VARs with drifting volatilities

Andrea Carriero

January 2018

VARs with drifting volatilities

- Consider the following VAR model with stochastic volatility:

$$\begin{aligned} y_t &= \Pi_0 + \Pi(L) y_{t-1} + v_t, \\ v_t &= A^{-1} \Lambda_t^{0.5} \epsilon_t, \ \epsilon_t \sim iid \ N(0, I_N), \end{aligned}$$

  where $\Pi(L) = \Pi_1 L + \Pi_2 L^2 + ... + \Pi_p L^p$, Λ_t is diagonal with generic j-th element $h_{j,t}$ and A^{-1} is lower triangular with ones on its main diagonal.
- The specification above implies

$$\Sigma_t \equiv Var(v_t) = A^{-1} \Lambda_t A^{-1\prime}$$

- The generic j-th element of the rescaled disturbances $\tilde{v}_t = A v_t$ is given by $\tilde{v}_{j,t} = h_{j,t}^{0.5} \epsilon_{jt}$. Taking logs of squares of $\tilde{v}_{j,t}$ yields the observation equations:

$$\ln \tilde{v}_{j,t}^2 = \ln h_{j,t} + \ln \epsilon_{j,t}^2, \ j = 1, \ldots, N. \tag{1}$$

- The model is completed by specifying laws of motion for the unobserved states:

$$\ln h_{j,t} = \ln h_{j,t-1} + e_{j,t}, \ j = 1, \ldots, N, \tag{2}$$

  where the vector of innovations to volatilities e_t is an iid Gaussian with a variance matrix Φ that is full

The Model - priors

- In a Bayesian setting, to estimate the model the likelihood needs to be combined with a prior distribution for the model coefficients

$$\Theta = \{\Pi, A, \Phi\}$$

  and the unobserved states Λ_t. The matrix Π collects the lag matrices $\Pi_0, \Pi_1, ..., \Pi_p$.

- The priors for the coefficients blocks of the model are as follows:

$$\begin{aligned} \text{vec}(\Pi) &\sim N(\text{vec}(\underline{\mu}_{\Pi}), \underline{\Omega}_{\Pi}); \\ A &\sim N(\underline{\mu}_{A}, \underline{\Omega}_{A}); \\ \Phi &\sim IW(\underline{d}_{\Phi} \cdot \underline{\Phi}, \underline{d}_{\Phi}). \end{aligned}$$

The Model - posteriors

- The conditional posteriors of the coefficients are:

$$\begin{aligned}
\text{vec}(\Pi)|A, \Lambda_T, y_T &\sim N(\text{vec}(\bar{\mu}_\Pi), \overline{\Omega}_\Pi); & (3)\\
A|\Pi, \Lambda_T, y_T &\sim N(\bar{\mu}_A, \overline{\Omega}_A); & (4)\\
\Phi|\Lambda_T, y_T &\sim IW((\underline{d}_\Phi + T)\cdot\bar{\Phi}, \underline{d}_\Phi + T), & (5)
\end{aligned}$$

where Λ_T and y_T denote the history of the states and data up to time T, and where the posterior moments $\bar{\mu}_\Pi$, $\overline{\Omega}_\Pi$, $\bar{\mu}_A$, $\overline{\Omega}_A$ and $\bar{\Phi}$ can be derived by combining prior moments and likelihood moments.

Gibbs sampler

1. Draw from $p(\Theta, s^T | \Lambda_T, y_T)$ relying on the factorization $p(\Theta, s^T | \Lambda_T, y) \propto p(s^T | \Theta, \Lambda_T, y) \cdot p(\Theta | \Lambda_T, y)$, that is by
   - a Drawing from the marginal posterior of the model parameters $p(\Theta | \Lambda_T, y_T)$
     - i Draw $\Phi | \Lambda_T, y_T$ using (5).
     - ii Draw $\text{vec}(\Pi) | A, \Lambda_T, y_T$ using (3).
     - iii Draw $A | \Pi, \Lambda_T, y_T$ using (4).
   - b Drawing from the conditional posterior of the mixture states $p(s^T | \Theta, \Lambda_T, y_T)$.
2. Draw from $p(\Lambda_T \, | \Theta, s^T, y_T)$ relying on the state space representation and Carter and Kohn (1994).

Note that step 1a and 1b are **not** intercheangeable since they constitute a draw from the joint of $p(\Theta, s^T | \Lambda_T, y_T)$

Drawing A

- Note that, in this model we have the additional steps of drawing the covariances A, which did not happen in the univariate case (for the obvious reason there is only 1 variable in that case)
- How we derive the posterior? Under knowledge of Π we can compute $v_t = A^{-1}\Lambda_t^{0.5}\epsilon_t$, and by pre-multiplying by $\Lambda_t^{-0.5}$ we have:

  $$\Lambda_t^{-0.5} v_t = A^{-1}\epsilon_t$$

  which is a system of unrelated regressions. The first one is

  $$\Lambda_{1,t}^{-0.5} v_{1,t} = \epsilon_{1,t}$$

  the second one is

  $$\Lambda_{2,t}^{-0.5} v_{2,t} = A_{21}^{-1}\epsilon_{1,t} + \epsilon_{2,t}$$

  and so on until the last equation:

  $$\Lambda_{n,t}^{-0.5} v_{n,t} = A_{n1}^{-1}\epsilon_{1,t} + ... + A_{nn-1}^{-1}\epsilon_{n-1,t} + \epsilon_{n,t}.$$

- Each of these equations - conditioning on the previous ones- is a CLRM. Hence by assumning priors on the elements of A^{-1} we can obtain the conditional posterior using standard results.

Drawing A - alternative method (in the code)

- What if we want to set a prior on the elements of A instead?
- Under knowledge of Π we can compute $v_t = A^{-1}\Lambda_t^{0.5}\epsilon_t$, and by pre-multiplying by A^{-1} we have:

$$Av_t = \Lambda_t^{0.5}\epsilon_t$$

which is a system of unrelated regressions. The first one is

$$v_{1,t} = \Lambda_{1,t}^{0.5}\epsilon_{1,t}$$

the second one is

$$A_{21}v_{1,t} + v_{2,t} = \Lambda_{2,t}^{0.5}\epsilon_{2,t}$$

and so on until the last equation:

$$A_{n,1}v_{1,t} + A_{n,2}v_{2,t} + ... + A_{n,n-1}v_{n-1,t} + v_{n,t} = \Lambda_{n,t}^{0.5}\epsilon_{n,t}$$

- This is a sequence of CLRM with generic equation:

$$v_{j,t} = -A_{j,1}v_{1,t} - A_{j,2}v_{2,t} - ... - A_{j,j-1}v_{j-1,t} + \Lambda_{j,t}^{0.5}\epsilon_{j,t}$$

- Each of these equations - conditioning on the previous ones- is a CLRM. Hence by assumning priors on the elements of A^{-1} we can obtain the conditional posterior using standard results.

Role of prior on A

- For the matrix A typically (Sims and Zha, Cogley and Sargent) is elicited a Gaussian independent prior element by element.
- This strategy implies that the ordering matters on the joint prior and -therefore- on the joint posterior.
- This effect, while present any time uses the LDL factorization, is likely to be empirically small, since the likelihood information should prevail.
- Primiceri (2005) models A as time varying. The ordering caveat applies also to his case
- The only way to avoid this is to specify time variation for the whole error variance matrix (e.g. as Shin and Zhong 2015)
- This has nothing to do with structural analysis considerations

Drawing the conditional mean coefficients

- Consider performing a draw Π^m from the conditional posterior of Π. One needs to draw a $N(Np+1)-$dimensional random vector denoted rand, and to compute:

$$\mathrm{vec}(\Pi^m) = \bar{\Omega}_\Pi \left\{ \mathrm{vec}\left(\sum_{t=1}^{T} X_t y_t' \Sigma_t^{-1} \right) + \underline{\Omega}_\Pi^{-1} \mathrm{vec}(\underline{\mu}_\Pi) \right\} + chol(\bar{\Omega}_\Pi) \times \mathrm{rand},$$

where $X_t = [1, y_{t-1}', ..., y_{t-p}']'$ is $(Np+1)$

- The calculation above requires to compute: i) the matrix $\bar{\Omega}_\Pi$ by inverting

$$\bar{\Omega}_\Pi^{-1} = \underline{\Omega}_\Pi^{-1} + \sum_{t=1}^{T} (\Sigma_t^{-1} \otimes X_t X_t');$$

ii) its Cholesky factor $chol(\bar{\Omega}_\Pi)$; iii) multiply the matrices obtained in i) and ii) by the vector in the curly brackets and the vector rand respectively.

- The computational complexity is $O(N^6)$.

Triangularization of large VARs with drifting volatiltities

Recall that in the step of the Gibbs sampler that involves drawing Π, all of the remaining model coefficients are given, and consider again the decomposition $v_t = A^{-1}\Lambda_t^{0.5}\epsilon_t$:

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \dots \\ v_{N,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a^*_{2,1} & 1 & & \dots \\ \dots & & 1 & 0 \\ a^*_{N,1} & \dots & a^*_{N,N-1} & 1 \end{bmatrix} \begin{bmatrix} h^{0.5}_{1,t} & 0 & \dots & 0 \\ 0 & h^{0.5}_{2,t} & & \dots \\ \dots & & \dots & 0 \\ 0 & \dots & 0 & h^{0.5}_{N,t} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \dots \\ \epsilon_{N,t} \end{bmatrix},$$

where $a^*_{j,i}$ denotes the generic element of the matrix A^{-1} which is available under knowledge of A. We will also denote by $\pi^{(i)}$ the vector of coefficients for equation i contained in row i of Π, for the intercept and coefficients on lagged y_t.

Triangularization of large VARs with drifting volatiltities

The VAR can be written as:

$$y_{1,t} = \pi_1^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{1,l}^{(i)} y_{i,t-l} + h_{1,t}^{0.5}\epsilon_{1,t}$$

$$y_{2,t} = \pi_2^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{2,l}^{(i)} y_{i,t-l} + a_{2,1}^{*} h_{1,t}^{0.5}\epsilon_{1,t} + h_{2,t}^{0.5}\epsilon_{2,t}$$

$$\dots$$

$$y_{N,t} = \pi_N^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{N,l}^{(i)} y_{i,t-l} + a_{N,1}^{*} h_{1,t}^{0.5}\epsilon_{1,t} + \cdots + a_{N,N-1}^{*} h_{N-1,t}^{0.5}\epsilon_{N-1,t} + h$$

with the generic equation for variable j:

$$y_{j,t} - (a_{j,1}^{*} h_{1,t}^{0.5}\epsilon_{1,t} + \dots + a_{j,,j-1}^{*} h_{j-1,t}^{0.5}\epsilon_{j-1,t}) = \pi_j^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{j,l}^{(i)} y_{i,t-l} + h_{j,t}\epsilon_{j,t}.$$

Consider estimating these equations in order from $j=1$ to $j=N$. When estimating the generic equation j the term on the left hand side is known.

Triangulularization of large VARs with drifting volatiltities

We can define:

$$y^*_{j,t} = y_{j,t} - (a^*_{j,1} h^{0.5}_{1,t} \epsilon_{1,t} + ... + a^*_{j,,j-1} h^{0.5}_{j-1,t} \epsilon_{j-1,t}), \tag{6}$$

and the equation:

$$y^*_{j,t} = \pi^{(0)}_j + \sum_{i=1}^{N} \sum_{l=1}^{p} \pi^{(i)}_{j,l} y_{i,t-l} + h_{j,t} \epsilon_{j,t}.$$

becomes a standard generalized linear regression model with i.i.d. Gaussian disturbances with mean 0 and variance $h_{j,t}$.
The full conditional posterior distribution of the conditional mean coefficients can be factorized as:

$$\begin{aligned} p(\Pi | A, \Lambda_T, y) \quad = \quad & p(\pi^{(N)} | \pi^{(N-1)}, \pi^{(N-2)}, \ldots, \pi^{(1)}, A, \Lambda_T, y) \\ & \times p(\pi^{(N-1)} | \pi^{(N-2)}, \ldots, \pi^{(1)}, A, \Lambda_T, y) \\ & \vdots \\ & \times p(\pi^{(1)} | A, \Lambda_T, y), \end{aligned}$$

Triangularization of large VARs with drifting volatiltities

- One can draw the coefficients of the matrix Π in separate blocks $\Pi^{\{j\}}$ which can be obtained from:

$$\Pi^{\{j\}}|\Pi^{\{1:j-1\}}, A, \Lambda_T, y \sim N(\bar{\mu}_{\Pi^{\{j\}}}, \overline{\Omega}_{\Pi^{\{j\}}})$$

with

$$\begin{aligned} \bar{\mu}_{\Pi^{\{j\}}} &= \overline{\Omega}_{\Pi^{\{j\}}} \left\{ \underline{\Omega}^{-1}_{\Pi^{\{j\}}} \underline{\mu}_{\Pi^{\{j\}}} + \sum_{t=1}^{T} X_{j,t} h^{-1}_{j,t} y^{*\prime}_{j,t} \right\} \\ \overline{\Omega}^{-1}_{\Pi^{\{j\}}} &= \underline{\Omega}^{-1}_{\Pi^{\{j\}}} + \sum_{t=1}^{T} X_{j,t} h^{-1}_{j,t} X'_{j,t}, \end{aligned}$$

where $y^*_{j,t}$ is defined in (6) and where $\underline{\Omega}^{-1}_{\Pi^{\{j\}}}$ and $\underline{\mu}_{\Pi^{\{j\}}}$ denote the prior moments on the j-th equation, given by the j-th column of $\underline{\mu}_{\Pi}$ and the j-th block on the diagonal of $\overline{\Omega}^{-1}_{\Pi}$.
- We have implicitly assumed here that the matrix $\underline{\Omega}^{-1}_{\Pi}$ is block diagonal. This assumption is frequent but can be easily relaxed.
- The ordering of the equations is immaterial as far as producing a draw from $\Pi|A, \Lambda_T, y$ is concerned.

VARs with drifting volatiltities and coefficients

- The VAR becomes:

$$\begin{aligned} y_t &= \Pi_{0_t} + \Pi_t(L)y_{t-1} + v_t, \\ v_t &= A_t^{-1}\Lambda_t^{0.5}\epsilon_t,\ \epsilon_t \sim iid\ N(0, I_N), \end{aligned}$$

  where $\Pi(L) = \Pi_{1,t}L + \Pi_{2,t}L^2 + ... + \Pi_{p,t}L^p$, Λ_t is diagonal with generic j-th element $h_{j,t}$ and A_t^{-1} is lower triangular with ones on its main diagonal.
- The specification above implies

$$\Sigma_t \equiv Var(v_t) = A_t^{-1}\Lambda_t A_t^{-1\prime}$$

- This specification is the one of Primiceri (2005). Cogley and Sargent (2005) consider a special case in which $A_t = A$.

VARs with drifting volatiltities and coefficients

- This VAR has more state variables: the Π_t and the A_t^{-1}.
- Typically, one assumes a RW (or AR) process for these state variables:

$$\begin{aligned} \ln \pi_{ij,t} &= \ln \pi_{ij,t-1} + \upsilon_{ij,t},\ j = 1, \ldots, Np, i = 1, ..., N \\ \ln a_{ij,t} &= \ln a_{ij,t-1} + \varsigma_{ij,t},\ j = 1, \ldots, N, i = 1, ..., N,\ j < i \end{aligned}$$

- These are in adddition to the volatility states:

$$\ln h_{j,t} = \ln h_{j,t-1} + e_{j,t},\ j = 1, \ldots, N,$$

VARs with drifting volatiltities and coefficients

- The shocks across all the state variables are mutually independent:

$$Var \begin{pmatrix} \epsilon_t \\ \upsilon_{ij,t} \\ \varsigma_{ij,t} \\ e_{j,t} \end{pmatrix} = \begin{pmatrix} I_N & 0 & 0 & 0 \\ 0 & \Omega_\upsilon & 0 & 0 \\ 0 & 0 & \Omega_\varsigma & 0 \\ 0 & 0 & 0 & \Phi \end{pmatrix} = \Omega$$

VARs with drifting volatiltities and coefficients

1. Draw from $p(\Omega, \Pi_{1:T}, A_T, s_T | \Lambda_T, y_T)$ relying on the factorization $p(\Omega, \Pi_{1:T}, A_T, s_T | \Lambda_T, y_T) \propto$ $p(s^T | \underbrace{\Omega, \Pi_{1:T}, A_T}, \Lambda_T, y) \cdot p(\underbrace{\Omega, \Pi_{1:T}, A_T} | \Lambda_T, y_T)$, that is by
   - a Draw from $p(\underbrace{\Omega, \Pi_{1:T}, A_T} | \Lambda_T, y_T)$
     - i Draw $\Phi | \Omega_\varsigma, \Omega_\upsilon, \Pi_{1:T}, A_T, \Lambda_T, y_T \ \propto \Phi | \Lambda_T, y_T$
     - ii Draw $\Omega_\upsilon | \Omega_\varsigma, \Phi, \Pi_{1:T}, A_T, \Lambda_T, y_T \ \propto \Omega_\upsilon | \Pi_{1:T}, y_T$
     - iii Draw $\Omega_\varsigma | \Omega_\upsilon, \Phi, \Pi_{1:T}, A_T, \Lambda_T, y_T \propto \Omega_\varsigma | A_T, y_T$
     - iv Draw the states $\Pi_{1:T} | A_T, \Lambda_T, \Omega, y_T$
     - v Draw the states $A_T | \Pi_{1:T}, \Lambda_T, \Omega, y_T$
   - b Draw from $p(s_T | \underbrace{\Omega, \Pi_{1:T}, A_T}, \Lambda_T, y_T)$
2. Draw from $p(\Lambda_T | s_T, \Pi_{1:T}, A_T, \Omega, y_T)$

Note that step 1a and 1b are **not** intercheangeable since they constitute a draw from the joint of $p(\Omega, \Pi_{1:T}, A_T, s_T | \Lambda_T, y_T)$