Comment

Michael JOHANNES
Graduate School of Business, Columbia University, New York, NY 10027 (*mj335@columbia.edu.*)

Nicholas POLSON
Graduate School of Business, University of Chicago, Chicago, IL 60637 (*ngp@gsbngp.uchicago.edu.*)

Pastorello, Patilea, and Renault (henceforth PPR) are to be congratulated for tackling an important and central question in financial econometrics: estimating nonlinear models in the presence of latent state variables and parameters from observed prices. This is challenging, because the state variables are typically unobserved and are related to the underlying asset prices in a complicated way through the pricing equation.

The basic idea of the article is straightforward: In many cases, an asset pricing model implies that state variables, X, or factors are known, that is, can be inverted conditional on the parameters, Θ, and observed asset prices, Y. Because of this, estimation methods can be tailored to take advantage of this property. PPR provide a general estimation strategy for simultaneous parameter and state variable estimation based on the intuition in the EM algorithm, and they illustrate the method with a term structure model.

In theory, if an asset pricing model is correct, the prices are fully revealing of both states and parameters. In practice, we rarely believe models to be exact, but instead we view them as useful approximations for learning about the nature of asset price dynamics. Thus in practice we often add pricing errors, breaking the stochastic singularity and formally recognizing that our models are approximations. A strength of PPR's approach is that it formally addresses the complicated issues that arise with pricing errors and nuisance parameters.

In their Section 1, PPR discuss the relative merits their approach vis-a-vis the Markov chain Monte Carlo approach as proposed and described by Johannes and Polson (2002). In our Comment, we therefore focus on a comparison of the two methodologies. To frame our discussion, we focus on financial applications and two specific cases: a stochastic volatility option pricing model and a single-factor term structure model. We conclude with a discussion of the futures challenges for both methods.

First, we describe the estimation problem with two concrete examples.

Square-Root Stochastic Volatility. Consider the Heston (1993) model of an equity price, S_t, whose variance, V_t, follows a square-root process,

$$\frac{dS_t}{S_t} = (r_t + \eta_v V_t)\,dt + \sqrt{V_t}\,dW_t^s(\mathbb{P}) \tag{1}$$

and

$$dV_t = \kappa_v(\theta_v - V_t)\,dt + \sigma_v\sqrt{V_t}\,dW_t^v(\mathbb{P}), \tag{2}$$

where W_t^s and W_t^v are two Brownian motions with correlation ρ_v and r_t is the instantaneous spot interest rate. Under an equivalent martingale measure, $\mathbb{Q}$, the dynamics are given by

$$\frac{dS_t}{S_t} = r_t\,dt + \sqrt{V_t}\,dW_t^s(\mathbb{Q}) \tag{3}$$

and

$$dV_t = [\kappa_v(\theta_v - V_t) + \lambda_v V_t]\,dt + \sigma_v\sqrt{V_t}\,dW_t^v(\mathbb{Q}). \tag{4}$$

We let $X = \{V_t\}_{t=1}^T$ denote the volatility states, $\Theta^{\mathbb{P}} = (\kappa_v, \theta_v, \sigma_v, \eta_v, \rho_v)$ denote the structural parameters, and $\Theta^{\mathbb{Q}} = \lambda_v$ denote the risk-premium parameter. Together with equity prices, one also observes option prices. Let the price of a call option on S_t, struck at K, with τ days to maturity be given by a function $C_t = g(S_t, V_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P}})$, where the dependence on contract variables is supressed and the function can be computed numerically (see, Duffie 2001). For later reference, we note that there typically exists a volatility state such that conditional on parameters and a market option, one can invert the mapping $g(\cdot)$ to solve for the "implied" spot volatility, V_t^{imp}. In this model, the goal of empirical asset pricing is to learn about X, $\Theta^{\mathbb{P}}$, and $\Theta^{\mathbb{Q}}$ from the observed from the observed asset prices, $Y = \{S_t, C_t\}_{t=1}^T$.

Vasicek's (1977) Term Structure Model. Vasicek's (1997) term structure model assumes that the instantaneous spot rate solves

$$dr_t = \kappa_r(\theta_r - r_t)\,dt + \sigma_r\,dW_t^r(\mathbb{P}) \tag{5}$$

where W_t^r is a scalar Brownian motion. Under an equivalent martingale measure, the spot rate's dynamics are

$$dr_t = [\kappa_r(\theta_r - r_t) + \lambda_r\sigma_r]\,dt + \sigma_r\,dW_t^r(\mathbb{Q}),$$

and continuously compounded, zero-coupon, τ-*maturity* bond yields are affine functions of the spot rate,

$$y_{t,\tau} = \alpha\big(\Theta^{\mathbb{Q}}, \Theta^{\mathbb{P}}, \tau\big) + \beta\big(\Theta^{\mathbb{P}}, \tau\big)r_t,$$

where α and β are known functions, $\Theta^{\mathbb{P}} = (\kappa_r, \theta_r, \sigma_r)$, and $\Theta^{\mathbb{Q}} = \lambda_r$. The authors later find that even on simulated data, it is not possible via MLE to estimate the risk premium parameter. With zero coupon yields, this parameter is likely not identified, see the discussion in the appendix of Dai and Singleton (2001). Again, the goal of empirical asset pricing is to learn about $X = \{r_t\}_{t=1}^T$, $\Theta^{\mathbb{P}}$, and $\Theta^{\mathbb{Q}}$ from an observed panel of bond yields, $Y = \{y_{t,\tau_1}, \ldots, y_{t,\tau_N}\}_{t=1}^T$. As in the case of options, conditional on parameters and a single yield, the latent state, r_t, is revealed.

1. THE MARKOV CHAIN MONTE CARLO APPROACH

MCMC is a simulation approach for implementing Bayesian inference. The Bayes solution to the inference problem is an inverse problem: Extract the information about the state variables (X) and parameters $\Theta = (\Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}})$ from observed asset

Journal of Business & Economic Statistics
???? 0, Vol. 0, No. 0
DOI 10.1198/00

prices, Y. The solution to the inverse problem is just the distribution of the parameters, Θ, and state variables, X, conditional on observed prices, Y, which we denote by $p(\Theta, X|Y)$. Bayesian inference merely uses the rules of conditional probability to characterize $p(\Theta, X|Y)$.

Summarizing inference via this distribution has two main advantages. First, it answers the obvious question of interest: What do the data tell the researcher about the parameters and states? Second, the Bayes approach is consistent with the theory of no arbitrage and provides the correct way to combine the information in the model and the sample. In this case, on observing Y, if one summarizes the information about Θ and X in any other probabilistic fashion, then one is left open to Dutch Book arbitrages (Ramsey 1926; de Finetti 1931; Shimony 1955). It is in this sense in which $p(\Theta, X|Y)$ provides the correct way to base inference for financial econometrics.

Bayes rule characterizes the posterior, $p(\Theta, X|Y)$, as

$$p(\Theta, X|Y) \propto p(Y|X, \Theta)p(X|\Theta)p(\Theta),$$

where $p(Y|X, \Theta)$ is the likelihood, $p(X|\Theta)$ summarizes the evolution of the state variables, and $p(\Theta)$ is the prior over the parameters. We disagree with PPR's criticism of the use of the prior $p(\Theta)$; it is just an implication of the laws of conditional probability! Moreover, it is an internally consistent way to overcome identification issues, impose parameter restrictions, and incorporate economic theory into the inference procedure. For example, simple models such as Merton's (1976) jump diffusion generate infinite likelihoods, which show the importance and necessity of imposing prior information.

To sample from $p(\Theta, X|Y)$, MCMC relies on the Clifford–Hammersley theorem, which states that $p(\Theta, X|Y)$ is completely characterized by the complete conditionals

$$p(X|\Theta, Y), \qquad p\big(\Theta^{\mathbb{P}}|\Theta^{\mathbb{Q}}, X, Y\big), \quad \text{and} \quad p\big(\Theta^{\mathbb{Q}}|\Theta^{\mathbb{P}}, X, Y\big).$$

MCMC algorithms iteratively sample from these distributions; given $X^{(j)}$ and $\Theta^{(j)}$, draw

$$X^{(j+1)} \sim p\big(X|\Theta^{(j)}, Y\big),$$

$$\big(\Theta^{\mathbb{P}}\big)^{(j+1)} \sim p\big(\Theta^{\mathbb{P}}|\big(\Theta^{\mathbb{Q}}\big)^{(j)}, X^{(j+1)}, Y\big),$$

and

$$\big(\Theta^{\mathbb{Q}}\big)^{(j+1)} \sim p\big(\Theta^{\mathbb{Q}}|\big(\Theta^{\mathbb{P}}\big)^{(j+1)}, X^{(j+1)}, Y\big).$$

Continuing in this fashion, this generates a sample, $\{X^{(j)}, \Theta^{(j)}\}_{j=1}^{J}$, whose equilibrium distribution is the posterior. The iterative procedure in MCMC is stochastic; at each stage we draw from a distribution. This implies that the algorithm converges to a distribution and not to a fixed point.

In the models described earlier, one additional facet must be addressed: how to deal with the fact that the latent states, conditional on the parameters, can be inverted. Formally, this implies that $p(X|\Theta^{(j)}, Y)$ is a degenerate distribution. To break the stochastic singularity, MCMC assumes a pricing error. In the case of the option pricing and term structure examples, one could assume additive pricing errors,

$$y_{t,\tau} = \alpha\big(\Theta^{\mathbb{Q}}, \Theta^{\mathbb{P}}, \tau\big) + \beta\big(\Theta^{\mathbb{P}}, \tau\big)r_t + \varepsilon_t \tag{6}$$

and

$$C_t = g\big(S_t, V_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P}}\big) + \varepsilon_t. \tag{7}$$

At this stage, we make three observations. First, note that when updating the structural parameters, drawing from $p(\Theta^{\mathbb{P}}|\Theta^{\mathbb{Q}}, X, Y)$, we use the information in both the vector of states and all of the observed asset prices. As we point out later, this is an important distinction between our approach and that of PPR. What does this imply? Consider the two foregoing examples. In the stochastic volatility model, consider the parameter σ_v. This parameter appears in the structural evolution of the volatility process and also in the option price. Our update uses the information in both of these sources to estimate the parameter and, in this sense, uses all of the available information. Similarly, in the term structure example, κ_r appears in both the bond yields (through the loading functions) and in the evolution of the spot rate. Efficient estimation of the structural parameters uses the information in both of these sources of information. Second, note that updating $\Theta^{\mathbb{Q}}$ uses only the information embedded in the equation with errors, either the option price equation or the yield equation. Conditional on the latent states, the structural evolution equation provides no information regarding the risk premiums.

2. THE PPR APPROACH

The methodology of PPR uses a multistage algorithm to estimate the parameters and states. As PPR point out, there are some similarities between their approach and MCMC algorithms. In this section we describe our interpretation of PPR's algorithm and then discuss important differences.

Consider the following algorithm:

1. Given parameters, $\Theta^{\mathbb{P}}$ and $\Theta^{\mathbb{Q}}$, and asset prices, "estimate" the latent states, $\widehat{X}_t$. For example, in the stochastic volatility or term structure model with a single observation, "estimate" should be interpreted as inversion or, in models with pricing errors, nonlinear least squares. These estimates are referred to as implied states.
2. Given the implied states, "estimate" the structural parameters assuming that the estimated states are observed without error using the structural evolution equations to get $\widehat{\Theta}^{\mathbb{P}}$. For example, in the stochastic volatility model, PPR's approach estimates $\Theta^{\mathbb{P}}$ using the structural system in (1) and (2) and the implied variances from the option prices. Similarly, in the term structure model, given spot rates, estimate $(\kappa_r, \theta_r, \sigma_r)$ via the transitions of (5). Note that in this stage neither option prices nor yields are used directly.
3. Given $\widehat{\Theta}^{\mathbb{P}}$ and $\widehat{X}_t$, estimate the risk-premium parameters using, for example, nonlinear least squares. In the option pricing and term structure models, one would minimize the pricing errors in (6) or (7). This produces estimates, $\widehat{\Theta}^{\mathbb{Q}}$.

After step 3, repeat from step 1 using $\widehat{\Theta}^{\mathbb{P}}$and $\widehat{\Theta}^{\mathbb{Q}}$. Iterating in this manner, one obtains a sequence of estimates. PPR prove that, in the EM algorithm, a properly specified algorithm converges to a fixed point. At first glance, this algorithm seems very similar to the MCMC algorithm. The MCMC algorithm iteratively draws or "estimates" X_t, $\Theta^{\mathbb{P}}$, and $\Theta^{\mathbb{Q}}$. What are the differences?

First, at one level, we know via the Clifford–Hammersley that theorem $p(X|\Theta, Y)$, $p(\Theta^{\mathbb{P}}|\Theta^{\mathbb{Q}}, X, Y)$, and $p(\Theta^{\mathbb{Q}}|\Theta^{\mathbb{P}}, X, Y)$ fully characterize the joint $p(\Theta, X|Y)$, and again, $p(\Theta, X|Y)$ fully characterizes all the information in the data regarding the parameters and states. How does the procedure proposed by PPR work? In the first step, it uses *some* of the information in $p(X|\Theta, Y)$ to estimate the states. It does not guarantee that the factor estimates are consistent with dynamics, for example. In the second step, it discards the information on option prices or yields and estimates based solely on the information in $p(\Theta^{\mathbb{P}}|X)$. In the case of the stochastic volatility model, it bases estimation on $p(\Theta^{\mathbb{P}}|X, S)$, discarding the option prices. Why do we throw out data at this stage? Clearly, this implies that the procedure is combining the data and the latent states in an inefficient manner. MCMC efficiently combines the information in all levels of the state-space model.

Second, at each stage in the algorithm, MCMC uses the information in entire conditional posteriors by drawing from the distributions, $p(X|\Theta, Y)$, $p(\Theta^{\mathbb{P}}|\Theta^{\mathbb{Q}}, X, Y)$, and $p(\Theta^{\mathbb{Q}}|\Theta^{\mathbb{P}}, X, Y)$. In contrast, the approach here optimizes at each stage, which, in the likelihood setting, is similar to taking the posterior model at each step. While leading to asymptotically valid inference, this may also lead to algorithms that degenerate or get stuck in a local optima. This is one of the reasons that care must be used when applying the EM algorithm; in difficult problems it converges, but to the wrong answer!

Third, the PPR procedure does not appear to be as general as MCMC in terms of estimating latent states. For example, how does the algorithm discriminate between state variables and factors? In equity price models used for option pricing, volatility is typically a state variable, whereas jumps in returns and jumps in volatility are just factors, because they are integrated out of the option price. Can the PPR procedure handle this case? There is no mapping between observed returns or volatility and jumps, and thus the iterative procedure would require that the latent factors (as opposed to states) be integrated out of the likelihood function. Once the latent factors are integrated out, they are difficult to estimate. Maybe PPR could discuss this case in more detail. The unobservability of any state or factor is no problem for MCMC. For example, in the case of equity price models with jumps, Eraker, Johannes, and Polson (2002) and Eraker (2002) have show how MCMC can estimate both states and factors.

In conclusion, the approach discussed in this article provides a general approach for estimating a class of models with certain types of latent states. The approach is similar in spirit to MCMC algorithms, because it is an iterative algorithm, but is substantively different along a number of dimensions. In the simple term structure model that PPR consider, the method appears to be nearly as efficient as ML. However, this example does not provide a strong test of the methodology. It therefore would be interesting to see how the algorithm performs on more difficult models, such as multifactor term structure models, equity index-option models with jumps, and regime-switching models, all of which can be easily tackled with MCMC (see, e.g., Johannes and Polson 2002; Lamoureux and Witte 2002; Polson and Stroud 2003; Eraker 2003). In the end, the real challenge for the PPR approach and others approaches like MCMC is to provide computationally attractive tools that solve problems of interest for empirical asset pricers.

[Received ??????. Revised ??????.]

ADDITIONAL REFERENCES

Abner, S. (1955), "Coherence and the Axioms of Confirmation," *The Journal of Symbolic Logic*, 20, 1–28.

de Finetti, B. (1931), "On the Subjective Meaning of Probability," in *Probabilitá e Induzione*, eds. P. Monari and D. Cocchi, Bologna, Italy: Clueb, 1993, pp. 291–321.

Duffie, D. (2001), *Dynamic Asset Pricing Theory* (3rd English ed.), Princeton, NJ: Princeton University Press.

Lamoureux, C., and Witte, D. (2002), "Empirical Analysis of the Yield Curve: The Information in the Data Viewed Through the Window of Cox, Ingersoll, and Ross," *Journal of Finance*, 57, 1479–1520.

Polson, N., and Stroud, J. (2003), "Bayesian Inference for Derivative Prices," in *Bayesian Statistics 7*, eds. J. Bernardo, et al., Oxford, U.K.: Oxford University Press, pp. 641–650.

Ramsey, F. P. (1931), "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, London: Routledge and Kegan Paul, pp. 156–198. Reprint: Kyberg, H. E. Jr., and Smokler, H. E. (eds.) (1964), *Studies in Subjective Probability*, New York: Wiley, pp. 61–92.