SV models. We feel that both are important and deserve more attention. We hope our comments will encourage other researchers to tackle some of the challenging problems nonlinear models present in a creative way. There is much work to be carried out in this area.

ADDITIONAL REFERENCES

Carter, C. K., and Kohn, R. (in press), "On Gibbs Sampling for State Space Models," *Biometrika*, 61.

De Jong, P., and Shephard, N. (1993), "Efficient Sampling From the Smoothing Density of Gaussian Time Series Models," working paper, Nuffield College, Oxford.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of the United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fruhwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 14, 15, 183–202.

Kim, S., and Shephard, N. (1993), "Stochastic Volatility: New Models and Optimal Likelihood Inference," working paper, Nuffield College, Oxford.

Phillips, P. C. B. (1991), "To Criticise the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333–364.

Schotman, P. C., and Van Dijk, H. K. (1991), "A Bayesian Analysis of the Unit Root in Real Exchange Rates," *Journal of Econometrics*, 49, 195–238.

Shephard, N. (1994), "Partial Non-Gaussian State Space," *Biometrika*, 81, 115–131.

Journal of Business & Economic Statistics, October 1994, Vol. 12, No. 4

Comment

Harald Uhlig
CentER for Economic Research, Tilburg University, 5000 LE Tilburg, The Netherlands

Jacquier, Polson, and Rossi (JPR) perform a "Bayesian analysis of stochastic volatility models" in their article. They accomplish three things. They show how these models can be analyzed numerically in an elegant fashion using an extension of the Metropolis-chain algorithm. They apply their method to a variety of financial data series that have been of interest to other researchers working with stochastic volatility models. Finally, they demonstrate that their method is superior to estimation by method of moments (MM) and by quasi-maximum likelihood (QML) methods.

I liked the article very much, and it presents a huge leap forward in our ability to conveniently and reliably analyze stochastic volatility by introducing Metropolis-chain algorithms to this literature. I wish the authors had given a bit more of a "cooking recipe" on how to do the cyclic Metropolis-chain algorithm: It is all in there but not in a plug-and-play version ready for the average research assistant. Moreover, the authors are amazingly muffled concerning the advantage of a Bayesian over a classical point of view: They mainly claim that the Metropolis-chain method yields superior estimators as compared to other numerical techniques (and another title for the article could have been "A Fast, Reliable Technique for Analyzing Stochastic Volatility Models"). Undoubtedly this is done to appeal to the majority of classical-minded readers—a laudable goal, but it is also an opportunity lost for telling these same readers why a Bayesian perspective can be so fruitful.

There are other and potentially better competitors than MM and QML, which may prove equally useful or useful as a companion device to the techniques developed in this article. Nonetheless, the techniques of JPR are here to stay. The article thus makes an important contribution, removing a major stumbling block in the analysis of stochastic volatility models. These techniques should be used by many applied researchers interested in analyzing situations with potential stochastic volatility.

1. THE CYCLIC METROPOLIS–CHAIN ALGORITHM

The key contribution of this article is technical—introducing a cyclic Metropolis-chain algorithm and bringing it to bear on the analysis of stochastic volatility models. This algorithm belongs to a series of recent breakthroughs regarding the numerical analysis of posteriors.

Bayesian analysis used to be hard. It has become simple and, in fact, is now simpler than the classical analysis of stochastic volatility models. To see why, note that a fully specified model will result in a likelihood function $L(\theta; y)$, where $\theta \in \Theta$ is the vector of parameters and y is the data. Bayesians tend to multiply this likelihood function with a prior $\pi(\theta)$, but never mind: With much data and few parameters for which one needs to choose a prior, as is the case in this article, the choice of a prior is very much a nonissue and one might as well just work with $L(\theta; y)$. A classical econometrician wants to know the maximum $\widehat{\theta}$ as well as some local information around it (like the second derivative matrix I at $\widehat{\theta}$) to do inference. A Bayesian wants a sample $\theta^{(1)}, \ldots, \theta^{(N)}$ in the parameter space Θ from the probability distribution $\alpha L(\theta; y)$, where α is the appropriate scaling fac-

tor, and then base inference on sample averages, where N can be chosen large enough to achieve a desired degree of accuracy. It used to be the case (and sometimes still is) that one needs $\widehat{\theta}$ and I to do that so that one could sample from, say, a t distribution with the same first and second moments as an approximation, weighing the draws appropriately (see Geweke 1989). Now, the algorithms used in this article can deliver the sample right away. This is useful for a classical econometrician as well because the likelihood function for stochastic volatility models is notoriously hard to maximize using conventional techniques (and this used to be the stumbling block in using these models): Infer $\widehat{\theta}$ and I from the sample $\theta^{(1)}, \ldots, \theta^{(N)}$ (I am sure there are techniques for that) and proceed as usual. But why bother? Once the sample is obtained, Bayesian inference is simpler.

How does their cyclic Metropolis-chain algorithm work? For the "cyclic" part, take one-dimensional (or low-dimensional) slices of the parameter space Θ and sample from each in turn. If $\theta^{(n)} = (\theta_1^{(n)}, \ldots, \theta_k^{(n)})$ was the last draw and $\theta_i^{(n)}$ the last updated component, sample a new $\tilde{\theta}_{i+1}$ (or $\tilde{\theta}_1$, if $i = k$) conditional on the rest of the parameter vector $\theta^{(n)}$. Then replace $\theta_{i+1}^{(n)}$ with $\tilde{\theta}_{i+1}$ to get the next draw $\theta^{(n+1)}$. To sample this $\tilde{\theta}_{i+1}$, use the "Metropolis-chain" part. To simplify notation, let $x = \theta_{i+1}^{(n)}$ and $\tilde{x} = \tilde{\theta}_{i+1}$. Suppose that, conditional on the rest of $\theta^{(n)}$, $\tilde{x}$ is to be drawn from a density proportional to $p(\tilde{x}) = .6$ on $[0; .5]$, $p(\tilde{x}) = 1.2$ on $(.5; 1]$ and 0 everywhere else and that one did not already know how to draw from that density. But suppose one knows how to draw a candidate y from the uniform density $f(y) \equiv 1$ on $[0; 1]$ (bad notation—y is not the data as in the preceding explanation, of course). Should you use $\tilde{x} = y$ or $\tilde{x} = x$? For that, calculate $\alpha(x, y)$ using the formula in the article. Here, $\alpha(x, y) = \frac{1}{2}$ if $y \leq .5 < x$ and $\alpha(x, y) = 1$ otherwise. In the latter case, use $\tilde{x} = y$ anyway. In the former case, flip a coin, and take $\tilde{x} = y$ if it lands heads but stay with $\tilde{x} = x$ if it lands tails. It is fun to recursively draw an entire sequence of new x's this way (and easy to see that one does indeed get the right distribution), but this is not done here: Once you have the new $\tilde{x}$, you move on to the next component θ_{i+2} and so forth.

This technique is fast, flexible, and not hard to use and thus a wonderful addition to the toolkit for analyzing stochastic volatility models.

2. STOCHASTIC VOLATILITY MODELS AND THEIR APPLICATIONS

The new literature on stochastic volatility models, including this article, provides an increasingly popular alternative to the widely used variants of ARCH models (see Bollerslev, Chou, and Kroner [1992] for an overview of the latter). Both ultimately aim at modeling the conditional distribution $y_t \mid y_{t-1}, y_{t-2}, \ldots$, where y_t are, say, stock returns. ARCH-type models do so by imposing directly a particular form for the conditional variance $\operatorname{var}(y_t \mid y_{t-1}, y_{t-2}, \ldots)$, whereas stochastic volatility models proceed indirectly, introducing an additional, unobserved state h_t with $\operatorname{var}(y_t \mid h_t, h_{t-1}, \ldots, y_{t-1}, y_{t-2}, \ldots) = h_t$, say, and modeling its evolution as a separate stochastic process. The distribution $y_t \mid y_{t-1}, y_{t-2} \ldots$ can then be found by projecting on the coarser information $y_{t-1}, y_{t-2}, \ldots$— a point forcefully raised by Andersen (1992). This projection can be figured out exactly when assuming a beta distribution for $\exp(v_t)$, a gamma distribution for h_{t-1}, and $\delta = 1$ (see Shephard 1994; Uhlig 1993), but not for the model in this article.

The distinction between ARCH models and stochastic volatility models is not a fundamental one but one of parsimony and of modeling elegance. As for the latter distinction, writing down stochastic processes and figuring out conditional distributions later is what we usually do elsewhere, and it has always seemed to me to be the more natural approach for modeling volatility. As for the former, there is only one "rat race" between these two approaches in the article (Figure 1) and I hope that the authors contribute more to that debate in future articles, using their techniques. Additional interesting comparisons can be found in the work of Kim (1993), for example, albeit using different methods. JPR chose their stochastic volatility model only for illustrating the approach, not as the best model to explain the data. Otherwise one would want to check for autocorrelations among the u_t's and v_t's, check their independence, check the normality assumption and check how well the stock market crash in 1987 is explained by the model and how much the subsequent high volatility is driving δ toward unity for the S&P 500 (see Nelson 1991). All of these things are not hard to do with their algorithm, and it would have been nice had the article contained some of that.

The distribution of $y_{t+j} \mid y_t, y_{t-1}, \ldots, j = 1, 2 \ldots$—that is, the predictive density—is something else I would have liked to see for, say, the end of the sample as well as before and after the stock market crash for the S&P 500. How much does posterior uncertainty regarding ω matter in that regard? These distributions are relevant in particular for pricing options correctly—arguably the most important "real world" application of these methods. They are easy to obtain with the methods of JPR but hard with classical methods. Let me explain. With usual time series models, knowledge of the parameters and the data suffices to calculate all of the residuals and to do out-of-sample forecasting. But here knowledge of ω and the data is not sufficient to back out all of the residuals, u_t's and v_t's and hence the h_t's. Thus they are insufficient if one wants to do out-of-sample forecasting in the obvious way—that is, by simply simulating the model forward—because h_T is not known. One needs to estimate h_T as well, but h_T will not be known precisely, even asymptotically for large T. The out-of-sample forecast needs to take account of that uncertainty in h_T without being able to rely on large-sample theory—something that should make classicists' stomachs turn. The ugly alternative for a classicist is to estimate the conditional distribution $y_{t+j} \mid y_t, u_{t-1}, \ldots$ directly—but then, why use stochastic volatility models? By contrast, it is completely natural for the Bayesian approach to take into account all of the uncertainty relevant for forecasting and to treat it on equal footing—that is, the posterior uncertainty about ω and h_T as well as the uncertainty about

future u_t's and v_t's. It requires hardly any extra work and can be as exact as one wants it to be with finite data. Furthermore, the Bayesian approach answers questions like "what is the probability that the stock price drops by 10%, given the observed data?" rather than "what is the estimator $\hat{p}$ for the probability p of a 10% drop in stock prices so that, given many realizations of the sample path, this estimator would be right on average?" Is it not obvious that an investor would usually be interested in the answer to the first rather than the second question? All of these are powerful arguments to switch to a Bayesian point of view, and I am a bit astonished that JPR do not say so aloud.

3. COMPETITORS

JPR show how their method does better than the MM or QML methods. That is good news, but it should not be all that surprising, given that MM is usually viewed as a more robust but less powerful method and that QML actually uses the "wrong" likelihood function by approximating the log-chi-squared distribution of $\epsilon_t \equiv \log u_t^2$ with a normal distribution. An interesting question is whether there are situations in which the robustness of MM is more valuable—for example, if one wants to make weak assumptions about the distributions of u_t and v_t. Note, however, that one needs distributional assumptions to even identify u_t and v_t. The comparison between the three techniques is somewhat odd anyway because Bayesian estimators are usually not meant to be, say, unbiased or have good repeated-sample properties. This comes about because with such a large sample there is (as usual) little distinction between Bayesian and classical inference regarding the parameter vector ω. In other words, these results show that the cyclic Metropolis-chain algorithm works better and not that one should be a Bayesian. There are other good reasons for that, and they have been explained previously.

I would like to mention two other competitors. One has been developed since JPR wrote their article and is based on approximating the distribution of ϵ_t not with one normal but with a mixture of normal distributions (see Kim 1993; Kim and Shephard 1993). This can be done to any desired degree of accuracy. Augment the data with random variables j_t, selecting the normal distribution from which ϵ_t is drawn. Conditional on the j's one can apply fast, standard Kalman-filtering techniques to find everything else, and conditional on everything else, it is not hard to sample the j_t's. The spirit of that approach is very Bayesian, although these authors try hard to be classical in the end. The method can be as accurate as desired, and it is likely to be fast due to the use of the Kalman filter.

The other approach, due to Shephard (1994) and Uhlig (1993), is dismissed perhaps a bit too quickly in JPR's introduction, but is easy to do even in the multivariate context $Y_t \in \mathbf{R}^m$, provided there are no intercept terms in the equation for Y_t (as is the case considered by JPR). Suppose that $Y_t = \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, H_t^{-1})$, where the precision matrix H_t evolves according to $H_t = \mathcal{U}(H_{t-1})'\Theta_t\mathcal{U}(H_{t-1})/\lambda$ with $\Theta_t \sim \mathcal{B}_m(\nu/2, 1/2)$ drawn from a multivariate, singular beta distribution, defined by Uhlig (1994), and where $\mathcal{U}(H)$ denotes the upper Cholesky factor of a positive definite matrix H. For the univariate case, take logs of the equation for H_t to see that this corresponds to the case of $\delta = 1$ and $\exp(v_t)$ following a univariate beta distribution in the notation of JPR. If the prior in ν, λ, and H_1 is proportional to $\psi_0(\nu, \lambda)\kappa_W((\nu S_0)^{-1}, \nu)\, f_W(H_1 \mid (\nu S_0)^{-1}, \nu)$ for some positive function ψ_0, then the posterior in ν, λ, and H_{T+1} is proportional to $\psi_T(\nu, \lambda)\kappa_W((\nu S_T)^{-1}, \nu)\, F_W(H_{T+1} \mid (\nu S_T)^{-1}, \nu)$, where $f_W(H \mid \Omega, \nu)$ denotes a Wishart density with mean $\nu\Omega$ and ν degrees of freedom, where $\kappa_W(\Omega, \nu) = 2^{m\nu/2}\,\Gamma_m(\nu/2)\,|\,\Omega\,|^{\nu/2}$ collects all of the constants of that density and where one needs to recursively update, for each λ and ν, the inverse of the mean of the Wishart densities according to $S_t = \lambda S_{t-1} + \lambda Y_t Y_t'/\nu$ and the function ψ according to $\psi_t(\nu, \lambda) = \Gamma_m((\nu+1)/2)/\Gamma_m(\nu/2)\,\lambda^{m\nu/2}\,|\,\nu S_t/\lambda\,|^{-1/2}\,\psi_{t-1}(\nu, \lambda)$. The posterior in ν, λ, and H_{T+1} is all one needs for calculating predictive densities. The predictive density for Y_{T+1}, for example, is a multivariate t distribution. For inference about ν and λ, simply drop the $f_W(H_{T+1} \mid (\nu S_T)^{-1}, \nu)$ piece. This is probably the quickest method and needs hardly any numerical techniques at all. Although the method does not easily generalize, whereas the techniques of JPR do, it can provide a useful benchmark and a starting point for understanding situations with stochastic volatility.

My guess is that stochastic volatility models and these three methods for analyzing them will prove useful in many applications to come. The main stumbling blocks for using these models have been removed.

ADDITIONAL REFERENCES

Andersen, T. G. (1992), "Volatility," Working Paper 144, University of Chicago, Kellog Graduate School of Management, Dept. of Finance.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Kim, S. (1993), "Persistence in Conditional Variance: Estimation Using a General Stochastic Volatility Framework," unpublished draft, Princeton University, Dept. of Economics.

Kim, S., and Shephard, N. G. (1993), "Stochastic Volatility: New Models and Optimal Likelihood Inference," unpublished draft, Nuffield College, Oxford University.

Shephard, N. G. (1994), "Local Scale Models—State Space Alternative to Integrated GARCH Processes," *Journal of Econometrics*, 60, 181–202.

Uhlig, H. (1993), "Bayesian Vectorautoregressions With Stochastic Volatility," unpublished draft, Princeton University, Dept. of Economics.

——— (1994), "On Singular Wishart and Singular Multivariate Beta Distributions," *The Annals of Statistics*, 22, 395–405.