TIME VARYING STRUCTURAL VECTOR AUTOREGRESSIONS AND MONETARY POLICY: A CORRIGENDUM

MARCO DEL NEGRO AND GIORGIO E. PRIMICERI

ABSTRACT. This note shows how to apply the procedure of Kim, Shephard and Chib (1998) to the estimation of VAR, DSGE, factor, and unobserved components models with stochastic volatility. In particular, it revisits the estimation algorithm of the time-varying VAR model of Primiceri (2005). The main difference of the new algorithm is the ordering of the various MCMC steps, with each individual step remaining the same.

JEL classification: C11, C15

Key words: Bayesian methods, time-varying volatility

1. THE MODEL IN SHORT

This note is a corrigendum of Primiceri (2005), but its lesson applies more broadly to several empirical macro models with stochastic volatility that are estimated using the approach of Kim, Shephard, and Chib (1998, KSC hereafter). Consider the time-varying VAR model of Primiceri (2005)

$$y_t = c_t + B_{1,t}y_{t-1} + ... + B_{k,t}y_{t-k} + A_t^{-1}\Sigma_t\varepsilon_t, \tag{1.1}$$

where y_t is an $n\times 1$ vector of observed endogenous variables; c_t is a vector of time-varying intercepts; $B_{i,t}$, $i = 1, ..., k$, are matrices of time-varying coefficients; A_t is a lower triangular matrix with ones on the main diagonal and time-varying coefficients below it; Σ_t is a diagonal matrix of time-varying standard deviations; ε_t is an $n \times 1$ vector of unobservable shocks with variance equal to the identity matrix. All the time-varying coefficients evolve as random walks, except for the diagonal elements of Σ_t, which behave as geometric random walks. All the innovations in the model (shocks to coefficients, log-volatilities and ε_t) are jointly normally distributed, with mean equal to zero and covariance matrix equal to V. The matrix V is block diagonal, with blocks corresponding to the time-varying elements of the B's, A, Σ and ε. The block structure of the matrix V is described in detail in Primiceri (2005).

Date: February 2015; First draft: April 2013.

2. The Original Algorithm of Primiceri (2005)

The unknown objects of the model are the history of the volatilities (Σ^T), the history of the coefficients (B^T and A^T), and the covariance matrix of the innovations (V). To simplify the notation, define $\theta \equiv \left[B^T, A^T, V\right]$. Primiceri (2005) proposed to simulate the posterior distribution of the model coefficients by Gibbs sampling, drawing the history of volatilities with the multi-move algorithm of KSC.

The difficulty with drawing Σ^T is that they enter the model multiplicatively. For given θ, however, simple algebraic manipulations of (1.1) yield a linear system in the log volatilities. A consequence of applying these transformations is that they also convert ε_t into $\log \varepsilon_t^2$, which is a vector of $\log \chi^2(1)$ random variables. The method of KSC relies on approximating each element of $\log \varepsilon_t^2$ with a mixture of normals. Conditioning on the mixture indicators makes it possible to use standard Gaussian state-space methods to conduct inference on the volatilities. As a consequence, the Gibbs sampler is augmented to include the mixture indicators $s^T \equiv \{s_t\}_{t=1}^T$ that select the component of the mixture for each variable at each date.

Primiceri (2005) adopts the following algorithm to obtain posterior draws for Σ^T, s^T and θ:

Algorithm 1 (original algorithm)

(1) Draw Σ^T from $\tilde{p}\left(\Sigma^T | y^T, \theta, s^T\right)$
(2) Draw s^T from $\tilde{p}\left(s^T | y^T, \Sigma^T, \theta\right)$
(3) Draw θ from $p\left(\theta | y^T, \Sigma^T\right)$,

where the "~" in step 1 and 2 indicates that the conditional posteriors of Σ^T and s^T correspond to the product of their conditional priors by $\tilde{p}(y^T|\Sigma^T, \theta, s^T)$, i.e. the likelihood of the data conditional on the components of the mixture-of-normals approximation of the $\log \chi^2(1)$ distribution for each date and variable. The conditional posterior of θ in step 3 is instead obtained using the true likelihood implied by model (1.1), i.e. $p\left(y^T|\Sigma^T, \theta\right)$.

There are two reasons why this algorithm does not yield draws from the correct posterior distribution of the model parameters. First of all, the algorithm alternates between the use of two different likelihood functions: steps 1 and 2 of the sampler make use of the mixture-of-normals approximation, to facilitate the draw of Σ^T; step 3, instead, uses the correct likelihood.

More important, the second problem with Algorithm 1 is related to the fact that it was conceived as a Gibbs sampler with "blocks" Σ^T, θ , and s^T. In a Gibbs sampler, one has to draw from each block conditional on all the others. However, the draw of θ in step 3 is not conditional on s^T. Primiceri (2005) erroneously assumed that conditioning on s^T in step 3 does not make a difference, but instead it does: the knowledge of which components of the mixture have been selected for each date and variable changes the likelihood of the data, thus affecting the conditional posterior of θ. This simple observation exposes the problems of Algorithm 1, even abstracting from the approximation error. In other words, Algorithm 1 would not yield draws from the correct posterior even if we used an arbitrarily large number of mixture components to make the approximation arbitrarily accurate.

3. A Gibbs Sampler with Different Blocking

Fixing this problem of Algorithm 1 by simply replacing step 3 with "Draw from $\tilde{p}\left(\theta|y^T,\Sigma^T,s^T\right)$" is not a viable option because $\varepsilon_t|s_t$ in not Gaussian, which precludes the possibility of drawing easily from $\tilde{p}\left(\theta|y^T,\Sigma^T,s^T\right)$. An alternative strategy is to use a Gibbs sampler with different blocking. Instead of using three blocks, Σ^T, θ , and s^T, one can use two blocks, i.e. Σ^T and (θ, s^T). The first step of the new sampler is to draw Σ^T conditional on (θ, s^T) and the data y^T. The second step is to draw from the joint distribution of (θ, s^T) conditional on Σ^T and the data. Of course, drawing from the joint of $\left(\theta, s^T\right)$ can be accomplished by drawing first from the marginal of θ and then from the conditional of s^T given θ. This yields the following algorithm (of which the online appendix presents a more formal treatment):

Algorithm 2 (correct algorithm under no approximation error)

(1) Draw Σ^T from $\tilde{p}\left(\Sigma^T|y^T,\theta,s^T\right)$
(2) Draw (θ, s^T) from $\tilde{p}\left(\theta,s^T|y^T,\Sigma^T\right)$, which is accomplished by
  (a) Drawing θ from $p\left(\theta|y^T,\Sigma^T\right)$
  (b) Drawing s^T from $\tilde{p}\left(s^T|y^T,\Sigma^T,\theta\right)$,

where the "$\tilde{}$" notation in steps 1 and 2b continues to indicate the use of the auxiliary approximating model—as opposed to the true likelihood—to facilitate the draw of the history of volatilities.

Like Algorithm 1, also Algorithm 2 alternates between the use of the correct and the approximate likelihood. However, unlike Algorithm 1, Algorithm 2 has the property that it would yield draws from the correct posterior in the hypothetical case in which the mixture of normals represented a perfect approximation for the $\log \chi^2(1)$ distribution, as we formally show in the online appendix. As we stress in the next section, in practice, the mixture of normals is of course only an approximation of the $\log \chi^2(1)$ distribution. We therefore think of Algorithm 2 as a sampler from an approximate posterior.

Finally, notice that the individual steps in Algorithms 1 and 2 are the same, but the order is different: in Algorithm 2 the indicators s^T are sampled *after* θ and *before* Σ^T. Since the individual steps remain the same, they can all be implemented as in Primiceri (2005).[1] Algorithm 2 is therefore equivalent to switching steps (d) and (e) in the algorithm summarized in Appendix A.5 of Primiceri (2005).[2] This order is key to derive Algorithm 2 as a Gibbs sampler based on the two blocks Σ^T and (θ, s^T), and thus to justify the draw of θ from a posterior that does not conditions on s^T.

4. Addressing the Approximation Problem

In this section we explicitly deal with the issue of the approximation error, recognizing the fact that the finite mixture of normals is only used as an approximation of the $\log \chi^2(1)$ distribution. Stroud et al. (2003) show how to address this problem by turning step 1 of Algorithm 2 into a Metropolis-Hastings step, where the distribution $\tilde{p}\left(\Sigma^T|y^T, \theta, s^T\right)$ is used as a proposal density. Specifically, we set up another algorithm, which we denote by *Algorithm 3 (correct algorithm)*. Steps 2a and 2b of Algorithm 3 are the same as in Algorithm 2. Step 1 is instead replaced with a candidate draw from the proposal density $\tilde{p}\left(\Sigma^T|y^T, \theta, s^T\right)$. This draw is then accepted with probability proportional to the ratio between the conditional density of the new and previous draw, re-weighted by the ratio between the proposal density of the previous and the new draw, as standard in each Metropolis-Hastings algorithm. If the candidate draw of Σ^T is not accepted, the draw of Σ^T is set equal to the previous draw. The functional form of the acceptance probability is shown in equation (11) of Stroud et al. (2003), and re-derived in our online appendix for the specific case of our model.

A formal illustration of Algorithm 3 requires some investment in notation and is therefore relegated to the online appendix. We stress that this sampler is correct (i.e. eventually yields the right posterior density of Σ^T and θ) regardless of the quality of the approximation, which matters only for its efficiency. We also emphasize that a key step in Algorithm 3, as in Algorithm 2, consists in integrating out the mixture components when drawing θ, which implies inverting the order of the draws of Σ^T and s^T relative to the original Gibbs sampler. This is the main difference relative to Primiceri (2005). The lesson of this note is that researchers using the KSC approach to estimate VAR, DSGE, or factor models with time-varying volatility need to make sure they sample the indicators s^T right before the history of volatilities. Examples of such papers are numerous in the past

[1] In particular, step (2) can be implemented by drawing from $p\left(B^T|y^T, A^T, V, \Sigma^T\right)$, $p\left(A^T|y^T, B^T, V, \Sigma^T\right)$ and $p\left(V|y^T, A^T, B^T, \Sigma^T\right)$.

[2] Section A.5 in Primiceri (2005) actually contains a typo: step (d) of the algorithm should be $p(s^T|y^T, \mathbf{B^T}, A^T, \Sigma^T, V)$ as opposed to $p(s^T|y^T, A^T, \Sigma^T, V)$. Unlike the conceptual mistake outlined in the previous section, this typo was inconsequential given that it is mechanically not possible to draw s^T without conditioning on B^T.

decade, e.g. Justiniano and Primiceri (2008).[3] This lesson also applies to unobserved components models with stochastic volatility (e.g., Stock and Watson, 2007).

5. Consequences for the Results

In the online appendix, we have applied Geweke's (2004) "Joint Distribution Tests of Posterior Simulators" to further confirm that Algorithm 3 is fully correct, Algorithm 2 provides a close approximation to the true posterior distribution, while Algorithm 1 provides a poor approximation. In addition, we have re-estimated the model of Primiceri (2005) using Algorithm 2 and 3, and compared the results to the original ones obtained with Algorithm 1.

Algorithm 2 generates results that are indistinguishable from those obtained with Algorithm 3, suggesting that the mixture-of-normals approximation error involved in the procedure of KSC is negligible in our application (as it was in theirs). The results based on Algorithm 2 and 3 are instead not the same as those obtained with Algorithm 1, albeit qualitatively similar. The main difference is that some estimates of the time-varying objects are now smoother. The full set of new results can be found in the online appendix.

REFERENCES

GEWEKE, J. (2004), "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, 99, 799–804.

JUSTINIANO, A. and PRIMICERI, G. E. (2008), "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–641.

KIM, S., SHEPHARD, N. and CHIB, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

PRIMICERI, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

STOCK, J.H. and WATSON, M.W. (2007), "Why Has U.S. Inflation Become Harder to Forecast?" Journal of Money, Credit, and Banking, 39(1), 3–33.

[3]The estimation algorithm of the DSGE model with stochastic volatility of Justiniano and Primiceri (2008) is correct, although their appendix describes an algorithm with the wrong order.

STROUD, J.R., MÜLLER, P. and POLSON, N.G. (2003), "Nonlinear State-Space Models With State-Dependent Variances," *Journal of the American Statistical Association*, 98, 377–386.

Federal Reserve Bank of New York

Northwestern University, CEPR and NBER