The Evolution of U.S. Monetary Policy: 2000-2007

Michael T. Belongia
University of Mississippi

Peter N. Ireland
Boston College

November 2016

1 Notes, Data, and Programs

These notes, data, and programs will help you reproduce the econometric results in our paper, which appears in the December 2016 issue of the *Journal of Economic Dynamics and Control*. The original data series are available in separate Excel files, as downloaded from the Federal Reserve Bank of Saint Louis' FRED database. The same data series can be read directly into MATLAB from the file `tvpvardata.dat`.

The main program for running the estimations is `tvpvarsim.m`. All options for estimating different model variants can be set at the top of that file. Please note that when identification is achieved by imposing sign restrictions on impulse responses, `tvpvarsim.m` can take several days or more to run.

The Gibbs sampling output saved by `tvpvarsim.m` is used by `counterfact.m` to run counterfactual simulations and `poststats.m` to compute posterior summary statistics; those programs need to be run in order, after `tvpvarsim.m`. The programs named `tablex.m` and `figurey.m`, `x` $\in \{1, 3, 4\}$ and `y` $\in \{1, 2, \ldots, 9\}$ can then be run to generate the statistics reported in the paper's tables and figures. The remaining MATLAB `.m` files are used by `tvpvarsim.m` for the Gibbs sampling routine; comments at the top of each file describe its function in more detail.

We hope you will find these programs useful in your own research; but if you do, please cite our paper as well as Primiceri (2005), Del Negro and Primiceri (2015*a*, 2015*b*), and the others on which we have built.

2 The Model

The model is from Primiceri (2005), with the MCMC algorithm corrected as described by Del Negro and Primiceri (2015*b*). Quarterly data on the inflation rate (measured using the GDP deflator, the PCE price index, or the CPI) Π_t, a gap variable (either the output gap or the unemployment rate) G_t, and the short-term nominal interest rate (either the three-month Treasury bill rate or the federal funds rate) R_t run from 1969:1 through 2007:4, with the sample period extended beyond the 2001:3 cut-off date from the original paper, but terminating before the most recent zero-interest rate episode.

These observable series are combined into the 3×1 vector

$$y_t = \begin{bmatrix} \Pi_t & G_t & R_t \end{bmatrix}',$$

which is assumed to follow a second-order vector autoregression with time-varying coefficients and a time-varying covariance matrix for the innovations. Thus, the model's reduced form is

$$y_t = b_t + B_{1,t} y_{t-1} + B_{2,t} y_{t-2} + u_t, \tag{1}$$

where

$$b_t = \begin{bmatrix} b_{\pi,t} & b_{g,t} & b_{r,t} \end{bmatrix}$$

is a 3×1 vector of time-varying constant terms,

$$B_{j,t} = \begin{bmatrix} b_{j,\pi\pi,t} & b_{j,\pi g,t} & b_{j,\pi r,t} \\ b_{j,g\pi,t} & b_{j,gg,t} & b_{j,gr,t} \\ b_{j,r\pi,t} & b_{j,rg,t} & b_{j,rr,t} \end{bmatrix},$$

$j = 1, 2$, are 3×3 matrices of time-varying coefficients, and

$$u_t = \begin{bmatrix} u_{\pi,t} & u_{g,t} & u_{r,t} \end{bmatrix}'$$

is a 3×1 vector of heteroskedastic shocks with covariance matrix Ω_t.

Without loss of generality, Ω_t can be decomposed as

$$\Omega_t = A_t^{-1} \Sigma_t \Sigma_t' (A_t')^{-1}, \tag{2}$$

where A_t is the lower triangular matrix

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{g\pi,t} & 1 & 0 \\ \alpha_{r\pi,t} & \alpha_{rg,t} & 1 \end{bmatrix} \tag{3}$$

and Σ_t is the diagonal matrix

$$\Sigma_t = \begin{bmatrix} \sigma_{\pi,t} & 0 & 0 \\ 0 & \sigma_{g,t} & 0 \\ 0 & 0 & \sigma_{r,t} \end{bmatrix}. \tag{4}$$

The reduced form (1) can therefore be represented equivalently as

$$y_t = b_t + B_{1,t} y_{t-1} + B_{2,t} y_{t-2} + A_t^{-1} \Sigma_t \varepsilon_t, \tag{5}$$

where $E\varepsilon_t \varepsilon_t' = I_3$. Stacking all the coefficients into the 21×1 vector

$$B_t = \text{vec} \left(\begin{bmatrix} b_t' \\ B_{1,t}' \\ B_{2,t}' \end{bmatrix} \right),$$

(5) can be rewritten as

$$y_t = X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \tag{6}$$

where

$$X_t' = I_3 \otimes \begin{bmatrix} 1 & \Pi_{t-1} & G_{t-1} & R_{t-1} & \Pi_{t-2} & G_{t-2} & R_{t-2} \end{bmatrix}.$$

Let

$$\alpha_t = \begin{bmatrix} \alpha_{g\pi,t} & \alpha_{r\pi,t} & \alpha_{rg,t} \end{bmatrix}'$$

be the vector of non-zero and non-one elements of A_t and

$$\sigma_t = \begin{bmatrix} \sigma_{\pi,t} & \sigma_{g,t} & \sigma_{r,t} \end{bmatrix}'$$

be the vector of diagonal elements of Σ_t. The dynamics of the time-varying parameters are specified as

$$B_t = B_{t-1} + \eta_t, \tag{7}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t, \tag{8}$$

and

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t. \tag{9}$$

In (6)-(9), all of the innovations are assumed to be jointly normally distributed with

$$V = E \begin{bmatrix} \varepsilon_t \\ \eta_t \\ \zeta_t \\ \eta_t \end{bmatrix} \begin{bmatrix} \varepsilon_t & \eta_t & \zeta_t & \eta_t \end{bmatrix} = \begin{bmatrix} I_3 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \tag{10}$$

where Q is 21×21, S is 3×3, and W is 3×3 and diagonal, so that the standard deviations in σ_t evolve as independent geometric random walks. Following Primiceri (2005), it will be assumed that S is block-diagonal, with one non-zero element in the first column of the first row and three distinct non-zero elements in the second and third columns of the second and third rows. Hence, Q has 231 distinct elements, S has 4 distinct elements, and W has three distinct elements.

In all that follows, let

$$\omega^\tau = \begin{bmatrix} \omega_1' & \dots & \omega_\tau' \end{bmatrix}'$$

denote the history of a generic vector of variables ω_t up to a generic time τ. And for a generic matrix of variables and constant terms M_t, let

$$M^\tau = \begin{bmatrix} m_1' & \dots & m_\tau' \end{bmatrix}'$$

where m_t is a column vector constructed from the time varying elements of M_t.

3 Prior Distributions

Following Cogley and Sargent(2005) and Primiceri (2005), classical estimates of the parameters obtained by applying a training sample consisting of the first ten years of data to a constant-parameter version of the model are used to calibrate the prior means and standard deviations for the time-varying parameters when estimated with the rest of the sample. Writing the constant-parameter version of the reduced form (1) as

$$y_t = b + B_1 y_{t-1} + B_2 y_{t-2} + u_t,$$

Hamilton (1994, Ch.11) and Lutkepohl (2006, Ch.9) show that estimates of the constant and slope coefficients in b, B_1, and B_2 can be obtained by applying ordinary least squares separately to each equation. Stacking these estimated coefficients into the 21×1 vector

$$\hat{B} = \text{vec}\left(\begin{bmatrix} \hat{b}' \\ \hat{B}_1' \\ \hat{B}_2' \end{bmatrix}\right),$$

and defining

$$x_t' = \begin{bmatrix} 1 & \Pi_{t-1} & G_{t-1} & R_{t-1} & \Pi_{t-2} & G_{t-2} & R_{t-2} \end{bmatrix},$$

standard errors can be computed using the formulas from Hamilton's (1994, Ch.11, pp.298-299) Proposition 11.1:

$$\text{Var}(\hat{B}) = \hat{\Omega} \otimes \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1},$$

where

$$\hat{\Omega} = (1/T)\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$$

is the estimated covariance matrix for the least squares residuals

$$\hat{u}_t = y_t - \hat{b} - \hat{B}_1 y_{t-1} - \hat{B}_2 y_{t-2}.$$

It is convenient to assume that the initial states for the coefficients, covariances, and log volatilities as well as the hyperparameters in V are all independent of each other. The priors for B_0, α_0, and $\log \sigma_0$ are assumed to be normal and the priors for Q, W, and the blocks of S are assumed to be distributed as independent inverse-Wishart. These assumptiosn together with (5)-(7) imply normal priors on the entire sequences B^T, α^T, and Σ^T.

Estimates $\hat{A}$ and $\hat{\Sigma}$ of A and Σ can then be obtained by decomposing $\hat{\Omega}$ as in (2):

$$\hat{\Omega} = \hat{A}^{-1}\hat{\Sigma}\hat{\Sigma}'(\hat{A}')^{-1}.$$

Standard errors for the non-zero, non-one elements $\hat{\alpha}$ and $\hat{\sigma}$ of $\hat{A}$ and $\hat{\Sigma}$ can be computed using the formulas in Lutkepohl's (2006, Ch.9, p.373) Proposition 9.5. Start by rewriting

$$\text{vec}(A) = R_A \alpha + r_A$$

and

$$\text{vec}(\Sigma) = R_\Sigma \sigma$$

where R_A is a 9×3 matrix consisting of zeros except for

$R_A(2,1) = 1$

$R_A(3,2) = 1$

$R_A(6,3) = 1,$

r_A is a 9×1 vector consisting of zeros except for

$r_A(1) = 1$

$r_A(5) = 1$

$r_A(9) = 1,$

and R_Σ is a 9×3 matrix consisting of zeros except for

$R_\Sigma(1,1) = 1$

$R_\Sigma(5,2) = 1$

$R_\Sigma(9,3) = 1.$

Next, let K_9 be the commutation matrix that, for any 3×3 matrix D, is such that

$$\text{vec}(D) = K\text{vec}(D').$$

Then, in particular, K_9 consists of zeros, except for

$K_9(1,1) = 1$

$K_9(2,4) = 1$

$K_9(3,7) = 1$

$K_9(4,2) = 1$

$K_9(5,5) = 1$

$K_9(6,8) = 1$

$K_9(7,3) = 1$

$K_9(8,6) = 1$

$K_9(9,9) = 1$

With

$$\mathcal{I}\left(\begin{bmatrix}\text{vec}(A)\\ \text{vec}(\Sigma)\end{bmatrix}\right) = \begin{bmatrix}A^{-1}\Sigma \otimes \Sigma^{-1}\\ -(I_3 \otimes \Sigma^{-1})\end{bmatrix}(I_9 + K_9)\left[[\Sigma(A')^{-1}] \otimes \Sigma^{-1} \quad -(I_3 \otimes \Sigma^{-1})\right],$$

and

$$\mathcal{I}\left(\begin{bmatrix}\alpha\\ \sigma\end{bmatrix}\right) = \begin{bmatrix}R_A' & 0_{3,9}\\ 0_{3,9} & R_\Sigma'\end{bmatrix}\mathcal{I}\left(\begin{bmatrix}\text{vec}(A)\\ \text{vec}(\Sigma)\end{bmatrix}\right)\begin{bmatrix}R_A & 0_{9,3}\\ 0_{9,3} & R_\Sigma\end{bmatrix},$$

it follows from Lutkepohl's proposition that

$$\text{Var}\left(\begin{bmatrix}\hat{\alpha}\\ \hat{\sigma}\end{bmatrix}\right) = \frac{1}{T}\left[\mathcal{I}\left(\begin{bmatrix}\alpha\\ \sigma\end{bmatrix}\right)\right]^{-1}.$$

Priors can now be selected along the same lines proposed by Cogley and Sargent (2005), Primiceri (2005) and Benati (2011). Specifically, for B_0, α_0, and $\log \sigma_0$, it is assumed that

$$B_0 \sim N(\hat{B}, k_B^2 V_B),$$

$$\alpha_0 \sim N(\hat{\alpha}, k_\alpha^2 V_\alpha),$$

and

$$\log \sigma_0 \sim N(\log \hat{\sigma}, k_\sigma^2 I_3),$$

where choices for the hyperparameters are tabulated below.

Training Sample Prior Hyperparameters

	k_B^2	V_B	k_α	V_α	k_σ^2
Cogley-Sargent	1	$\text{Var}(\hat{B})$	10000	I_3	10
Primiceri	4	$\text{Var}(\hat{B})$	4	$\text{Var}(\hat{A})$	1
Benati	4	$\text{Var}(\hat{B})$	$\sqrt{10}$	$\text{diag}(\hat{\alpha})$	10

Note that (7)-(9) imply that

$$B_t | B_{t-1}, Q \sim N(B_{t-1}, Q),$$

$$\alpha_t | \alpha_{t-1}, S \sim N(\alpha_{t-1}, S),$$

and

$$\log \sigma_t | \sigma_{t-1}, W \sim N(\log \sigma_{t-1}, W).$$

Hence, priors for the entire sequences B^T, α^T, and Σ^T are

$$p(B^T | B_0, Q) = \prod_{t=1}^{T} p(B_t | B_{t-1}, Q),$$

$$p(\alpha^T | \alpha_0, Q) = \prod_{t=1}^{T} p(\alpha_t | \alpha_{t-1}, S),$$

and

$$p(\Sigma^T | \Sigma_0, Q) = \prod_{t=1}^{T} p(\log \sigma_t | \log \sigma_{t-1}, W).$$

For Q and the two blocks of S, the inverse Wishart priors are calibrated as

$$Q \sim IW(d_Q k_Q^2 V_Q, d_Q),$$

$$S_1 \sim IW(d_{S1} k_S^2 V_{S1}, d_{S1}),$$

and

$$S_2 \sim IW(d_{S2} k_S^2 V_{S2}, d_{S2}).$$

Finally, for each diagonal element $w_{i,i}$, $i = 1, 2, 3$, of W, the inverse Gamma prior used by Cogley and Sargent (2005) and Benati (2011) can also be expressed as an inverse Wishart:

$$w_{i,i} \sim IG\left(\frac{d_W}{2}, \frac{d_W k_W^2}{2}\right) = IW(d_W k_W^2, d_W).$$

Choices for the hyperparameters are tabulated below. Cogley and Sargent (2005) do not allow for time-variation in the elements of A.

Time-Varying Parameter Prior Hyperparameters

	k_Q^2	V_Q	d_Q	k_S^2	V_{S1}	d_{S1}	V_{S2}	d_{S2}	k_W^2	d_W
Cogley-Sargent	0.00035	$\text{Var}(\hat{B})$	22	–	–	–	–	–	0.0001	1
Primiceri	0.0001	$\text{Var}(\hat{B})$	40	0.01	$V_\alpha^{1,1}$	2	$V_\alpha^{2:3,2:3}$	3	0.0001	2
Benati	0.00035	$\text{Var}(\hat{B})$	22	0.001	$\hat{\alpha}^1$	2	$\text{diag}(\hat{\alpha}^{2:3})$	3	0.0001	1

where $V_\alpha^{1,1}$ is the element from the first row and column of V_α, $V_\alpha^{2:3,2:3}$ is the matrix formed from the last two rows and columns of V_α, and $\hat{\alpha}^1$ and $\hat{\alpha}^{2:3}$ correspond to the first and the second through third elements of the vector $\hat{\alpha}$.

4 The Markov Chain Monte Carlo Algorithm

The algorithm gets initialized by choosing initial draws for α^T, σ^T, and V from the prior distributions described above. The Gibbs sampling algorithm then loops through the following steps.

4.1 Drawing the Coefficient States

Conditional on (α^T, σ^T, V), the observation equation (4) in linear and has Gaussian innovations with known variance. As shown in Carter and Kohn (1994) and Fruhwirth-Schnatter (1994), the density can be factored as

$$p(B^T|y^T, \alpha^T, \sigma^T, V) = p(B_T|y^T, \alpha^T, \sigma^T, V) \prod_{t=1}^{T-1} p(B_t|B_{t+1}, y^t, \alpha^T, \sigma^T, V),$$

where

$$B_t|B_{t+1}, y^t, \alpha^T, \sigma^T, V \sim N(B_{t|t+1}, P_{t|t+1}),$$
$$B_{t|t+1} = E(B_t|B_{t+1}, y^t, \alpha^T, \sigma^T, V),$$

and

$$P_{t|t+1} = \text{Var}(B_t|B_{t+1}, y^t, \alpha^T, \sigma^T, V).$$

The vector of B's can be drawn easily because $B_{t|t+1}$ and $P_{t|t+1}$ can be computed using forward and backward recursions on the Kalman filter as follows.

The measurement equation for this step is (6), rewritten as

$$y_t = X_t' B_t + u_t \tag{11}$$

where $u_t = A_t^{-1}\Sigma_t \varepsilon_t$, $Eu_t u_t' = \Omega_t$ and $\Omega_t = A_t^{-1}\Sigma_t \Sigma_t' (A_t^{-1})'$, and the state transition equation is (7):

$$B_t = B_{t-1} + \nu_t, \tag{7}$$

where $E\nu_t \nu_t' = Q$. Let

$$B_{t|s} = E(B_t|y^s, X^s, \Omega^s, Q)$$

and

$$P_{t|s} = \text{Var}(B_t|y^s, X^s, \Omega^s, Q).$$

Then, given $B_{0|0} = \hat{B}$ and $P_{0|0} = k_B^2 V_B$, the Kalman filter implies

$$B_{t|t-1} = B_{t-1|t-1},$$

$$P_{t|t-1} = P_{t-1|t-1} + Q,$$
$$K_t = P_{t|t-1}X_t(X_t'P_{t|t-1}X_t + \Omega_t)^{-1},$$
$$B_{t|t} = B_{t|t-1} + K_t(y_t - X_t'B_{t|t-1}),$$

and

$$P_{t|t} = P_{t|t-1} - K_t X_t' P_{t|t-1}.$$

The last elements from these recursions are $B_{T|T}$ and $P_{T|T}$, which are the mean and variance of the normal distribution used to make a draw for B_T. The draw for B_T and the output of the filter can now be used for the first step of the backward recursions

$$B_{t|t+1} = B_{t|t} + P_{t|t}P_{t+1|t}^{-1}(B_{t+1} - B_{t|t}) = B_{t|t} + P_{t|t}(P_{t|t} + Q)^{-1}(B_{t+1} - B_{t|t})$$

and

$$P_{t|t+1} = P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t} = P_{t|t} - P_{t|t}(P_{t|t} + Q)^{-1}P_{t|t},$$

which are the means and variances used to make the draws for B_t, $t = T-1, T-2, \ldots, 1$.

4.2 Drawing Covariance States

The system of equations in (6) can be rewritten as

$$A_t(y_t - X_t'B_t) = A_t u_t = \Sigma_t \varepsilon_t, \tag{12}$$

where, taking B^T as given, u_t is observable from (11). Since A_t is a lower triangular matrix with ones on the main diagonal, (12) can be rewritten as

$$u_t = Z_t \alpha_t + \Sigma_t \varepsilon_t, \tag{13}$$

where α_t is defined in (8) and Z_t is the following 3×3 matrix:

$$Z_t = \begin{bmatrix} 0 & 0 & 0 \\ -u_{\pi,t} & 0 & 0 \\ 0 & -u_{\pi,t} & -u_{g,t} \end{bmatrix}$$

The model given by (13) and (8) has a Gaussian but nonlinear state space representation. The problem is that the dependent variable of the observation equation, u_t, also appears on the right-hand size in Z_t. Therefore, the vector $\begin{bmatrix} u_t & \alpha_t \end{bmatrix}$ is not jointly normal and, as a consequence, the conditional distributions cannot be computed using the standard Kalman filter recursions. However, under the additional maintained assumption that S is block diagonal, this problem can be solved by applying the Kalman filter and the backward recursion equation by equation.

Thus, consider the second equation from (13), which can be written

$$u_{g,t} = Z_{1t}\alpha_{g\pi,t} + \sigma_{g,t}\varepsilon_{g,t}, \tag{14}$$

where $Z_{1t} = -u_{\pi,t}$ and $\varepsilon_{g,t} \sim iidN(0,1)$. Taking B^T and σ^T as given, u_{gt} and Z_{1t} are observable and $\sigma_{g,t}$ is given as well. Equation (14) can serve as the observation equation and the first equation from (8),

$$\alpha_{g\pi,t} = \alpha_{g\pi,t-1} + \zeta_{1,t} \tag{15}$$

as the state transition equation, where $\xi_{1,t} \sim N(0, S_1)$, with S_1 given as well.

Thus, given $\alpha_{g\pi,0|0} = \hat{\alpha}_{g\pi}$ and $P_{0|0} = k_\alpha^2 V_\alpha^{1,1}$, the Kalman filter implies

$$\alpha_{g\pi,t|t-1} = \alpha_{g\pi,t-1|t-1},$$

$$P_{t|t-1} = P_{t-1|t-1} + S_1,$$
$$K_t = P_{t|t-1}Z_{1t}'(Z_{1t}P_{t|t-1}Z_{1t}' + \sigma_{g,t}^2)^{-1},$$
$$\alpha_{g\pi,t|t} = \alpha_{g\pi,t|t-1} + K_t(u_{g,t} - Z_{1t}\alpha_{g\pi,t|t-1}),$$

and

$$P_{t|t} = P_{t|t-1} - K_t Z_{1t} P_{t|t-1}.$$

The last elements from these recursions are $\alpha_{g\pi,T|T}$ and $P_{T|T}$, which are the mean and variance of the normal distribution used to make a draw for $\alpha_{g\pi,T}$. The draw for $\alpha_{g\pi,T}$ and the output of the filter can now be used for the first step of the backward recursions

$$\alpha_{g\pi,t|t+1} = \alpha_{g\pi,t|t} + P_{t|t}P_{t+1|t}^{-1}(\alpha_{g\pi,t+1} - \alpha_{g\pi,t|t}) = \alpha_{g\pi,t|t} + P_{t|t}(P_{t|t} + S_1)^{-1}(\alpha_{g\pi,t+1} - \alpha_{g\pi,t|t})$$

and

$$P_{t|t+1} = P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t} + P_{t|t} - P_{t|t}(P_{t|t} + S_1)^{-1}P_{t|t},$$

which are the means and variances used to make the draws for $\alpha_{u\pi,t}$, $t = T-1, T-2, \ldots, 1$.

Now consider the third equation from (13), which can be written

$$u_{r,t} = Z_{2t}\alpha_{2,t} + \sigma_{r,t}\varepsilon_{r,t}, \tag{16}$$

where $Z_{2t} = \begin{bmatrix} -u_{\pi,t} & -u_{g,t} \end{bmatrix}$, $\alpha_{2,t} = \begin{bmatrix} \alpha_{r\pi,t} & \alpha_{rg,t} \end{bmatrix}'$, and $\varepsilon_{r,t} \sim iidN(0,1)$. Taking B^T and σ^T as given, u_{rt} and Z_{2t} are observable and $\sigma_{r,t}$ is given as well. Equation (16) can serve as the observation equation and last two equations from (8),

$$\alpha_{2,t} = \alpha_{2,t-1} + \xi_{2,t} \tag{17}$$

as the state transition equation, where $\xi_{2,t} \sim N(0, S_2)$, with S_2 given as well.

Thus, given $\alpha_{2,0|0} = \begin{bmatrix} \hat{\alpha}_{r\pi} & \hat{\alpha}_{rg,t} \end{bmatrix}'$ and $P_{0|0} = k_\alpha^2 V_\alpha^{2:3,2:3}$, the Kalman filter implies

$$\alpha_{2,t|t-1} = \alpha_{2,t-1|t-1},$$

$$P_{t|t-1} = P_{t-1|t-1} + S_2,$$
$$K_t = P_{t|t-1}Z_{2t}'(Z_{2t}P_{t|t-1}Z_{2t}' + \sigma_{r,t}^2)^{-1},$$

$$\alpha_{2,t|t} = \alpha_{2,t|t-1} + K_t(u_{r,t} - Z_{2t}\alpha_{2,t|t-1}),$$

and

$$P_{t|t} = P_{t|t-1} - K_t Z_{2t} P_{t|t-1}.$$

The last elements from these recursions are $\alpha_{2,T|T}$ and $P_{T|T}$, which are the mean and variance of the normal distribution used to make draws for $\alpha_{r\pi,T}$ and $\alpha_{rg,T}$. The draw for $\alpha_{2,T}$ and the output of the filter can now be used for the first step of the backward recursions

$$\alpha_{2,t|t+1} = \alpha_{2,t|t} + P_{t|t}P^{-1}_{t+1|t}(\alpha_{2,t+1} - \alpha_{2,t|t}) = \alpha_{2,t|t} + P_{t|t}(P_{t|t} + S_2)^{-1}(\alpha_{2,t+1} - \alpha_{2,t|t})$$

and

$$P_{t|t+1} = P_{t|t} - P_{t|t}P^{-1}_{t+1|t}P_{t|t} = P_{t|t} - P_{t|t}(P_{t|t} + S_2)^{-1}P_{t|t},$$

which are the means and variances used to make the draws for $\alpha_{u\pi,t}$, $t = T-1, T-2, \ldots, 1$.

4.3 Drawing Volatility States

Consider next the system of equations

$$A_t(y_t - X_t'B_t) = y_t^* = \Sigma_t \varepsilon_t, \tag{18}$$

where, taking B^T and α^T as given, y_t^* is observable. This is a system of nonlinear measurement equations, but can be converted into a linear one by squaring and taking logs of every element of (18). Due to the fact that $y_{i,t}^2$ can be very small, an offset constant can be used to make the estimation procedure more robust. This leads to the following approximating state space form:

$$y_t^{**} = 2h_t + e_t \tag{19}$$

and

$$h_t = h_{t-1} + \eta_t, \tag{20}$$

where $y_{i,t}^{**} = \log[(y_{i,t}^*)^2 + \bar{c}]$, $\bar{c}$ is the offset constant, set equal to 0.001, $e_{i,t} = \log(\varepsilon_{i,t}^2)$, and $h_{i,t} = \log \sigma_{i,t}$.

This system has a linear, but non-Gaussian, state space form because the innovations in the measurement equations are distributed as $\log \chi^2(1)$. In order to further transform the system into a Gaussian one, a mixture of normals approximation of the $\log \chi^2$ distribution is used, as described by Kim, Shephard, and Chib (1998). This involves selecting a mixture of seven normal densities with component probabilities q_j, means $m_j - 1.2704$ and variances v_j^2, where the constants are chosen to match a number of moments of the $\log \chi^2(1)$ distribution as reported in Table 4 (p.371) from their paper:

Distribution for the Indicator Variables

ω	$q_j = Pr(\omega = j)$	m_j	v_j^2
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Define $s^T = \begin{bmatrix} s_1 & \dots & s_T \end{bmatrix}'$, the matrix of indicator variables selecting at every point in time which member of the mixture of the normal approximation will be used for each element of e. Given $(y^{**})^T$ and h^T, each element $s_{i,t}$ of S^T is sampled from the discrete density defined by

$$Pr(s_{i,t} = j | y^{**}_{i,t}, h_{i,t}) \propto q_j f_N(y^{**}_{i,t} | 2h_{i,t} + m_j - 1.2704, v_j^2),$$

$j = 1, 2, \dots, 7$, where $f_N(\cdot|\mu, v^2)$ denotes the pdf for a normal random variable with mean μ and variance v^2. Conditional on B^T, A^T, V and s^T, the system has an approximate linear and Gaussian state space form, where each element $e_{i,t}$ of e_t in (17) can now be viewed as being distributed as normal with mean $m_j - 1.2704$ and variance v_j^2 if $s_{i,t} = j$.

For each $t = 1, 2, \dots, T$, let m_t denote the 3×1 vector consisting of the means $m_j - 1.2704$ of each element of e_t as determined above and let V_t denote the 3×3 matrix with the corresponding variances v_j^2 along its diagonal. Finally, define $x_t = y_t^{**} - m_t + 1.2704$. Now (19) can be rewritten as the observation equation

$$x_t = 2h_t + e_t, \tag{21}$$

where $e_t \sim N(0, V_t)$ and

$$h_t = h_{t-1} + \eta_t \tag{20}$$

remains as the state transition equation, with $\eta_t \sim N(0, W)$.

Given $h_{0|0} = \log \hat{\sigma}$ and $P_{0|0} = k_\sigma^2 I_3$, the Kalman filter implies

$$h_{t|t-1} = h_{t-1|t-1},$$

$$P_{t|t-1} = P_{t-1|t-1} + W,$$
$$K_t = 2P_{t|t-1}(4P_{t|t-1} + V_t)^{-1},$$
$$h_{t|t} = h_{t|t-1} + K_t(x_t - 2h_{t|t-1}),$$

and

$$P_{t|t} = P_{t|t-1} - 2K_t P_{t|t-1}.$$

The last elements from these recursions are $h_{T|T}$ and $P_{T|T}$, which are the mean and variance of the normal distribution used to make a draw for h_T. The draw for h_T and the output of the filter can now be used for the first step of the backward recursions

$$h_{t|t+1} = h_{t|t} + P_{t|t}P^{-1}_{t+1|t}(h_{t+1} - h_{t|t}) = h_{t|t} + P_{t|t}(P_{t|t} + W)^{-1}(h_{t+1} - h_{t|t})$$

and

$$P_{t|t+1} = P_{t|t} - P_{t|t}P^{-1}_{t+1|t}P_{t|t} = P_{t|t} - P_{t|t}(P_{t|t} + W)^{-1}P_{t|t},$$

which are the means and variances used to make the draws for h_t, $t = T-1, T-2, \dots, 1$.

Del Negro and Primiceri (2015*b*) note that, strictly speaking, because the mixture-of-normals distribution used in the Kim-Shephard-Chib algorithm is only an approximation to the true distribution of the innovations in the measurement equation (19), each draw selected using this algorithm should be used as a proposal in a Metropolis-Hastings step, following the general analysis in Stroud, Muller, and Polson (2003). With y_t^* and $y_{i,t}^{**}$ defined as above, let $\tilde{\Sigma}_t$ and Σ_t^{old} be the latest and previous draws for the volatility state for period

$t = 1, 2, \ldots, T$, and let $\tilde{\sigma}_{i,t}$ and $\sigma_{i,t}^{old}$ be the ith diagonal elements of $\tilde{\Sigma}_t$ and Σ_t^{old}. Del Negro and Primiceri (2015a) show that in the Metropolis step, the new draw should be accepted with probability α, where

$$\alpha = \frac{\left[\prod_{t=1}^{T} F_N(y_t^*|0_{3,1}, \tilde{\Sigma}_t\tilde{\Sigma}_t')\right]\left[\prod_{t=1}^{T}\prod_{i=1}^{3}\prod_{j=1}^{7} q_j f_N(y_{i,t}^{**}|2\sigma_{i,t}^{old} + m_j - 1.2704, v_j^2)\right]}{\left[\prod_{t=1}^{T} F_N(y_t^*|0_{3,1}, \Sigma_t^{old}(\Sigma_t^{old})')\right]\left[\prod_{t=1}^{T}\prod_{i=1}^{3}\prod_{j=1}^{7} q_j f_N(y_{i,t}^{**}|2\tilde{\sigma}_{i,t} + m_j - 1.2704, v_j^2)\right]},$$

and $F_N(\cdot|\mu, V)$ is the pdf for the multivariate normal distribution with mean μ and covariance matrix V.

4.4 Drawing Hyperparameters

The hyperparameters are the diagonal blocks of V, each of which has an inverse-Wishart posterior distribution. Conditional on B^T, α^T, σ^T, and y^T, it is easy to draw from these posteriors because the innovations are observable. Use (7) to compute

$$\nu_t = B_t - B_{t-1},$$

use (15) to compute

$$\zeta_{1,t} = \alpha_{u\pi,t} - \alpha_{u\pi,t-1},$$

use (17) to compute

$$\zeta_{2,t} = \alpha_{2,t} - \alpha_{2,t-1},$$

and use (9) to compute

$$\eta_t = \log\sigma_t - \log\sigma_{t-1}.$$

Then a new draw for Q can be taken from the inverse-Wishart posterior distribution with scale matrix

$$d_Q k_Q^2 V_Q + \sum_{t=1}^{T} \nu_t\nu_t',$$

and degrees of freedom $d_Q + T$, a new draw for S_1 can be taken from the inverse-Wishart posterior distribution with scale matrix

$$d_{S1} k_S^2 V_{S1} + \sum_{t=1}^{T} \zeta_{1,t}\zeta_{1,t}',$$

and degrees of freedom $d_{S1} + T$, a new draw for S_2 can be taken from the inverse-Wishart posterior distribution with scale matrix

$$d_{S2} k_S^2 V_{S2} + \sum_{t=1}^{T} \zeta_{2,t}\zeta_{2,t}',$$

and degrees of freedom $d_{S2} + T$, and new draws for each diagonal element of W can be taken from the inverse-Wishart posterior distributions with scale matrix

$$d_W k_W^2 + \sum_{t=1}^{T} \eta_t\eta_t',$$

which in this case is a scalar, and degrees of freedom $d_W + T$.

4.5 Assessing Convergence

To assess the convergence of the MCMC algorithm, Primiceri (2005) recommends initializing the chain from different, randomly selected starting points, to verify that none of the results is affected. A related but slightly more formal approach is suggested by Geweke (1992). For any model statistic θ, which may be an element of B^T, A^T, Σ^T, V, or any function of these parameters, calculate the means $\bar{\theta}_A$ and $\bar{\theta}_B$ from two disjoint subsamples of the Gibbs sampling output: Geweke suggests letting subsample A be formed from the first 10 percent of the draws and subsample B from the last 50 percent of the draws. The numerical standard errors of the means $\bar{\theta}_A$ and $\bar{\theta}_B$ are given by

$$\left(\frac{1}{N_A}\right)[2\pi S_{\theta,A}(0)] \text{ and } \left(\frac{1}{N_B}\right)[2\pi S_{\theta,B}(0)],$$

where $S_{\theta,A}(0)$ and $S_{\theta,B}(0)$ denote the spectral densities of $\hat{\theta}_A$ and $\hat{\theta}_B$ at frequency zero, which can be estimated using Newey and West's (1987) Bartlett weighting scheme as

$$S_{\theta,A}(0) = \frac{1}{2\pi}\left[\upsilon_{\theta,A,0} + 2\sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\upsilon_{\theta,A,j}\right]$$

and

$$S_{\theta,B}(0) = \frac{1}{2\pi}\left[\upsilon_{\theta,B,0} + 2\sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\upsilon_{\theta,B,j}\right],$$

where $\upsilon_{\theta,A,j}$ and $\upsilon_{\theta,B,j}$ are the jth autocovariances of the draws for θ in subsamples A and B. Geweke's convergence diagnostic

$$CD(\theta) = \frac{\bar{\theta}_A - \bar{\theta}_B}{\{N_A^{-1}[2\pi S_{\theta,A}(0)] + N_B^{-1}[2\pi S_{\theta,B}(0)]\}^{1/2}} \Rightarrow N(0,1),$$

which, as shown, has the standard normal distribution as $N_A \to \infty$ and $N_B \to \infty$.

To gauge the extent to which the chain mixes, Primiceri (2005) and Benati (2011) compute inefficiency factors, described in more detail by Chib (1992, pp.3579-3580). The inefficiency factor for any individual statistic θ, which may again be an element of B^T, A^T, Σ^T, V, or any function of these parameters, is defined as the inverse of Geweke's (1992) measure of relative numerical efficiency:

$$IF(\theta) = \frac{2\pi S_\theta(0)}{\mathrm{Var}(\theta)} = \frac{2\pi S_\theta(0)}{\int_{-\pi}^{\pi} S_\theta(\omega)\mathrm{d}\omega},$$

where $S_\theta(\omega)$ is the spectral density of θ at frequency ω so that, in particular, $S_\theta(0)$ is the spectral density of θ at frequency zero. Primiceri notes that

$$IF(\theta) = 1 + 2\sum_{j=1}^{\infty}\rho_{\theta,j},$$

where $\rho_{\theta,k}$ is the j-th autocorrelations of the draws for θ. Hence, $IF(\theta)$ will generally be larger than one, and lower values of $IF(\theta)$ reflect less autocorrelation in the draws. In computing $IF(\theta)$, Newey and West's (1987) estimator

$$S_\theta(0) = \frac{1}{2\pi}\left[\upsilon_{\theta,0} + 2\sum_{j=1}^{m}\left(1 - \frac{j}{m+1}\right)\upsilon_{\theta,j}\right]$$

can be used for the numerator, while the denominator is simply the variance $\upsilon_{\theta,0}$ across all draws for θ.

5 Identification of Monetary Policy Shocks

5.1 The Identification Problem

Two approaches can be taken to identify monetary policy shocks from the estimated reduced form. The first uses assumptions about the timing with which monetary policy disturbances affect inflation and the gap variable to re-interpret the triangular factorization of the reduced-form covariance matrix shown in (2) as a mapping between the reduced-form and structural models – an approach that dates back to Sims (1980). The second uses sign restrictions to identify monetary policy shocks based on their implied impulse responses. Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005) propose and develop the idea that sign restrictions can serve a source of identifying assumptions in VARs, and Benati (2011) implements the particular scheme used here in a similar VAR framework with time-varying parameters.

Details on each of the two identification strategies follows, but each works to factor the reduced-form covariance matrix as

$$\Omega_t = C_t^{-1} D_t D_t' (C_t')^{-1}, \tag{22}$$

where C_t and D_t are 3×3 matrices of the form

$$C_t = \begin{bmatrix} 1 & -c_{\pi g,t} & -c_{\pi r,t} \\ -c_{g\pi,t} & 1 & -c_{gr,t} \\ -c_{r\pi,t} & -c_{rg,t} & 1 \end{bmatrix} \tag{23}$$

and

$$D_t = \begin{bmatrix} \delta_{\pi,t} & 0 & 0 \\ 0 & \delta_{g,t} & 0 \\ 0 & 0 & \delta_{r,t} \end{bmatrix} \tag{24}$$

Equations (22)-(24) provide the general mapping between the reduced-form (1) and the structural model, which can now be written as

$$C_t y_t = \gamma_t + \Gamma_{1,t} y_{t-1} + \Gamma_{2,t} y_{t-2} + D_t \xi_t, \tag{25}$$

where $\gamma_t = C_t b_t$, $\Gamma_{j,t} = C_t B_{j,t}$ for $j = 1, 2$, and ξ_t is a 3×1 vector of structural disturbances, normally distributed with zero mean and $E\xi_t\xi_t' = I_3$. The third row from (25) takes the

form of a monetary policy rule

$$\begin{aligned} R_t = \gamma_{r,t} + c_{r\pi,t}\Pi_t + \gamma_{1,r\pi,t}\Pi_{t-1} + \gamma_{2,r\pi,t}\Pi_{t-2} \\ + c_{rg,t}G_t + \gamma_{1,rg,t}G_{t-1} + \gamma_{2,rg,t}G_{t-2} + \gamma_{1,rr,t}R_{t-1} + \gamma_{2,rg,t}R_{t-2} + \delta_{r,t}\xi_t^{mp}, \end{aligned} \tag{26}$$

where the constant term and the coefficients on lagged values of inflation, the gap variable, and the interest rate are those from the third rows of γ_t, $\Gamma_{1,t}$ and $\Gamma_{2,t}$, and ξ_t^{mp} represents the identified monetary policy shock. This policy rule takes the same general form as Taylor's (1993), in that it prescribes a setting for the short-term interest rate with reference to the changing values of inflation and the gap variable. However, (26) allows for considerably flexibility in the dynamic response of the short-term interest rate to changes in inflation and the gap variable and, through the inclusion of lagged interest rate terms on the right-hand side, captures as well the Federal Reserve's tendency to smooth interest rate movements over time. Deviations in the actual short-term interest rate away from the value dictated by the current and lagged values of inflation, the gap variable, and the interest rate get picked up as monetary policy shocks in (26). Finally, (26) allows for time-variation in all of the response coefficients and in the standard deviation $\delta_{r,t}$ of the monetary policy shocks.

Comparing (3) and (4) to (23) and (24) highlights the identification problem: together, the matrices A_t and Σ_t of reduced-form parameters contain 6 elements not equal to zero or one, whereas the matrices C_t and D_t of structural parameters have 9 such elements. Each of the two identification schemes described next imposes more structure on the matrix C_t to solve this problem.

5.2 Triangular Identification Based on Timing Assumptions

The factorization of the symmetric, positive definite reduced-form covariance matrix Ω_t shown in (2)-(4) always exists and is unique; hence, the model can be written in this form without any loss of generality. However, under the additional assumptions – made throughout much of the literature on VARs that builds on Sims (1980) – that inflation and the output gap respond to monetary policy shocks only after a one-period lag, the reduced-form parameters from the third rows of (2)-(4) are linked to structural parameters from the third rows on (22)-(24) via

$$c_{r\pi,t} = -\alpha_{r\pi,t},$$

$$c_{rg,t} = -\alpha_{rg,t},$$

and

$$\delta_{r,t} = \sigma_{r,t}$$

and the structural monetary policy shock ξ_t^{mp} from (25) and (26) is identified as the third element of the vector ε_t from (5).

5.3 Sign Restrictions for the Variables that Respond to Monetary Policy

An alternative approach to identification builds on work by Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005) by associating monetary policy shocks with the effects they

have on observable variables. Following Benati (2011), suppose that the first element of ξ_t corresponds to a supply shock that moves inflation and the output gap in opposite directions or inflation and the unemployment rate in the same direction. Suppose that the second element of ξ_t is a non-monetary demand shock, that moves the short-term interest rate and inflation in the same direction and the interest rate and the output gap in the same direction or the interest rate and the unemployment rate in opposite directions. Finally, suppose that the third element of ξ_t corresponds to a monetary policy shock that moves the short-term interest rate and inflation in opposite directions and the interest rate and the output gap in opposite directions or the interest rate and the unemployment rate in the same direction. Rubio-Ramirez, Waggoner, and Zha (2010) and Arias, Rubio-Ramirez, and Waggoner (2014) emphasize that sign restrictions of this form do not suffice to identify structural disturbances in the classical sense, but develop a Bayesian algorithm for characterizing the set of parameter values implying impulse responses that satisfy these restrictions.

Let the index $i = 1, 2, \ldots, N$ keep track of the number of desired draws. For $i = 1, 2, \ldots, N$, the algorithm loops through the following steps.

1. Draw (A^T, Σ^T) from their posterior distribution during the Gibbs sampling stage.

2. For each $t = 1, 2, \ldots, T$, construct A_t and Σ_t based on the draw for (A^T, Σ^T). Then let $L_t = A_t^{-1}\Sigma_t$, so that the reduced-form error covariance matrix is given by $\Omega_t = L_t L_t'$.

3. Draw $\tilde{X}$, a 3×3 random matrix with each element having an independent standard normal distribution. Then factor $\tilde{X} = Q_X R_X$, where Q_X is an orthogonal matrix and R_X is upper triangular with positive diagonal elements.

4. Let $\tilde{L}_t = L_t Q_X'$, and note that

$$\tilde{L}_t \tilde{L}_t' = L_t Q_X' Q_X L_t' = L_t L_t' = \Omega_t,$$

by virtue of the fact that Q_X is orthogonal. This highlights that multiplying the structural model (25) through by D_t^{-1} and then Q_X results in an observationally-equivalent rotation of the model's three equations. Suppressing for convenience explicit reference to the constant and lagged terms in (25), the candidate structural model based on the specific draw for Q_X can be written as

$$y_t = \tilde{L}_t \xi_t,$$

since

$$E[(\tilde{L}_t \xi_t)(\tilde{L}_t \xi_t)'] = E(\tilde{L}_t \xi_t \xi_t' \tilde{L}_t') = \tilde{L}_t E(\xi_t \xi_t') \tilde{L}_t' = \tilde{L}_t \tilde{L}_t' = \Omega_t.$$

Thus, the matrix $\tilde{L}_t$ contains impact coefficients linking the structural shocks in ξ_t to the observable variables in y_t. The sign restriction used to identify the supply, demand, and monetary policy shocks as the first, second, and third elements of ξ_t require the elements of $\tilde{L}_t$ to have the sign patterns

$$\tilde{L}_t = \begin{bmatrix} (+) & (+) & (-) \\ (-) & (+) & (-) \\ (?) & (+) & (+) \end{bmatrix}$$

if the gap variable is measured by the output gap and

$$\tilde{L}_t = \begin{bmatrix} (+) & (-) & (-) \\ (+) & (+) & (+) \\ (?) & (-) & (+) \end{bmatrix}$$

if the gap variable is measured by the unemployment rate. If these restrictions are not satisfied for any $t = 1, 2, \ldots, T$, the draws for (A^T, Σ^T) and $\tilde{X}$ are discarded and the algorithm returns to step one. If the restrictions are satisfied, then $\tilde{L}_t$ is renormalized as $\tilde{L}_t = C_t^{-1} D_t$, where C_t and D_t have the forms shown in (23) and (24), these draws are saved, and the Gibbs sampling algorithm moves on.

5.4 Other Statistics

Once draws are obtained for the structural parameters using one of three identification schemes, impulse responses can be generated from (25) after multiplying through by C_t^{-1}. These computations can be simplified by writing the system in companion form as

$$Y_t - \bar{\mu}_t + B_{12,t}(Y_{t-1} - \bar{\mu}_t) + F_t \xi_t, \tag{27}$$

where

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix},$$

$$B_{12,t} = \begin{bmatrix} B_{1,t} & B_{2,t} \\ I_3 & 0_{3,3} \end{bmatrix}, \tag{28}$$

$$\bar{\mu}_t = (I_6 - B_{12,t})^{-1} \begin{bmatrix} b_t \\ 0_{3,1} \end{bmatrix}, \tag{29}$$

and

$$F_t = \begin{bmatrix} C_t^{-1} D_t \\ 0_{3,3} \end{bmatrix}. \tag{30}$$

Following Cogley and Sargent (2005), the first element of the vector $\bar{\mu}_t$ defined in (29) can also be used as a measure of core, or target, inflation and, in the case where the gap variable is measured by the unemployment rate, the second element of $\bar{\mu}_t$ can be interpreted as an estimate of the natural rate of unemployment.

Since (27) implies

$$Y_{t+k} - E_t Y_{t+k} = F_t \xi_{t+k} + B_{12,t} F_t \xi_{t+k-1} + \ldots + B_{12,t}^{k-1} F_t \xi_{t+1},$$

the k-step ahead forecast error variances for the elements of Y_t are

$$E[(Y_{t+k} - E_t Y_{t+k})(Y_{t+k} - E_t Y_{t+k})'] = F_t F_t' + B_{12,t} F_t F_t' B_{12,t}' + \ldots + B_{12,t}^{k-1} F_t F_t' (B_{12,t}^{k-1})'. \tag{31}$$

Forecast error variances decompositions can be found by using (28), (30), and (31) to compute the total variances and then by using these same equations with the first two diagonal elements of D_t set equal to zero to find the variances attributable to monetary policy shocks alone.

Finally, it may be helpful to consider the "artificial" long-run responses, as defined by Sims and Zha (2006, p.62), of the interest rate to changes in inflation and the gap variable. From (26), these are

$$\gamma_{r\pi,t} = \frac{c_{r\pi,t} + \gamma_{1,r\pi,t} + \gamma_{2,r\pi,t}}{1 - \gamma_{1,rr,t} - \gamma_{2,rr,t}}$$

for inflation and

$$\gamma_{rg,t} = \frac{c_{rg,t} + \gamma_{1,rg,t} + \gamma_{2,rg,t}}{1 - \gamma_{1,rr,t} - \gamma_{2,rr,t}}.$$

for the gap variable.

6 References

Arias, Jonas E., Juan F. Rubio-Ramirez, and Daniel F. Waggoner. "Inference Based on SVARs Identified with Sign and Zero Restrictions: Theory and Applications." Manuscript. Durham: Duke University, September 2014.

Benati, Luca. "Would the Bundesbank Have Prevented the Great Inflation in the United States?" *Journal of Economic Dynamics and Control* 35 July (2011): 1106-1125.

Canova, Fabio and Gianni De Nicolo. "Monetary Disturbances Matter for Business Fluctuations in the G-7" *Journal of Monetary Economics* 49 (September 2002): 1131-1159.

Carter, C.K. and R. Kohn. "On Gibbs Sampling for State Space Models." *Biometrika* 81 (August 1994): 541-553.

Chib, Siddhartha. "Markov Chain Monte Carlo Methods: Computation and Inference." In *Handbook of Econometrics*, Volume 5, edited by James J. Heckman and Edward Leamer. Amsterdam: Elsevier, 1992, pp.3569-3649.

Cogley, Timothy and Thomas J. Sargent. "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US." *Review of Economic Dynamics* 8 (April 2005): 262-302.

Del Negro, Marco and Giorgio E. Primiceri. "Time Varying Structural Vector Autoregressions and Monetary Policy: Appendix to the Corrigendum." Manuscript. Evanston: Northwestern University, February 2015*a*.

Del Negro, Marco and Giorgio E. Primiceri. "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum." *Review of Economic Studies* 82 (October 2015*b*): 1342-1345.

Faust, Jon. "The Robustness of Identified VAR Conclusions About Money." *Carnegie-Rochester Conference Series on Public Policy* 49 (December 1998): 207-244.

Fruhwirth-Schnatter, Sylvia. "Data Augmentation and Dynamic Linear Models." *Journal of Time Series Analysis* 15 (March 1994): 183-202.

Geweke, John. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics 4*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith. Oxford: Oxford University Press, 1992, pp.169-193.

Hamilton, James. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib. "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies* 65 (July 1998): 361-393.

Lutkepohl, Helmut. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag, 2006.

Newey, Whitney K. and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (May 1987): 703-708.

Primiceri, Giorgio E. "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* 72 (July 2005): 821-852.

Rubio-Ramirez, Juan F., Daniel F. Waggoner, and Tao Zha. "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference." *Review of Economic Studies* 77 (April 2010): 665-696.

Sims, Christopher A. "Macroeconomics and Reality." *Econometrica* 48 (January 1980): 1-48.

Sims, Christopher A. and Tao Zha. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96 (March 2006): 54-81.

Stroud, Jonathan R., Peter Muller, and Nicholas G. Polson. "Nonlinear State-Space Models With State-Dependent Variances." *Journal of the American Statistical Association* 98 (June 2003): 377-386.

Taylor, John B. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39 (December 1993): 195-214.

Uhlig, Harald. "What Are the Effects of Monetary Policy On Output? Results From an Agnostic Identification Procedure." *Journal of Monetary Economics* 52 (March 2005): 381-419.