s = 1, \dots, n_{sim}$: draw a pair $(\Phi^{(s)}, \Sigma^{(s)})$ from its conditional MNIW posterior distribution given $\theta^{(s)}$. $\square$

DSGE-VARs: Posterior of λ

- 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

Specification	Data Set 1	Data Set 2
DSGE Model	-196.66	-279.38
DSGE-VAR $\lambda = \infty$	-196.88	-277.49
DSGE-VAR $\lambda = 5.00$	-198.87	-270.46
DSGE-VAR $\lambda = 1.00$	-206.57	-258.25
DSGE-VAR $\lambda = 0.75$	-209.53	-257.53
DSGE-VAR $\lambda = 0.50$	-215.06	-258.73
DSGE-VAR $\lambda = 0.25$	-231.20	-269.66

DSGE-VARs: Comparison of DSGE and VAR

- 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 θ draw compare responses of the state-space version of the DSGE to the DSGE-VAR($\lambda = \infty$) version.

DSGE-VARs: Comparison of DSGE and VAR

- 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 (Φ, Σ, θ) 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(λ) and DSGE-VAR$(\lambda = \infty)$. We use these differences to compute a posterior mean and 90% probability bands.

DSGE-VARs: Comparison of DSGE and VAR

- Suppose we rewrite the structural equations in a New Keynesian DSGE model as follows:

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

- 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(λ) response.

- We overlay the right-hand-side for DSGE and DSGE-VAR.

DSGE-VARs: An Application

- 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:

$$\pi_{*,t} = \pi_{*,t-1} + \epsilon_{\pi,t}.$$

Evaluating a Monetary DSGE Model

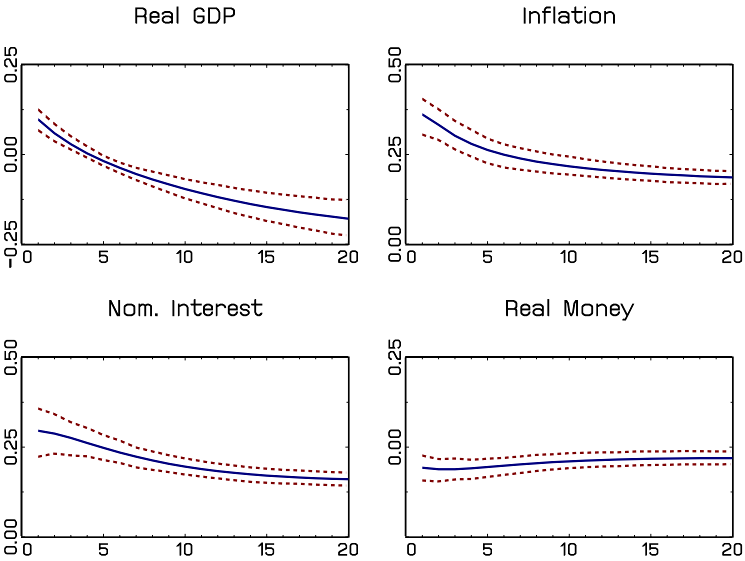
- Suppose we compute the posterior odds of the DSGE model versus a VAR of the form

$$\begin{aligned} y_{1,t} &= \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Psi \Delta y_{2,t} + u_{1,t} \\ y_{2,t} &= y_{2,t-1} + \sigma_{\pi_*} \epsilon_{\pi_*,t}, \end{aligned}$$

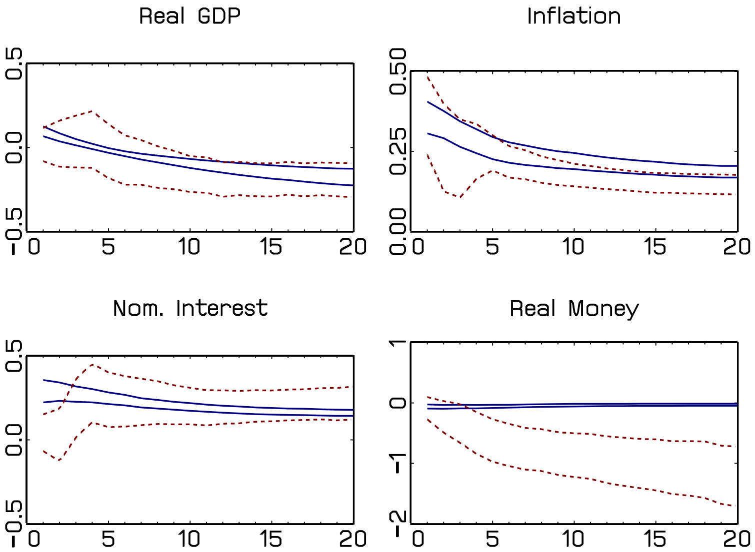
- $y_{1,t}$: output, inflation, interest rates, inverse velocity
- $y_{2,t}$: target inflation rate constructed from inflation expectations and low-freq band-pass filtered inflation.
- $u_{1,t} \sim \mathcal{N}(0, \Sigma_{11})$ and is independent of $\epsilon_{\pi_*,t}$. This identifies $\epsilon_{\pi_*,t}$.
- For now the VAR is equipped with Minnesota prior (see Del Negro and Schorfheide, 2010).
- Posterior odds of VAR versus DSGE model

$$\frac{\pi_{V,T}}{\pi_{D,T}} = \frac{\pi_{V,0}}{\pi_{D,0}} \underbrace{\frac{p(Y|\mathcal{M}_V)}{p(Y|\mathcal{M}_D)}}_{e^{25}}$$

Response to Target Inflation Shock



Response to Target Inflation Shock – DSGE (Blue) versus VAR (Red) IRFs



DSGE-VAR Estimation

- 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 $\lambda \longrightarrow 0$
  2. We make adjustments to account for the unit root in the DSGE model

Combining Minnesota Prior and DSGE Model Prior

- 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.:

$$X_*(\theta)'X_*(\theta) = (\lambda_D T)\Gamma^D_{XX}(\theta) + X_*^{M'}X_*^M$$

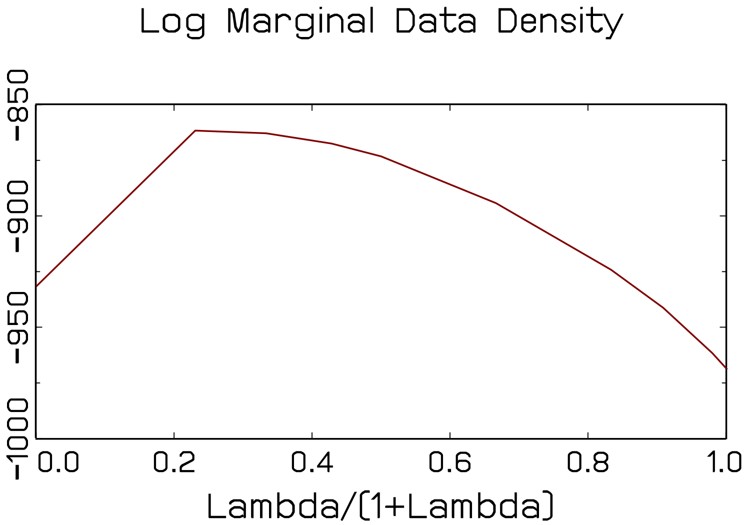
- Similarly for $Y_*(\theta)'X_*(\theta)$ and $Y_*(\theta)'Y_*(\theta)$.

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 $\Gamma^D_{XX}(\theta)$ 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 Γ^D_{XX}, Γ^D_{XY}, and Γ^D_{YY} 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 $\tau = 40$.


Log Marginal Data Density
-850
-900
-950
-1000
0.0
0.2
0.4
0.6
0.8
1.0
Lambda/(1+Lambda)

Effect of a Change in Target Inflation (as Function of λ)