Bayesian Analysis (2014) **9**, Number 4, pp. 819–822

Rejoinder

Jesse Windle * and Carlos M. Carvalho †

A good deal of thanks go out to all of the discussants for their insightful and interesting comments, to the referees for their help in improving the paper, and to the editors for enabling this discourse.

Broadly speaking, the criticisms and suggestions of the discussants pointed to several theoretical and applied weaknesses. Without respect to any specific modeling goal, the Uhlig extension as originally presented suffers from several unappealing features: it uses only a few fixed parameters to model large, time-varying quantities, which is a bit too parsimonious, and the evolution of the hidden states is problematic. When examining the model within the context of finance, the model misses many key features and disagreements between observed statistical regularities and those captured by the model are brought into relief.

Thankfully, the discussants not only identified many shortcomings, but also provided many solutions.

We very much appreciate the proposed improvements in the discussions of both Forbes and Casarin. Within the context of a financial time series, Forbes has shown how to elegantly incorporate jumps into the dynamics of the price process while preserving all of the machinery of the Uhlig extension; Casarin has shown not only why one should want to use time-varying parameters, but how to incorporate them via a Markov-switching approach.

Regarding Casarin's suggested improvements (see Section 3), one must be careful when letting the degrees of freedom parameters change. He suggests

$$\begin{cases} \mathbf{Y}_t \sim W_m(k_t, (k_t \mathbf{X}_t)^{-1}), \\ \mathbf{X}_t \sim \mathbf{T}'_{t-1} \mathbf{\Psi}_t \mathbf{T}_{t-1} / \lambda, \qquad \mathbf{\Psi}_t \sim \beta_m\left(\frac{n_t}{2}, \frac{k_t}{2}\right), \\ \mathbf{T}_{t-1} = \text{upper chol } \mathbf{X}_{t-1} \end{cases}$$

where m is the order of the matrices involved. Given the initial distribution $(\mathbf{X}_0 \,|\, \mathcal{D}_0) \sim W_m(n_0 + k_0, (k_0 \mathbf{\Sigma}_0)^{-1})$, the filtered and predictive distributions evolve in the following way (we continue to use the notation $\mathcal{D}_t = (\mathbf{Y}_1, \ldots, \mathbf{Y}_t)$ and we implicitly condition on

*Duke University jesse.windle@stat.duke.edu
†The University of Texas at Austin carlos.carvalho@mccombs.utexas.edu

 DOI:10.1214/14-BA923

$\{n_t\}_{0=1}^T$, $\{k_t\}_{0=1}^T$, and $\{\lambda_t\}_{t=1}^T$):

$$\begin{aligned}
(\mathbf{X}_0 \,|\, \mathcal{D}_0) &\sim W_m(n_0 + k_0, (k_0 \mathbf{\Sigma}_0)^{-1}) \\
(\mathbf{X}_1 \,|\, \mathcal{D}_0) &\sim W_m(n_0 + k_0 - k_1, (\lambda_1 k_0 \mathbf{\Sigma}_0)^{-1}) \\
(\mathbf{X}_1 \,|\, \mathcal{D}_1) &\sim W_m(n_0 + k_0, (k_0 \mathbf{\Sigma}_1)^{-1}) \\
(\mathbf{X}_2 \,|\, \mathcal{D}_1) &\sim W_m(n_0 + k_0 - k_2, (\lambda_2 k_0 \mathbf{\Sigma}_1)^{-1}) \\
(\mathbf{X}_2 \,|\, \mathcal{D}_2) &\sim W_m(n_0 + k_0, (k_0 \Sigma_2)^{-1}) \\
&\cdots,
\end{aligned}$$

where the recursion for $\mathbf{\Sigma}_t$ is now

$$\mathbf{\Sigma}_t = \frac{k_t}{k_0} \mathbf{Y}_t + \lambda \mathbf{\Sigma}_{t-1}.$$

Examining the parameters of the Wishart distributions, one can see that if $\mathbf{X}_t$ has full rank, then $n_0 + k_0 - k_t$ and k_t should never drop below $m - 1$. Further, in the rank deficient case, the degrees of freedom need to remain fixed as k corresponds to the rank of $\mathbf{Y}_t$, though one may let the smoothing parameter λ vary.

As Casarin points out, we failed to explore other methods for examining the evolution of the variance process; he suggests looking at the dynamics of the first and second moments. Mercifully, Konno (1988) has derived the moments of the the multivariate beta distribution and its inverse (see Theorems 3.2 and 3.4 on p. 129). Using those moments one can define recurrence relations for the predicted means and covariances of $\mathbf{X}_t$ for the fixed parameters n, k, and λ. For instance, by the law of iterated expectations and Konno's work one can show that the h-step ahead prediction $\mathbb{E}[\mathbf{X}_{t+h}^{-1} \,|\, \mathbf{X}_t^{-1}]$ is defined by

$$M_h^{(1)} = \frac{n + k - m - 1}{n - m - 1} \, \lambda \, M_{h-1}^{(1)}$$

where $M_0^{(1)} = \mathbf{X}_t^{-1}$. If the constraint

$$\lambda = \frac{n - m - 1}{n + k - m - 1} \tag{1}$$

holds, which corresponds to constraint (3) in the main paper, then the h-step ahead prediction is the current state. Otherwise, the prediction grows or decays without bound.

As we note in the main manuscript, the process $\mathbf{X}_t$ is not stationary, which is an undesirable property. The growth or decay of the first moment is an artifact of this feature. Another example of this degeneracy can be found by examining the condition number of the process. The condition number of a matrix is the ratio of the largest to smallest eigenvalue. Mathematically non-singular matrices become numerically singular when the condition number is large. A cursory search turned up nothing so useful as Konno (1988), but Theorem 3.3.4 in Muirhead (1982) suggests that one could find a recursive description of the evolution of the eigenvalues of $\mathbf{X}_t$. We conjecture that one

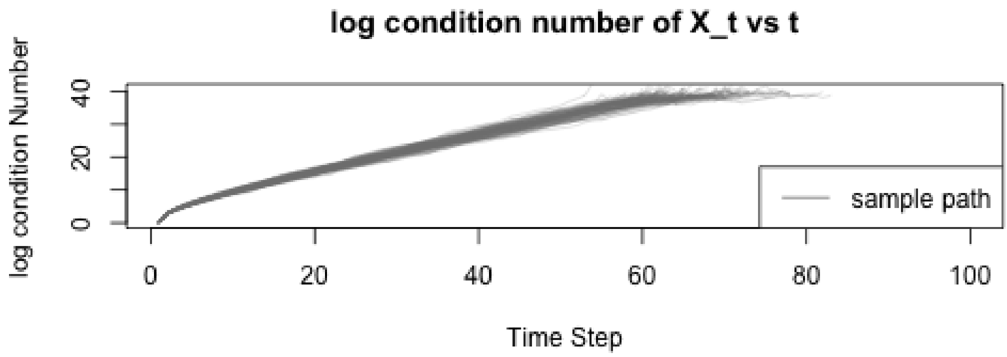

Figure 1: Sample paths of the log condition number of $\mathbf{X}_t$, $t = 1, \ldots, 100$. When the log condition number reaches roughly 40, the matrices become numerically singular and the simulation breaks down. We set the parameters to match those estimated in the realized covariance example from the main text with $m = 30$, $k = 67$, and $n = 393$. The parameter λ was set to satisfy contraint (1). In this case, the predicted values $\mathbb{E}[\mathbf{X}_{t+h}^{-1} \,|\, \mathbf{X}_t^{-1}]$ are constant in h; but the predicted values $\mathbb{E}[\mathbf{X}_{t+h} \,|\, \mathbf{X}_t]$ are not.

could show that the log condition number of $\mathbf{X}_t$ diverges, in which case simulating from $\{\mathbf{X}_t\}_{t=1}^T$ becomes problematic. Evidence in support of this conjecture can be seen by simulating the log condition number of $\mathbf{X}_t$, as seen in Figure 1.

Building on Casarin's Markov switching suggestion, it would be interesting to examine how one might use Markov switching to force the evolution of $\mathbf{X}_t$ to be more well behaved. For instance, one could introduce some value to describe the center of $\{\mathbf{X}_t\}_{t=1}^T$, call it $\mathbf{M}$, and then make the distribution of the parameters n_t, k_t, and λ_t depend on the location of $\mathbf{X}_t$ in relation to $\mathbf{M}$. Looking at Konno's moment relationships the expected value of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ is

$$\mathbb{E}[\mathbf{X}_t|\mathbf{X}_{t-1}] = \frac{n_t}{n_t + k_t}\frac{1}{\lambda_t}\mathbf{X}_{t-1}.$$

Thus, the expected growth or decay of $\mathbf{X}_t$ is determined by the ratio

$$\frac{n_t}{n_t + k_t}\frac{1}{\lambda_t}.$$

One could model n_t, k_t, and λ_t so that this ratio is on average greater than or less than unity depending on whether $\mathbf{X}_{t-1}$ is less than or greater than $\mathbf{M}$, respectively, under some suitably defined ordering. The requisite dynamics of n_t, k_t, and λ_t is left to future work.

The comprehensive analysis by ter Horst and Molina is essential reading for anyone interested in the application of stochastic volatility models to real world scenarios. We did not adequately address a number of issues they raised: the model ignores overnight returns, jumps, dividends, and asset specific news; has no leverage effect; is based on a single asset class; and, within that asset class, makes use of assets that trade at similar frequencies. In short, we ignored many of the practical issues one must face when doing complex volatility modeling.

Similarly, we avoided many subtle issues when collecting the data and constructing the realized covariance matrices. We did not think deeply about the opening and closing effects of the market, nor about asynchronicity or the impact of considering different intraday frequencies; instead we simply used the methods of Barndorff-Nielsen et al. (2011, 2009) and assumed that they had accounted for all of those issues. The sample period was chosen without much scrutiny either. To collect the data, we wrote a SAS script to query the Trades and Quotes database from Wharton Research Data Services. The script initially used ticker symbols. From time to time a company changes its ticker symbol. If such a change occurred over the period we queried, our script would break. Thus, the date February 27, 2007 simply corresponded to a starting date that did not break the script.

Regarding the portfolio comparisons, we did in fact use 24-hour returns when computing the minimum variance portfolio measures. We estimated and then predicted the daily covariance matrix of open to close returns by treating realized covariance matrices as data in the Uhlig extension. We then used those predictions to generate portfolios. The realized returns of those portfolios were computed on a 24-hour basis using the same data that was used in the factor stochastic volatility models.

Finally, to repeat the sentiment at the end of ter Horst and Molina's critique, it would indeed be interesting to see how factor stochastic volatility handled high-frequency returns directly. We leave that to future work.

References

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). "Realized Kernels in Practice: Trades and Quotes." *Econometrics Journal*, 12(3): C1–C32. 822

— (2011). "Multivariate Realized Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading." *Journal of Econometrics*, 162: 149–169. 822

Konno, Y. (1988). "Exact Moments of the Multivariate F and Beta Distributions." *Journal of the Japanese Statistical Society*, 18: 123–130. 820

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley. 820