Comment on Garland B. Durham and A. Ronald Gallant's "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes"

Siddhartha Chib
John M. Olin School of Business, Washington University, St Louis, MO 63130, USA
Neil Shephard
Nuffield College, Oxford OX1 1NF, UK

August 12, 2001

1 Introduction

This paper proposes an interesting approach for estimating the parameters of nonlinear diffusion models with discretely sampled data. The parameter estimates are obtained by maximizing an approximate likelihood function that is obtained by a Monte Carlo importance sampling method. As the authors point out, the elements of their approach are not substantially new. In particular, the idea of approximating the transition density of the process by integrating out the "missing values" between each successive observation is due to Pedersen (1995); the use importance sampling to estimate the likelihood in which the missing values are drawn from a "tied-down" distribution is due to Elerian, Chib, and Shephard (2001); and finally, the idea of transforming the process to one with a constant diffusion coefficient in order to improve the accuracy of the Euler approximation scheme is due to Doss (1977).

As we see it, the paper makes two main contributions. First, it provides a detailed comparison of various extant methods for estimating the likelihood function of univariate diffusion models. Elerian (1999) has done related work along the same lines. Taken together these two papers have enhanced our understanding of what methods are effective for approximating the likelihood function. We are pleased that the value of such comparative work has been recognized by this journal. Second, the paper provides a new proposal density for the importance sampling step

that appears to be numerically quite efficient. Because this is perhaps the central contribution of the paper we begin our discussion by presenting the Durham – Gallant (DG) proposal density in a form that has intuitive appeal. We then conclude with comments on the connections of this work to other recent research.

2 DG proposal density

2.1 Importance sampler

To begin, in the notation of the paper, write the Markov process for the auxiliarly observations as

$$u_{m+1}|u_m \sim \mathrm{N}(u_m + \mu(u_m;\theta)\delta, \sigma^2(u_m;\theta)\delta)\,, \tag{1}$$

where δ is the time gap. The objective is to evaluate the integral in equation (3) of the paper. Pedersen (1995) suggests that the integral be approximated by drawing samples on $(u_1, ..., u_{M-1})$ by simulating the above process. The implied importance sampling function is, therefore, just the product of these one-step ahead transition densities. Elerian, Chib, and Shephard (2001) pointed out that the Pedersen importance function ignores the end-point information u_M and that a better importance function can be developed by utilizing this information. The optimal importance function is given by the non-linear or tied-down process,

$$p(u_1, ..., u_{M-1}|u_0, u_M) = \left\{ \prod_{m=1}^{M} p(u_m|u_{m-1}) \right\} / p(u_M|u_0), \tag{2}$$

under which the importance sampling estimate of the likelihood would have zero variance even with a single draw. Of course this choice is not available because it requires knowledge of $p(u_M|u_0)$, which if known, would directly lead to an estimate of the likelihood. In practice we need to approximate $p(u_1, ..., u_{M-1}|u_0, u_M)$ with a distribution that is both easy to sample and compute. One possibility, proposed by Elerian, Chib, and Shephard (2001), is a multivariate Gaussian or Student-t approximation based on the mode and Hessian of the target distribution. This approximation is quite efficient in general although the search for the mode imposes a computational burden that can be high when M is large.

In the current paper another approximation is suggested which we think is worth discussing

in full. From (2) we can write (noting the cancelation in the first line)

$$\begin{aligned} p(u_1,...,u_{M-1}|u_0,u_M) &= \left\{\prod_{m=1}^{M-1} p(u_m|u_{m-1})\right\}\left[\left\{\prod_{m=1}^{M-1} p(u_M|u_m)\right\}/\left\{\prod_{m=1}^{M-1} p(u_M|u_{m-1})\right\}\right] \quad (3)\\ &= \prod_{m=1}^{M-1}\{p(u_m|u_{m-1})p(u_M|u_m)/p(u_M|u_{m-1})\}\\ &= \prod_{m=1}^{M-1} p(u_m|u_{m-1},u_M). \quad (4)\end{aligned}$$

Next DG work with

$$p(u_m|u_{m-1},u_M) = p(u_m|u_{m-1})p(u_M|u_m)/p(u_M|u_{m-1}), \quad (5)$$

were $p(u_m|u_{m-1})$ is known and Gaussian but we cannot, by definition, know $p(u_M|u_m)$ or $p(u_M|u_{m-1})$, as it is the same problem as calculating $p(u_M|u_0)$. As a result the above is not feasible. At this point DG approximate the above by using a standard Euler approximation to $u_M|u_{m-1}$, that is

$$u_M|u_{m-1} \sim \mathrm{N}(u_{m-1}+\mu(u_{m-1};\theta)\delta^+, \sigma^2(u_{m-1};\theta)\delta^+), \quad (6)$$

but a more non-standard

$$u_M|u_m \sim \mathrm{N}(u_m+\mu(u_{m-1};\theta)\delta^*, \sigma^2(u_{m-1};\theta)\delta^*). \quad (7)$$

The latter approximation is used in order to get tractability. We write these approximate densities as $\overline{p}(u_M|u_{m-1})$ and $\widehat{p}(u_M|u_m)$. By Bayes theorem, we now have a Gaussian approximation to $u_m|u_{m-1},u_M$ which can be written as

$$\overline{p}(u_m|u_{m-1},u_M) = p(u_m|u_{m-1})\widehat{p}(u_M|u_m)/\overline{p}(u_M|u_{m-1}), \quad (8)$$

leading to the feasible importance sampler

$$\overline{p}(u_1,...,u_{M-1}|u_0,u_M) = \prod_{m=1}^{M-1}\overline{p}(u_m|u_{m-1},u_M) \quad (9)$$

This is attractive as this is a Gaussian density which is easy to simulate and evaluate. Note that we can rewrite the latter density as

$$\overline{p}(u_1,...,u_{M-1}|u_0,u_M) = \left\{\prod_{m=1}^{M-1} p(u_m|u_{m-1})\right\}\left[\left\{\prod_{m=1}^{M-1}\widehat{p}(u_M|u_m)\right\}/\left\{\prod_{m=1}^{M-1}\overline{p}(u_M|u_{m-1})\right\}\right] \quad (10)$$

under which the resulting estimate of the likelihood contribution is

$$\begin{aligned} p(u_M|u_0) &= \int \frac{p(u_1,...,u_{M-1},u_M|u_0)}{\overline{p}(u_1,...,u_{M-1}|u_0,u_M)}\overline{p}(u_1,...,u_{M-1}|u_0,u_M)\mathrm{d}u_1...,\mathrm{d}u_{M-1} \\ &= \int \overline{p}(u_M|u_{M-1})\frac{\prod_{m=1}^{M-1}\overline{p}(u_M|u_{m-1})}{\prod_{m=1}^{M-1}\widehat{p}(u_M|u_m)}\overline{p}(u_1,...,u_{M-1}|u_0,u_M)\mathrm{d}u_1...,\mathrm{d}u_{M-1} \\ &= \widehat{p}(u_M|u_0)\int \frac{\prod_{m=0}^{M-1}\overline{p}(u_M|u_m)}{\prod_{m=0}^{M-1}\widehat{p}(u_M|u_m)}\overline{p}(u_1,...,u_{M-1}|u_0,u_M)\mathrm{d}u_1...,\mathrm{d}u_{M-1} \\ &= \widehat{p}(u_M|u_0)\mathrm{E}_{u|u_0,u_M}\left\{\frac{\prod_{m=0}^{M-1}\overline{p}(u_M|u_m)}{\prod_{m=0}^{M-1}\widehat{p}(u_M|u_m)}\right\}. \end{aligned} \tag{11}$$

This has the rather elegant interpretation of being the Euler approximation of $p(u_M|u_0)$, which we have written as $\widehat{p}(u_M|u_0)$, times the expected value of a likelihood ratio of two predictive models. In practice we evaluate the expectation by simulating from $\overline{p}(u_1,...,u_{M-1}|u_0,u_M)$, which is straightforward.

2.2 Markov chain Monte Carlo based on DG's proposal density

Of course the DG proposal density could also be used in the context of a Bayesian analysis of the non-linear diffusion model. Indeed it fits rather naturally into the Markov chain Monte approach of Elerian, Chib, and Shephard (2001). Suppose the current state of the auxiliary points is $u_1^c,...,u_{M-1}^c$ and that we make a proposal $u_1^{new},...,u_{M-1}^{new}$ from

$$\overline{p}(u_1,...,u_{M-1}|u_0,u_M), \tag{12}$$

then the Metropolis-Hastings acceptance probability (Chib and Greenberg (1995)) would be

$$\begin{aligned} & \min\left[1,\frac{p(u_1^{new},...,u_{M-1}^{new},u_M|u_0)}{p(u_1^c,...,u_{M-1}^c,u_M|u_0)}\frac{\overline{p}(u_1^c,...,u_{M-1}^c|u_0,u_M)}{\overline{p}(u_1^{new},...,u_{M-1}^{new}|u_0,u_M)}\right] \\ = & \min\left[1,\frac{\prod_{m=0}^{M-1}\overline{p}(u_M|u_m^{new})}{\prod_{m=0}^{M-1}\widehat{p}(u_M|u_m^c)}\right]. \end{aligned} \tag{13}$$

Elerian, Chib, and Shephard (2001) also consider Markov chain algorithms in which $(u_1,...,u_{M-1})$ are updated in smaller, random blocks. The above proposal density can be adapted for that case as well. The advantage of DG proposal density over the Elerian, Chib, and Shephard (2001) proposal density is that it avoids the numerical evaluation of the mode.

3 Other points

Section 8 of the paper extends the analysis to cover the case of a partially observed diffusion — stochastic volatility. Unfortunately we found the material concerning propagation difficult to follow. We believe that the authors are employing the adapted particle filter approach discussed by Pitt and Shephard (1999). Here we give a slightly different derivation of this solution, adapting the above notation to the bivariate case with $u_0 = (X_{i-1}, H_{i-1})$ and $u_M = (X_i, H_i)$.

In a particle filter we approximate the density of $H_{i-1}|\mathcal{F}_{i-1}$ by a sample of particles (think bootstrap!), written $H_{i-1}^1, ..., H_{i-1}^N$. If we let $u_0^j = (X_{i-1}, H_{i-1}^j)$, then our task is to find a way of simulating from $H_i|\mathcal{F}_i$ in order to produce a desired new sample $H_i^1, ..., H_i^N$. We do this by augmentation, sampling N times from

$$\widetilde{p}(j, u_1, ..., u_M|\mathcal{F}_i) \propto p(u_M|u_{M-1})...p(u_2|u_1)p(u_1|u_0^j), \quad j = 1, 2, ..., N, \tag{14}$$

(keeping X_i fixed) and then throwing away all the draws except $u_M^j = (X_i, H_i^j)$. As N goes to infinity these draws converge to samples from $H_i|\mathcal{F}_i$ for this is a particle approximation to

$$\widetilde{p}(u_1, ..., u_M|\mathcal{F}_i) \propto p(u_M|u_{M-1})...p(u_2|u_1) \int p(u_1|u_0)p(u_0|\mathcal{F}_{i-1})\mathrm{d}u_0. \tag{15}$$

The only remaining propagation question concerns the sampling of $\widetilde{p}(j, u_1, ..., u_M|\mathcal{F}_i)$.

Sampling from $\widetilde{p}$ can be carried out under the DG proposal by using a bridge process for the observations (as above) and an unconstrained sampler for the volatilities. If we write the DG sampler as $\overline{p}(u_1, ..., u_{M-1}, u_M|u_0, X_i)$ then the auxiliary particle filter method has the following algorithmic form

1. Set $k = 1$.

2. Choose j randomly from $1, ..., N$. Call this draw j_k.

3. Sample and store

$$u_1^{j_k}, ..., u_{M-1}^{j_k}, u_M^{j_k} \quad \text{from} \quad \overline{p}(u_1, ..., u_{M-1}, u_M|u_0^{j_k}, X_i) \tag{16}$$

and compute

$$w_k = \frac{p(u_M^{j_k}|u_{M-1}^{j_k})...p(u_2^{j_k}|u_1^{j_k})p(u_1^{j_k}|u_0^{j_k})}{\overline{p}(u_1^{j_k}, ..., u_{M-1}^{j_k}, u_M^{j_k}|u_0^{j_k}, X_i)}. \tag{17}$$

4. Repeat 1 until $k = R$.

5. Resample the from the particles $\left(u_1^{j_1}, ..., u_{M-1}^{j_1}, u_M^{j_1}\right), ..., \left(u_1^{j_R}, ..., u_{M-1}^{j_R}, u_M^{j_R}\right)$ with probabilities proportional to $w_1, ..., w_R$.

Typically we choose $R = 10N$, then the theory of particle filters implies that as N goes to infinity this provides a valid sample of size N from $H_i|\mathcal{F}_i$.

The resampling operation given above is important but it introduces considerable variability in output from the particle filter, implying that the corresponding estimated likelihood function will not be differentiable everywhere. Recently Pitt and Walker (2001) gave a solution to this problem in the context of the univariate non-Gaussian OU based SV models developed by Barndorff-Nielsen and Shephard (2001) (which has such a convenient structure that data augmentation can be avoided entirely). Their method, which has been developed for state space models with univariate states by Michael Pitt in unpublished work, seems similar to the use of Hermite functions at the end of Section 8. Pitt's solution does not easily extend to multivariate states. Could the authors comment on whether this will be true of their method?

Finally, we conclude by noting that the use of augmentation to deal with likelihood based inference for continuous time SV models was first suggested by Kim, Shephard, and Chib (1998, Section 6.2).

4 Acknowledgments

Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839.

References

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican 49*, 327–35.

Doss, H. (1977). Liens entre equations differentielles stochastiques et ordinaires. *Annals de Institute Henri Poincare 13*, 99–125.

Elerian, O. (1999). Simulation estimation of continuous-time models with applications to finance. Unpublished D.Phil. thesis, Nuffield College, Oxford.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69*, 959–993.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations on discrete observations. *Scandinavian Journal of Statistics 27*, 55–71.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Pitt, M. K. and S. Walker (2001). Comment on Ole E Barndorff-Nielsen and Neil Shephard's paper on non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 228–229.