Macroeconometrics

Lecture 6

TVP-VAR

Reduce form VAR

- Given a N multivariate series $\mathbf{y}_t = (y_{1,t}, \cdots, y_{N,t})'$, a Vector Autoregressive process is.

$$\mathbf{y}_t = A(L)\mathbf{y}_t + \mathbf{u}_t \quad \mathbf{u}_t \sim i.i.d.f(0, \Omega), \; t = 1, \cdots, T \tag{1}$$

  The function $f()$ is usually specified as a normal distribution. VAR(p) in less compact form:

$$\mathbf{y}_t = A_0 + \sum_{i=1}^{p} A_i \mathbf{y}_{t-i} + \mathbf{u}_t \quad \mathbf{u}_t \sim i.i.d.f(0, \Omega)$$

- Assume $N = 2$, the model assumes that $y_{t,1}$ and $y_{t,2}$ move jointly following a linear specification, which depends on previous L lags and the residuals $\mathbf{u}_t$.
- The model can be easily estimated by using frequentist inference (MLE-SUR) or Bayesian inference.

VAR: Bayesian estimation

- **Priors:** Normal-Whishart (diffuse) prior.

$$A \sim MN(\mathbf{0}, \underline{\Sigma})$$

$$\Omega \sim IW(\underline{\nu}, \underline{S})$$

- The likelihood is:

$$L(A, \Omega) = |\Omega|^{T/2} \exp\left[-\frac{1}{2}(\alpha - \hat{\alpha})^{'}(\Omega \otimes (X^{'}X)^{-1})(\alpha - \hat{\alpha})\right]$$

$$\exp\left[-\frac{1}{2} tr(\Omega^{-1} S)\right]$$

where $S = (Y - XA)^{'}(Y - XA)$, $\alpha = vec(A)$, $\hat{\alpha} = vec(\hat{A})$, $\hat{A} = (X^{'}X)^{-1})X^{'}Y$, $\otimes$ is the Kroeneker product and

$$X = \begin{bmatrix} 1 & \mathbf{y}_0' & \dots & \mathbf{y}_{-p}' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{y}_{T-1}' & \dots & \mathbf{y}_{T-p}' \end{bmatrix}.$$

VAR: Bayesian estimation

- Under the Normal-diffuse prior, we estimate the VAR(p) model with a two-step Gibbs sampler, detailed in such studies as Kadiyala and Karlsson (1997).
  - Step 1: Draw the vector of VAR coefficients A conditional on the error variance-covariance matrix Ω. We draw the VAR coefficients from a conditional posterior distribution that is multi-variate normal, as in equation (16a) of Kadiyala and Karlsson (1997).
  - Step 2: Draw the error covariance matrix Ω conditional on the VAR coefficients A.
- We draw the error variance matrix from a conditional posterior distribution that is inverse Wishart, as in equation (16b) of Kadiyala and Karlsson (1997).

Bayesian VAR

- In-sample fit (tuning)
  The applied researcher typically wishes to have a model with plausible in-sample fit, i.e. able to replicate important features of the data (sd, means, cross- and auto-correlations)
  VARs have very good fit. No need of 'fine tuning'
  DSGE needs to be 'tuned', by choosing and appropriately transforming the observable variables, and by tailoring the priors on parameters when necessary.
- Out-of-sample performance (tuning)
  In many situations we are more interested in how the model fares in predicting future path of the endogenous variables
  VARs need to be 'tuned' in order to obtain good forecasts

Bayesian VAR

- Forecasts with empirical and structural models
  In this short lecture we describe how to produce unconditional forecasts with an industry standard DSGE (structural) model, with a Bayesian VAR (empirical) model and with an DSGE-VAR model (hybrid).
- We characterize the predictive density of the data, compute point forecasts as well as the uncertainty surrounding them.
- We compare the forecast output during the Great Recession using the SW(2007) model and an empirical multivariate autoregressive model (BVAR)

Book and References

- Geweke, J. (2005), Contemporary Bayesian Econometrics and Statistics, Wiley.
- Geweke, J. and Whiteman (2006), Bayesian Forecasting, in in Handbook of Economic Forecasting, ed. by G. Elliott, C. Granger, and A. Timmermann, vol. 1 of Handbooks in Economics 24, pp. 3 to 80. North Holland, Amsterdam.
- Del Negro, M., F. Schorfheide (2012): DSGE Model-Based Forecasting, In: Graham Elliott and Allan Timmermann (eds.): Handbook of Economic Forecasting, Vol 2A, 2013, Handbooks in Economics, Elsevier / North-Holland, 57-140.
- Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): On the Fit of New Keynesian Models, Journal of Business and Economic Statistics, 25(2), 123 162.
- Del Negro, M., F. Schorfheide (2012): Bayesian Macroeconometrics, In J. Geweke, G. Koop, and H. van Dijk (eds.) "The Oxford Handbook of Bayesian Econometrics," 2011, Oxford University Press, 293-389.
- Villemot, S. (2012): BVAR models a la Sims in Dynare, Dynare Documents

Some background notions of theory

- Recall the posterior distribution of the parameters

$$p(\theta|y_{1:T}) = \frac{p(y_{1:T}|\theta)p(\theta)}{p(y_{1:T})}, \quad p(y_{1:T}) = \int p(y_{1:T}|\theta)p(\theta)d\theta$$

- For forecasting purposes we are less interested in the posterior distribution of the parameters rather in the **predictive density**, i.e.

$$p(y_{T+1:T+h}|y_{1:T}) = \int p(y_{T+1:T+h}|y_{1:T};\theta)p(\theta|y_{1:T})d\theta$$

Some background notions of theory

- This decomposition highlights that draws from the predictive density can be obtained by simulating the structural or empirical model conditional on posterior parameter draws and the observations $y_{1:T}$.
- Suppose you have $\{\theta^i\}_1^N$ draws from the posterior distribution, then one could generate a sequence $y^{(i)}_{T+1:T+h}$ for $j = 1, ..., N$ that represent draw from the predictive distribution. These draws can then be used to obtain numerical approximations of moments, quantiles or probability density functions.

(B)VAR models(1)

- A vector of autoregression with p lags can be expressed as

$$y_t = \Psi_1 y_{t-1} + ... + \Psi_p y_{t-p} + \Psi_0 + u_t$$

  where y_t is $n \times 1$ vector, Ψ_j suitable matrices and u_t are i.i.d. zero mean shocks with covariance matrix Σ. Lets collect the vector of parameters in $\vartheta = vec(\Psi_1, ..., \Psi_p, \Psi_0, \Sigma)$.
- In Bayesian statistic we combine prior information and the likelihood of the data. As for the latter we need distributional assumption regarding the error term. Widespread assumption: normality

(B)VAR models(2)

- As for the former, the choice of the prior is crucial for forecasting performance. Non-informative priors tend to maximize the in-sample fit (overfitting) at the cost of poor out-of-sample performances.
- People use Minnesota prior (introduced by Litterman (1980) and Doan, Litterman and Sims (1984)) to mitigate this problem. The underlying logic of this prior is that individual time series can be represented a priori as random walks. This specification is selected because univariate unit root models are typically good at forecasting macroeconomic time series.
- Notice that the random walk hypothesis is imposed a priori: a posteriori, each time series will follow a more complicated process if there is sufficient information in the data.

BVAR Jeffrey Prior

- Let $k = np + 1$ and define the $k \times n$ matrix $\Psi = [\Psi_1, ..., \Psi_p, \Psi_0]'$. The conditional likelihood function can be conveniently expressed if the VAR is written as a multivariate linear regression model in matrix notation:

$$Y = X\Psi + U$$

where

$$Y = \begin{pmatrix} y_1' \\ \vdots \\ y_T' \end{pmatrix} \quad X = \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix} \quad x_t' = [y_1', ..., y_T', 1] \quad Y = \begin{pmatrix} u_1' \\ \vdots \\ u_T' \end{pmatrix}$$

BVAR Jeffrey Prior

- The conditional likelihood function can be conveniently expressed (see Zellner (1971))

$$p(Y|X,\Psi,\Sigma) \propto |\Sigma|^{-T/2}$$
$$\exp\left(-1/2tr(\Sigma^{-1}\widehat{S})\right) \times \exp\left(-1/2tr(\Sigma^{-1}(\Psi - \widehat{\Psi})'X'X(\Psi\right.$$
$$= p(\Sigma|Y,X)p(\Psi|\Sigma,Y,X)$$

where $\widehat{\Psi} = (X'X)^{-1}X'Y$ and $\widehat{S} = (Y - X\widehat{\Psi})'(Y - X\widehat{\Psi})$

BVAR Jeffrey Prior

- If we combine the the likelihood with the Jeffery prior $p(\Psi, \Sigma) = |\Sigma|^{-(n+1)/2}$, we obtain the familiy posterior distribution for, $\Psi, \Sigma | Y \sim MNIW\left(\widehat{\Psi}, (X'X)^{-1}, \widehat{S}, T-k\right)$, i.e.

$$p(\Psi, \Sigma | Y, X) = IW(\widehat{S}, T-k) \times N(\widehat{\Psi}, \Sigma \otimes (X'X)^{-1})$$

  Draws from this distribution can be obtained by:

  1. draw Σ^j from an $IW(\widehat{S}, T-k)$
  2. conditional on Σ^j draw Ψ^j from $N(\widehat{\Psi}, \Sigma^j \otimes (X'X)^{-1})$

BVAR with dummy priors(1)

- A popular way to introduce priors is through dummy observations
- Suppose T^* dummy observations are collected in matrices Y^* and X^*, and we use the likelihood function associated with the VAR to relate the dummy observations to the parameters (Ψ, Σ)
- Using the same arguments has before, we deduce that up to a constant the product $p(Y^*|\Psi, \Sigma) * |\Sigma|^{-(n+1)/2}$ can be interpreted as

$$\Psi, \Sigma | Y^* \sim MNIW\left(\underline{\Psi}, (X^{*'} X^*)^{-1}, \underline{S}, T^* - k\right)$$

same as before but use Y^* and X^*. Provided that $T^* > k + n$, $X^{*'} X^*$ is inventible and the prior distribution is proper.

BVAR with dummy priors(2)

- Let $\overline{T} = T + T^*$, $\overline{Y} = [Y^{*'}, Y']'$ and $\overline{X} = [X^{*'}, X']'$ and $\overline{\Psi}, \overline{\Sigma}$ the analog of $\widehat{\Psi}, \widehat{\Sigma}$
- we obtain the posterior distribution for

$$\Psi, \Sigma | \overline{Y} \sim MNIW\left(\overline{\Psi}, (\overline{X}'\overline{X})^{-1}, \overline{S}, \overline{T} - k\right)$$

- Hence, dummy observations lead to a conjugate priors, i.e. the prior distribution family coincides with the posterior one.

BVAR with Minnesota priors

- The flat Jeffrey prior allows us to produce forecast, but with poor prediction quality. We wish to introduce the notion that individual TS are a priori (close to) random walk.
- While it is fairly straightforward to choose prior means and variances for the elements of Ψ, it tends to be difficult to elicit beliefs about the correlation between elements of the Ψ matrix. After all, there are $nk(nk+1)/2$ of them.
- At the same time, setting all these correlations to zero potentially leads to a prior that assigns a lot of probability mass to parameter combinations that imply quite unreasonable dynamics for the endogenous variables.

BVAR with Minnesota priors

- The Minnesota prior offers a parsimonious way of introducing plausible correlations between parameters. Prior distribution is a function of a small vector of hyperparameters
  1. τ: the overall tightness of the prior. Large values imply a small prior covariance matrix.
  2. d: the decay factor for scaling down the coefficients of lagged values.
  3. ω controls the tightness for the prior on Σ. Must be an integer.
  4. λ and μ: additional tuning parameters, cross correlations and constants

BVAR with Minnesota priors

- The Minnesota prior for Ψ and Σ belongs to the MN-IW family. Conjugate priors.
- Prior for Ψ, first moments

$$E(\psi_{ij,k} \mid \Sigma) = \begin{cases} 1 & \text{if } j = i \text{ and } k = 1; \\ 0 & \text{else.} \end{cases}$$

- Prior for Ψ, second moments for the first lag, Ψ_1

$$cov(\psi_{ij,1}, \psi_{hg,1} \mid \Sigma) = \begin{cases} \Sigma_{ih}/(\tau \underline{s_j})^2 & \text{if } g = j; \\ 0 & \text{else.} \end{cases}$$

- Prior for Ψ, second moments for distant lags, Ψ_ℓ with $\ell = 2, ..., p$

$$cov(\psi_{ij,\ell}, \psi_{hg,\ell} \mid \Sigma) = \begin{cases} \Sigma_{ih}/(\tau \underline{s_j} 2^d)^2 & \text{if } g = j; \\ 0 & \text{else.} \end{cases}$$

- The sum of coefficient and the co-persistence are introduced though dummy observations

BVAR with Minnesota dummy priors(1)

- Let $y_{-\tau:0}$ be a presample, and let $\underline{y}$ and $\underline{s}$ be the $n \times 1$ vectors of means and standard deviations.
- To simplify the exposition, suppose that $n = 2$ and $p = 2$. The dummy observations are interpreted as observations from the regression model. We begin with dummy observations that generate a prior distribution for Ψ_1.

$$\begin{pmatrix} \tau \underline{s_1} & 0 \\ 0 & \tau \underline{s_2} \end{pmatrix} = \begin{pmatrix} \tau \underline{s_1} & 0 & 0 & 0 & 0 \\ 0 & \tau \underline{s_2} & 0 & 0 & 0 \end{pmatrix} \Psi + \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

BVAR with Minnesota dummy priors(2)

- Prior for Ψ_1
  The first row implies $\tau \underline{s_1} = \tau \underline{s_1} \psi_{11,1} + u_{11}$ and $0 = \tau \underline{s_1} \psi_{21,1} + u_{12}$, which given the normality assumption of u implies unit root behavior, i.e.

$$
\begin{aligned}
\psi_{11,1} &\sim N(1, \Sigma_{11}/(\tau^2 \underline{s_1}^2)) \\
\psi_{21,1} &\sim N(0, \Sigma_{22}/(\tau^2 \underline{s_1}^2))
\end{aligned}
$$

  with covariation

$$
\begin{aligned}
E(\psi_{11,1}\psi_{21,1}) &= E((1 - u_{11}/(\tau \underline{s_1}))(-u_{12}/(\tau \underline{s_1})) = \Sigma_{12}/(\tau^2 \underline{s_1}^2) \\
E(\psi_{11,1}\psi_{22,1}) &= E(\psi_{11,1}\psi_{12,1}) = 0
\end{aligned}
$$

  and similarly for $\psi_{12,1}$ and $\psi_{22,1}$.

BVAR with Minnesota dummy priors(3)

- Prior for Ψ_2

  The prior distribution for Ψ_2, where d scales the coefficients associated to the lags

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \tau \underline{s_1} 2^d & 0 & 0 \\ 0 & 0 & 0 & \tau \underline{s_2} 2^d & 0 \end{pmatrix} \Psi + \begin{pmatrix} u_{31} & u_{32} \\ u_{41} & u_{42} \end{pmatrix}$$

  This implies

$$\begin{aligned} \psi_{11,2} &\sim N(0, \Sigma_{11}/(\tau \underline{s_1} 2^d)^2)) \\ \psi_{21,2} &\sim N(0, \Sigma_{22}/(\tau \underline{s_1} 2^d)^2)) \\ E(\psi_{11,2}\psi_{21,2}) &= E((-u_{31}/(\tau \underline{s_1} 2^d))(-u_{32}/(\tau \underline{s_1} 2^d))) = \Sigma_{12}/(\tau \underline{s_1} 2^d)^2 \end{aligned}$$

BVAR with Minnesota dummy priors(4)

- *Sums-of-coefficients* dummy

$$\begin{pmatrix} \lambda \underline{y_1} & 0 \\ 0 & \lambda \underline{y_2} \end{pmatrix} = \begin{pmatrix} \lambda \underline{y_1} & 0 & \lambda \underline{y_1} & 0 & 0 \\ 0 & \lambda \underline{y_2} & 0 & \lambda \underline{y_2} & 0 \end{pmatrix} \Psi + U$$

$$\begin{aligned} (\psi_{11,1} + \phi_{11,2}) &\sim N(1, \Sigma_{11}/(\lambda \underline{y_1})^2) \\ (\psi_{21,1} + \phi_{21,2}) &\sim N(0, \Sigma_{22}/(\lambda \underline{y_1})^2) \\ E((\psi_{11,1} + \phi_{11,2})(\psi_{21,1} + \phi_{21,2})) &= \Sigma_{12}/(\lambda \underline{y_1})^2 \end{aligned}$$

and similarly for the second variable.
This prior implies that when the lagged values of $y_{1,t}$ are at $\underline{y_1}$, then the value $\underline{y_1}$ is a good forecaster of $y_{1,t}$.

BVAR with Minnesota dummy priors(5)

- *Co-persistence* dummy

  The remaining sets of dummy observations provide a prior for the intercept

$$\begin{pmatrix} \mu \underline{y}_1 & \mu \underline{y}_2 \end{pmatrix} = \begin{pmatrix} \mu \underline{y}_1 & \mu \underline{y}_2 & \mu \underline{y}_1 & \mu \underline{y}_2 & \mu \end{pmatrix} \Psi + U$$

  which implies a sum of independent normals

$$\psi_{1,0} = \underline{y}_1 - \underline{y}_1(\psi_{11,1} + \phi_{11,2}) - \underline{y}_2(\psi_{12,1} + \phi_{12,2}) - 1/\mu u_{71}$$
$$\psi_{1,0} \sim N(0, \Sigma_{11}(\lambda^{-2} + \mu^{-2}) + \Sigma_{22}\lambda^{-2})$$

  The sum-of-coefficients prior is not consistent with cointegration, Sims (1993). The co-persistence states that a no-change forecast for all variables is a good forecast at the beginning of the sample

BVAR with Minnesota dummy priors(6)

- The prior distribution for Σ is centered at the matrix with elements equals to the pre-sample variance of y_t

$$\begin{pmatrix} \underline{s_1} & \underline{s_{12}} \\ \underline{s_{12}} & \underline{s_2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Psi + U$$

these observations are replicated ω times.

- Number of dummy observations

$$T^* = n \times p + n \times \omega + 1 + n = n(p + \omega + 1) + 1$$

BVAR forecasts

- A vector of autoregression with p lags can be expressed as

$$y_t = \Psi_1 y_{t-1} + ... + \Psi_p y_{t-p} + \Psi_0 + u_t$$

where y_t is $n \times 1$ vector, Ψ_j suitable matrices and u_t are i.i.d. zero mean normal shocks with covariance matrix Σ. Lets collect the vector of parameters in $\vartheta = vec(\Psi_0, ..., \Psi_p, \Sigma)$.

- Assuming Minnesota dummy prior we obtain the posterior distribution for

$$\Psi, \Sigma | Y \sim MNIW\left(\overline{\Psi}, (\overline{X}'\overline{X})^{-1}, \overline{S}, \overline{T} - k\right)$$

- **Algorithm 1**

  Given the posterior distribution of the parameters, $p(\vartheta | y_{1:T})$, one can generate trajectories simply by

  - draw $\vartheta^j = vec(\Psi_0^j, ..., \Psi_p^j, \Sigma^j)$ from $MNIW$
  - draws from u, i.e. $u^k_{T+1:T+h}$ iterate on the VAR representation, i.e.

$$y_\tau = \Psi_0^j + \Psi_1^j y_{\tau-1} + ... + \Psi_p^j y_{\tau-p} + u_\tau$$

These draws can then be used to obtain numerical approximations of moments, quantiles or probability density functions. For point forecasts, you need a loss function. The most widely-used: the quadratic forecast error loss function. Mean across all trajectories.

Homoskedastic TVP-VAR

- Time-varying parameter Vector autoregressive with homoskedastic errors are:

$$\mathbf{y}_t = X_t B_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma)$$
$$B_t = B_{t-1} + n_t, \quad n_t \sim N(0, Q)$$

  The VAR coefficients follow random walk processes, with innovations that are allowed to be correlated across coefficients.

- Priors?

Homoskedastic TVP-VAR: estimation step 1

- Draw the time series of the vector of VAR coefficients B_t conditional on the history of Σ and Q.
- As detailed in Primiceri (2005), drawing the VAR coefficients involves using the Kalman filter to move forward in time, a backward smoother to obtain posterior means and variances of the coefficients at each point in time, and then drawing coefficients from the posterior normal distribution.

TVPVAR-SV: estimation step 2

- Draw the elements of Σ conditional on the history of B_t and Q.
- Following Cogley and Sargent (2005) and Primiceri (2005), the sampling of Σ, the variance-covariance matrix of innovations to the VAR errors, is based on inverse Wishart priors and posteriors. The scale matrix of the posterior distribution is the sum of the prior mean $\times$ the prior degrees of freedom and $\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}'_t$, where $\hat{\epsilon}_t$ denotes the innovations to the posterior draws of coefficients obtained in step 1.

TVPVAR-SV: estimation step 3

- Draw the variance matrix Q conditional on the history of B_t and Σ.
- Following Cogley and Sargent (2005) and Primiceri (2005), the sampling of Q, the variance-covariance matrix of innovations to the VAR coefficients, is based on inverse Wishart priors and posteriors. The scale matrix of the posterior distribution is the sum of the prior mean $\times$ the prior degrees of freedom and $\sum_{t=1}^{T} \hat{n}_t \hat{n}_t'$, where $\hat{n}_t$ denotes the innovations to the posterior draws of coefficients obtained in step 1.

TVPVAR-SV

- Time-varying parameter Vector autoregressive with stochastic volatility (note change notation):

$$
\begin{aligned}
\mathbf{y}_t &= X_t B_t + v_t \\
B_t &= B_{t-1} + n_t, \ \text{var}(n_t) = Q \\
v_t &= A^{-1} \Lambda_t^{0.5} \epsilon_t, \ \epsilon_t \sim N(0, I_k), \ \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \ldots, \lambda_{k,t}) \quad (2) \\
\log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t}, \ i = 1, k \\
\nu_t &\equiv (\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{k,t})' \sim \ N(0, \Phi),
\end{aligned}
$$

  where $A =$ a lower triangular matrix with ones on the diagonal and non-zero coefficients below the diagonal. The VAR coefficients follow random walk processes, with innovations that are allowed to be correlated across coefficients.
- Priors?

TVPVAR-SV: estimation (1)

- Step 1: Draw the time series of the vector of VAR coefficients B_t conditional on the history of Λ_t, Q, A, and Φ. As detailed in Primiceri (2005), drawing the VAR coefficients involves using the Kalman filter to move forward in time, a backward smoother to obtain posterior means and variances of the coefficients at each point in time, and then drawing coefficients from the posterior normal distribution.

TVPVAR-SV: estimation (2)

- Step 2: Draw the elements of A conditional on the history of B_t, the history of Λ_t, Q, and Φ.
- Following Cogley and Sargent (2005), rewrite the VAR as

$$A(y_t - X_t' B_t) = A\hat{y}_t \equiv \tilde{y}_t = \Lambda_t^{0.5} \epsilon_t, \tag{3}$$

  where, conditional on B_t, $\hat{y}_t$ is observable. This system simplifies to a set of $i = 2, \ldots, k$ equations, with equation i having as dependent variable $\hat{y}_{i,t}$ and as independent variables $-1 \cdot \hat{y}_{j,t}, j = 1, \ldots, i-1$, with coefficients a_{ij}. Multiplying equation i by $\lambda_{i,t}^{-0.5}$ eliminates the heteroskedasticity associated with stochastic volatility.
- Then, proceeding separately for each transformed equation i, draw the i'th equation's vector of coefficients a_i (a vector containing a_{ij} for $j = 1, \ldots, i-1$) from a normal posterior distribution.

TVPVAR-SV: estimation(3)

- Step 3: Draw the elements of the states for the mixture distribution used to approximate the χ^2 distribution under the Kim, Shephard, and Chib (1998) algorithm, conditional on the history of B_t, A, the history of Λ_t, Q, and Φ. See Primiceri (2005) for details. However, we depart from Primiceri by using a 10 state approximation of the χ^2 distribution from Omori, et al. (2007) instead of the 7-state approximation from Kim, Shephard, and Chib (1998).

TVPVAR-SV: estimation(4)

- Step 4: Draw the elements of the variance matrix Λ_t conditional on the history of B_t, A, Q, Φ, and the mixture states.
- Following Primiceri (2005), the VAR can be rewritten as

$$A(y_t - X_t' B_t) \equiv \tilde{y}_t = \Lambda_t^{0.5} \epsilon_t,$$

where $\epsilon_t \sim N(0, I_k)$. Taking logs of the squares yields

$$\log \tilde{y}_{i,t}^2 = \log \lambda_{i,t}^2 + \log \epsilon_{i,t}^2, \;\; i = 1, \ldots, k.$$

- The conditional volatility process is

$$\log(\lambda_{i,t}^2) = \log(\lambda_{i,t-1}^2) + \nu_{i,t}, \;\; i = 1, \ldots, k.$$

TVPVAR-SV: estimation(5)

- The estimation of the time series of $\lambda_{i,t}^2$ uses the vector of the measured log $\tilde{y}_{i,t}^2$ and Primiceri's (2005) version of the Kim, Shephard, and Chib (1998) algorithm; see Primiceri for further detail (we depart from Primiceri by using the Durbin-Koopman simulation smoother instead of the Carter-Kohn smoother).
- Step 5: Draw the variance matrix Q conditional on the history of B_t, the history of Λ_t, A, and Φ.
- Following Cogley and Sargent (2005) and Primiceri (2005), the sampling of Q, the variance-covariance matrix of innovations to the VAR coefficients, is based on inverse Wishart priors and posteriors. The scale matrix of the posterior distribution is the sum of the prior mean $\times$ the prior degrees of freedom and $\sum_{t=1}^{T} \hat{n}_t \hat{n}_t'$, where $\hat{n}_t$ denotes the innovations to the posterior draws of coefficients obtained in step 1.

TVPVAR-SV: estimation(6)

- Step 6: Draw the variance matrix Φ, conditional on the history of B_t, the history of Λ_t, A, and Q.
- Following Primiceri (2005), the sampling of Φ, the variance of innovations to the log variances, is based on inverse Wishart priors and posteriors. The scale matrix of the posterior distribution is the sum of the prior mean $\times$ the prior degrees of freedom and $\sum_{t=1}^{T} \hat{\nu}_t \hat{\nu}_t'$, where $\hat{\nu}_t$ denotes the vector of innovations to the posterior draw of the volatilities for the set of variables.