Priors from General Equilibrium Models for VARs

by

Marco del Negro and Frank Schorfheide

Presenter: Keith O'Hara

March 10, 2014

The Main Idea of the Paper

- Use the implied moments of a DSGE model as the prior for a Bayesian VAR (a 'DSGE-VAR(λ)').
  - This is similar to a 'dummy observation' approach.
- We can view the policy functions of a DSGE model,

$$S_t = \mathcal{F} S_{t-1} + \mathcal{G} \varepsilon_t,$$

  as a VAR(1) with tight cross-equation restrictions.
- The parameter $\lambda \in (0, \infty)$ controls how 'close' the DSGE-VAR matches the DSGE model dynamics; it corresponds to the ratio of 'dummy observations' to actual data. As $\lambda \to \infty$, we approach the DSGE model.

Solving and Estimating a DSGE Model

Solving a DSGE Model

- We can express a log-linearized DSGE model as a system of linear expectational difference equations:

$$A_\theta S_t = B_\theta \mathbb{E}_t[S_{t+1}] + C_\theta S_{t-1} + D_\theta \varepsilon_t$$

  which we 'solve' to get a first-order VAR for the state's transition:

$$S_t = \mathcal{F} S_{t-1} + \mathcal{G} \varepsilon_t .$$

- $\mathcal{F}$ solves the following matrix quadratic, which implies a solution for $\mathcal{G}$:

$$0 = B_\theta \mathcal{F}^2 - A_\theta \mathcal{F} + C_\theta$$
$$\mathcal{G} = \left(A_\theta - B_\theta \mathcal{F}\right)^{-1} D_\theta$$

- We solve the matrix polynomial as a generalized eigenvalue problem.

The Log-Linearized Model

- For the log-linearized model in this paper, this corresponds to

$$x_t = \mathbb{E}_t x_{t+1} - \tau^{-1}(R_t - \mathbb{E}_t[\pi_{t+1}]) + (1-\rho_g)g_t + \rho_z \frac{1}{\tau} z_t$$
$$\pi_t = \frac{\gamma}{r^*}\mathbb{E}_t[\pi_{t+1}] + \kappa(x_t - g_t)$$
$$R_t = \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1 \pi_t + \psi_2 x_t\right) + \nu_t$$

with shock processes:

$$z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_{z,t}$$
$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t}$$
$$\nu_t = \sigma_R \varepsilon_{R,t}$$

- So $S_t = [x_t, \pi_t, R_t, z_t, g_t, \nu_t]^\top$.

Solving a DSGE Model

- Three cases to consider: (1) In the case of as many explosive eigenvalues (λ_e) as forward-looking equations, we have a unique solution to the problem.

(2) If there are more stable eigenvalues (λ_s) than forward-looking equations, there are many stable solutions for $\mathcal{F}$, one for each block-partition of λ_s. Here, equilibrium is indeterminate, and we face the issue of equilibrium selection.

(3) If there are less stable eigenvalues than forward-looking equations, then there are no non-explosive solutions, as there is no block-partition of λ_s such that all λ are stable.

Estimating a DSGE Model

- With our solution

$$S_t = \mathcal{F} S_{t-1} + \mathcal{G} \varepsilon_t,$$

  we have the state equation of a filtering problem.
- Assuming Gaussian disturbances, which, coupled with our linear problem, implies a Kalman filter approach with measurement equation

$$\mathcal{Y}_t = \mathcal{H}^\top S_t + \varepsilon_{\mathcal{Y},t}$$

- We proceed in three short steps, repeated for all t in $\{1, \ldots, T\}$:
  - predict the state at time $t+1$ given information available at time t;
  - update the state with new $\mathcal{Y}_{t+1}$ information; and
  - calculate the likelihood at $t+1$ based on forecast errors of $\mathcal{Y}_{t+1}$ and the covariance matrix of these forecasts.
- Classical ML and Bayesian estimation procedures are standard, with the latter being particularly popular; probably due to identification.

DSGE-VAR Details: Setup and Prior

VAR Model

- The standard VAR(p) model is denoted by

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, \quad u_t | y^{t-1} \sim N(\mathbf{0}, \Sigma_u)$$

- Let $k = 1 + n \times p$. In stacked form, we have

$$Y = X\Phi + U$$

with likelihood function

$$p(Y|\Phi,\Sigma_u) \propto |\Sigma_u|^{-T/2} \times \exp\left\{ -\frac{1}{2} \mathrm{tr} \left[\Sigma_u^{-1} (Y^\top Y - \Phi^\top X^\top Y - Y^\top X\Phi + \Phi^\top X^\top X\Phi) \right] \right\}$$

The Prior

- We look at the prior in terms of a 'dummy observation' approach.
- The prior is of the form

$$p(\Phi, \Sigma_u|\theta) = c^{-1}(\theta)|\Sigma_u|^{-\frac{\lambda T+n+1}{2}} \times$$
$$\exp\left\{-\frac{1}{2}\text{tr}\left[\lambda T \Sigma_u^{-1}(\Gamma^*_{yy}(\theta) - \Phi^\top \Gamma^*_{xy}(\theta) - \Gamma^*_{yx}(\theta)\Phi + \Phi^\top \Gamma^*_{xx}(\theta)\Phi)\right]\right\}$$

where $\Gamma^*_{yy}(\theta) := \mathbb{E}_\theta[y_t y_t^\top]$, etc, are the population moments.
- The form of the normalizing term $c^{-1}(\theta)$ is a little complicated.

The Prior

- For a model solution of the form

$$S_t = \mathcal{F} S_{t-1} + \mathcal{G}\varepsilon_t$$
$$\mathcal{Y}_t = \mathcal{H}^\top S_t$$

we first compute the steady state covariance matrix of the state by solving the discrete Lyapunov equation

$$\Omega_{ss} = \mathcal{F}\Omega_{ss}\mathcal{F}^\top + \mathcal{G}Q\mathcal{G}^\top$$

using a doubling algorithm, where $\mathbb{E}_\theta[S_t S_t^\top] = \Omega_{ss}$.

- Then we compute the Γ^* matrices with

$$\Gamma^*_{yy}(\theta) = \mathcal{H}^\top \Omega_{ss} \mathcal{H}$$
$$\Gamma^*_{yx_h}(\theta) = \mathcal{H}^\top \mathcal{F}^h \Omega_{ss} \mathcal{H}$$

Computational Aside: Solving for Ω_{ss}

- For completeness... we solve

$$\Omega_{ss} = \mathcal{F}\Omega_{ss}\mathcal{F}^\top + \mathcal{G}Q\mathcal{G}^\top$$

by iteration. Let $\mathcal{Q} := \mathcal{G}Q\mathcal{G}^\top$. Set $\Omega_{ss}(0) = \mathcal{Q}$ and $B(0) = \mathcal{F}$. Then, for $i = 1, 2, \ldots$

$$\Omega_{ss}(i+1) = \Omega_{ss}(i) + B(i)\Omega_{ss}(i)B(i)^\top$$
$$B(i+1) = B(i)B(i)^\top$$

- Continue until the difference between $\Omega_{ss}(i+1)$ and $\Omega_{ss}(i)$ is 'small'.
- Note: Ω_{ss} is a symmetric positive-definite matrix, so the relevant matrix norm here is the largest singular value (from a SVD).
- Could also use the vec-Kronecker trick: $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$.

The Prior

- Let

$$\Phi^*(\theta) = [\Gamma^*_{xx}(\theta)]^{-1}\Gamma^*_{xy}(\theta)$$
$$\Sigma^*_u(\theta) = \Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)[\Gamma^*_{xx}(\theta)]^{-1}\Gamma^*_{xy}(\theta)$$

  - Interpretation: If the data were generated by the DSGE model at hand, $\Phi^*(\theta)$ is the coefficient matrix of the VAR(p) that minimizes the one-step-ahead QFE loss.

- Given a θ, the prior distribution is of the usual IW-N form:

$$\Sigma_u|\theta \sim \mathcal{IW}\left(\lambda T \Sigma^*_u(\theta), \lambda T - k, n\right)$$
$$\Phi|\Sigma_u, \theta \sim \mathcal{N}\left(\Phi^*(\theta), \Sigma_u \otimes \left(\lambda T \Gamma^*_{xx}(\theta)\right)^{-1}\right)$$

- The joint prior is then given by

$$p(\Phi, \Sigma_u, \theta) = p(\Phi, \Sigma_u|\theta)p(\theta)$$

DSGE-VAR Posterior

The Posterior: Block 1

- The joint posterior distribution is factorized similarly:

$$p(\Phi, \Sigma_u, \theta | Y) = p(\Phi, \Sigma_u | Y, \theta) p(\theta | Y)$$

- The ML estimates are

$$\widehat{\Phi}(\theta) = \left[\lambda T \Gamma^*_{xx}(\theta) + X^\top X\right]^{-1} [\lambda T \Gamma^*_{xy} + X^\top Y]$$

$$\widehat{\Sigma}_u(\theta) = \frac{1}{(\lambda+1)T} \left[(\lambda T \Gamma^*_{yy}(\theta) + Y^\top Y)\right] - \frac{1}{(\lambda+1)T} \times$$
$$\left[(\lambda T \Gamma^*_{yx}(\theta) + Y^\top X)(\lambda T \Gamma^*_{xx}(\theta) + X^\top X)^{-1}(\lambda T \Gamma^*_{xy}(\theta) + X^\top Y)\right]$$

- The prior and likelihood are conjugate, so

$$\Sigma_u | Y, \theta \sim \mathcal{IW}\left((\lambda+1)T\widehat{\Sigma}_u(\theta), (1+\lambda)T - k, n\right)$$
$$\Phi | Y, \Sigma_u, \theta \sim \mathcal{N}\left(\widehat{\Phi}(\theta), \Sigma_u \otimes (\lambda T \Gamma^*_{xx}(\theta) + X^\top X)^{-1}\right)$$

The Posterior: Block 2

- The posterior distribution of the DSGE parameters is

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

where the marginal likelihood is

$$p(Y|\theta) = \int p(Y|\Phi, \Sigma_u)p(\Phi, \Sigma_u)d(\Phi, \Sigma_u) \tag{1}$$

- The authors show (in the appendix) that the closed form for (1) is

$$\begin{aligned} p(Y|\theta) &= \frac{p(Y|\Phi, \Sigma)p(\Phi, \Sigma|\theta)}{p(\Phi, \Sigma|Y)} \\ &= \frac{|\lambda T \Gamma^*_{xx}(\theta) + X^\top X|^{-\frac{n}{2}} |(\lambda+1)T\widehat{\Sigma}_u(\theta)|^{-\frac{(\lambda+1)T-k}{2}}}{|\lambda T \Gamma^*_{xx}(\theta)|^{-\frac{n}{2}} |\lambda T \Sigma^*_u(\theta)|^{-\frac{\lambda T - k}{2}}} \\ &\quad \times \frac{(2\pi)^{-nT/2} 2^{\frac{n((\lambda+1)T-k)}{2}} \prod_{i=1}^n \Gamma[((\lambda+1)T - k + 1 - i)/2]}{2^{\frac{n(\lambda T - k)}{2}} \prod_{i=1}^n \Gamma[(\lambda T - k + 1 - i)/2]} \end{aligned}$$

Sampling Algorithm

- Our previous discussion implies that a Metropolis-within-Gibbs MCMC algorithm would be appropriate.
- Given some value for θ, we sample Σ_u from

$$\Sigma_u | Y, \theta \sim \mathcal{IW}\left((\lambda + 1) T \widehat{\Sigma}_u(\theta), (1 + \lambda) T - k, n \right)$$

- Then, given θ and Σ_u, sample

$$\Phi | Y, \Sigma_u, \theta \sim \mathcal{N}\left(\widehat{\Phi}(\theta), \Sigma_u \otimes (\lambda T \Gamma^*_{xx}(\theta) + X^\top X)^{-1} \right)$$

- Given Φ and Σ_u, we evaluate a new θ draw using a Random Walk Metropolis MCMC algorithm.

Random Walk Metropolis Sampling Algorithm

- Given some initial θ (perhaps the posterior mode), draw a proposal $\theta^{(*)}$ from a jumping distribution,

$$\mathcal{N}(\theta^{(h-1)}, c \cdot \Sigma_m)$$

where Σ_m is the inverse of the Hessian computed at the posterior mode and c is a scaling factor.

- Compute the acceptation ratio,

$$\nu = \frac{p(Y|\theta^{(*)})p(\theta^{(*)})}{p(Y|\theta^{(h-1)})p(\theta^{(h-1)})}$$

- Finally, we accept or reject the proposal according to

$$\theta^{(h)} = \begin{cases} \theta^{(*)} & \mathbb{P} = \min\{\nu, 1\} \\ \theta^{(h-1)} & \text{else} \end{cases}$$

- Given this $\theta^{(h)}$, draw a new Σ_u, and so on.

Choosing λ

- How to choose λ?
- Construct a grid Λ and select the λ that maximizes the marginal data density

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

  That is,

$$\hat{\lambda} = \underset{\lambda \in \Lambda}{\operatorname{argmax}}\, p_\lambda(Y)$$

- $\hat{\lambda}$ is roughly 0.6, but forecasting exercises show $p_\lambda(Y)$ to be relatively flat between 0.5 and 2.

The Log-Linearized Model

The Log-Linearized Model

- Three main equations:

$$x_t = \mathbb{E}_t x_{t+1} - \tau^{-1}(R_t - \mathbb{E}_t[\pi_{t+1}]) + (1-\rho_g)g_t + \rho_z \frac{1}{\tau} z_t$$
$$\pi_t = \frac{\gamma}{r^*}\mathbb{E}_t[\pi_{t+1}] + \kappa(x_t - g_t)$$
$$R_t = \rho_R R_{t-1} + (1-\rho_R)\left(\psi_1 \pi_t + \psi_2 x_t\right) + \sigma_R \varepsilon_{R,t}$$

- Shock processes:

$$z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_{z,t}$$
$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t}$$

- Measurement equations:

$$\Delta \ln X_t = \ln \gamma + \Delta x_t + z_t$$
$$\Delta \ln P_t = \ln \pi^* + \pi_t$$
$$\ln R_t^a = 4\left[(\ln r^* + \ln \pi^*) + R_t\right]$$

Priors

Θ	Meaning	*Prior* Density	$\mathcal{P}_1$	$\mathcal{P}_2$
$\ln \gamma$	Technology Scaling	$\mathcal{N}$	0.500	0.250
$\ln \pi^*$	SS Inflation	$\mathcal{N}$	1.000	0.500
$\ln r^*$	γ/β	$\mathcal{G}$	0.500	0.250
κ	NKPC Slope	$\mathcal{G}$	0.300	0.150
τ	Coef. RRA	$\mathcal{G}$	2.000	0.500
ψ_1	MP Inflation	$\mathcal{G}$	1.500	0.250
ψ_2	MP Output Gap	$\mathcal{G}$	0.125	0.100
ρ_R	Interest Smoothing	$\mathcal{B}$	0.500	0.200
ρ_g	AR Gov. Spending	$\mathcal{B}$	0.800	0.100
ρ_z	AR Technology	$\mathcal{B}$	0.300	0.100
σ_R	SD Int.	$\mathcal{IG}$	0.251	0.139
σ_g	SD Gov.	$\mathcal{IG}$	0.630	0.323
σ_z	SD Tech.	$\mathcal{IG}$	0.875	0.430

Posterior CI for Different λ

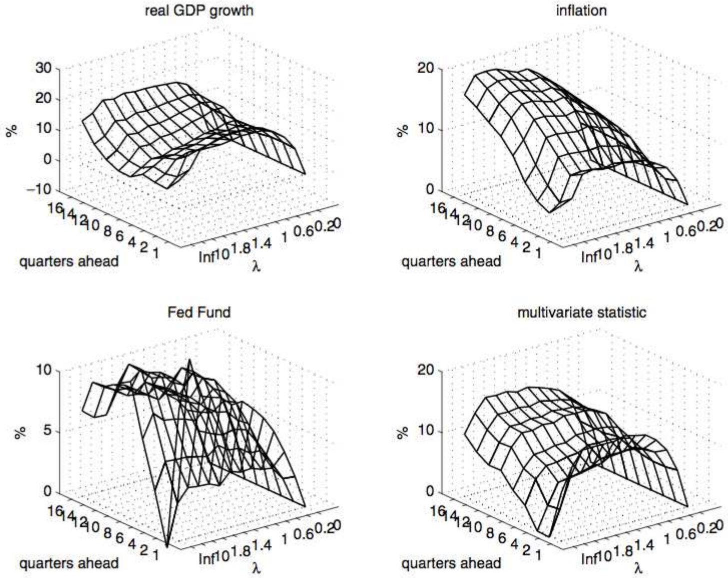
TABLE 2

POSTERIOR OF DSGE MODEL PARAMETERS: 1959:III–1979:II

	Prior		Posterior, $\lambda = 1$		Posterior, $\lambda = 10$	
Name	CI (Low)	CI (High)	CI (Low)	CI (High)	CI (Low)	CI (High)
$\ln \gamma$	0.101	0.922	0.473	1.021	0.616	1.045
$\ln \pi^*$	0.219	1.863	0.433	1.613	0.553	1.678
$\ln r^*$	0.132	0.880	0.113	0.463	0.126	0.384
κ	0.063	0.513	0.101	0.516	0.081	0.416
τ	1.197	2.788	1.336	2.816	1.684	3.225
ψ_1	1.121	1.910	1.011	1.559	1.009	1.512
ψ_2	0.001	0.260	0.120	0.497	0.150	0.545
ρ_R	0.157	0.812	0.530	0.756	0.550	0.747

NOTES: We report 90% confidence intervals (CI) based on the output of the Metropolis–Hastings Algorithm. The model parameters $\ln \gamma$, $\ln \pi^*$, and $\ln r^*$ are scaled by 100 to convert them into percentages.

RMSFE versus unrestricted VAR(4)



NOTES: The plot shows the percentage gain (loss) in RMSEs relative to an unrestricted VAR. The rolling sample is 1975:III–1997:III (90 periods). At each date in the sample, 80 observations are used

RMSFE versus unrestricted VAR(4) and BVAR(4)

TABLE 3

PERCENTAGE GAIN (LOSS) IN RMSES: DSGE PRIOR VERSUS UNRESTRICTED VAR (V-UNR) AND MINNESOTA PRIOR (V-MINN)

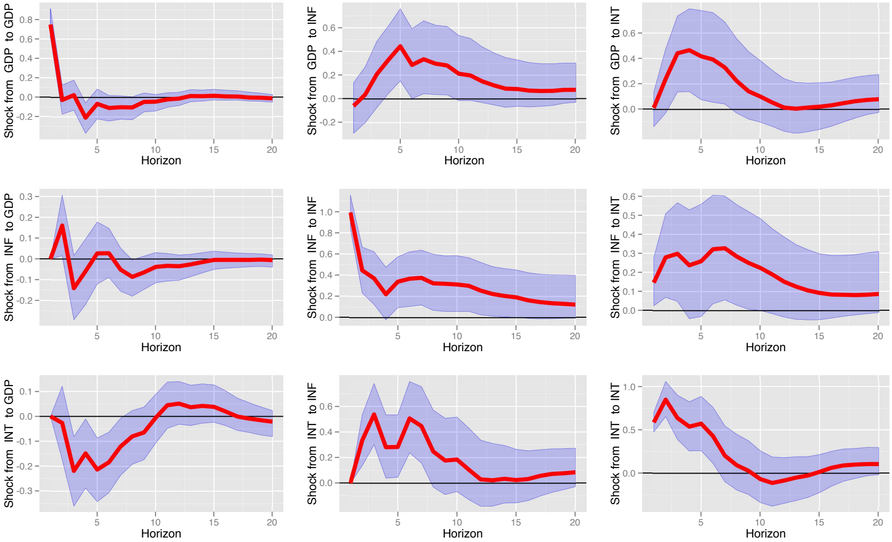
	rGDP Growth		Inflation		Fed Funds		Multivariate	
Horizon	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn
1	17.335	1.072	8.389	1.653	7.250	−7.593	12.842	0.943
2	16.977	6.965	7.247	1.339	5.024	−4.895	10.993	2.884
4	15.057	5.803	8.761	4.767	5.008	−1.878	9.630	3.959
6	14.116	3.452	10.460	7.240	6.648	−0.713	10.388	4.290
8	12.387	4.230	11.481	7.794	8.420	−0.204	11.023	5.187
10	14.418	7.986	12.261	8.351	8.242	−0.639	12.864	6.463
12	15.078	12.512	12.626	9.011	6.404	0.726	12.419	7.537
14	16.236	17.233	12.995	9.634	6.059	1.146	12.611	8.481
16	19.122	21.575	13.238	10.116	5.823	2.389	13.428	9.512

NOTES: The rolling sample is 1975:III–1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR. The forecasts are computed based on the values $\hat{\lambda}$ and $\hat{\iota}$ that have the highest posterior probability based on the estimation sample.

Impulse Response Functions

- Not entirely obvious how to relate the structural shocks from a DSGE model (ε_t) to the one-step ahead forecast errors of a VAR (u_t).
- We're used to seeing IRFs of the form...

IRFs from VAR(4)



Impulse Response Functions

- Let $\Sigma_{\text{tr}} = \text{Chol}(\Sigma_u)$. We have

$$u_t = \Sigma_{\text{tr}} \Omega \epsilon_t \tag{2}$$

where $\Omega^\top \Omega = \mathbb{I}$.

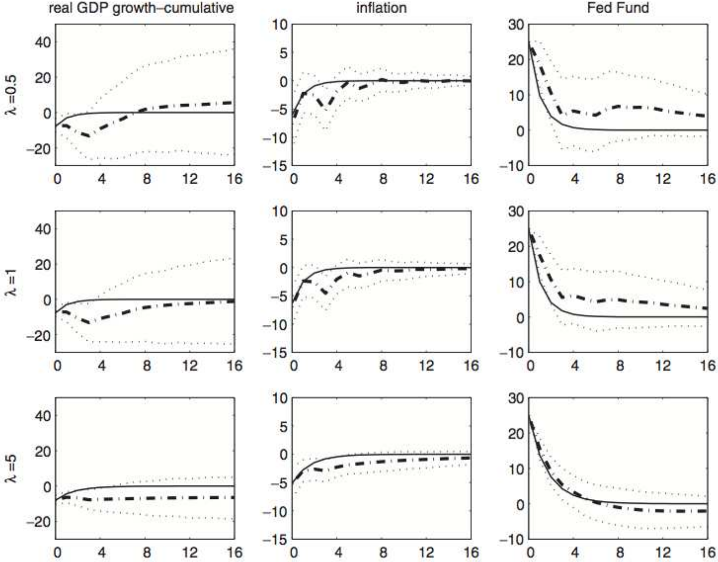
$$\left(\frac{\partial y_t}{\partial \varepsilon_t} \right)_{\text{VAR}} = \Sigma_{\text{tr}} \Omega$$

- For a DSGE model, there is a unique $A(\theta)$ that determines the contemporaneous effect of ε on y_t.

$$\left(\frac{\partial y_t}{\partial \varepsilon_t} \right)_{\text{DSGE}} = A_0(\theta) = \Sigma^*_{\text{tr}}(\theta) \Omega^*(\theta)$$

- Use a QR decomposition to get A_0. Keep Σ_{tr} in (2) but replace Ω with $\Omega^*(\theta)$.

IRFs



NOTES: The dashed-dotted lines represent the posterior means of the VAR impulse response functions. The dotted lines are 90% confidence bands. The solid lines represent the mean impulse responses from the DSGE model. The impulse responses are based on the sample 1981:IV–2001:III.

Conclusion

- Good forecasting performance.
- Perhaps better to compare forecasts from a DSGE/DSGE-VAR against a steady state BVAR. Mattias Villani has some nice work on this.
- DSGE-VAR extended to higher-order expansions?