Comment on ”Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes” by G. B. Durham and A. R. Gallant

Bjørn Eraker

September 17, 2001

1 Introduction

I am grateful for the opportunity to comment on this very readable and interesting paper by professors Durham and Gallant. The problem of estimation and inference for discretely observed diffusions has been studied extensively. Since the paper by Durham and Gallant also uses methods that have previously been used in other contexts, it is surprising to see the extent to which their method seems to improve upon the computational speed of existing methods. This is an important achievement. In particular, for academics or practitioners that needs to compute parameters in large scale models or across a high number of time-series, their contribution may prove especially relevant.

In my comment I will focus on two aspects of the methods discussed in Durham and Gallant: First, I discuss the three importance sampling algorithms used. Of particular interest is the question of whether or not the importance densities in question provide sufficiently good approximations to the target densities to guarantee convergence of the algorithms. The second point of this comment is the construction of volatility filtering algorithms for continuous time SV models. Here I am particularly interested in how the particle filtering approach favored by DG compares to (posterior) smoothing algorithms used in Bayesian inference. Unless otherwise noted, I use the notation in DG.

2 Convergence

It is well recognized that importance sampling techniques and MCMC algorithms such as the Metropolis Hastings algorithms, require certain regularity conditions to converge. In Durham and Gallant's article, the important regularity condition is given by assumption 4. The following passage from Geweke (1989), p. 1319, illustrates the importance of this condition in the context of a Bayesian estimation problem:

> *Loosely speaking, the importance sampling density should mimic the posterior density, and its especially important that the tails of $I(\theta)$ not decay more quickly than the tails of $p(\theta)$. ... Investigators have experienced substantial difficulties in tailoring importance sampling densities to the problem at hand. This is an important failing, for the approach is neither attractive of methodical if special arcane problems in numerical analysis have to be resolved in each application.*

Therefore, it is somewhat unsatisfactory that DG assume that this condition holds, rather than explicitly show that is holds for the importance densities they use.

As mentioned in Geweke, the regularity condition [assumption 4] amounts to a restriction on the decay rate of the tails of the two densities. In the case of diffusions, this means that the tails of the transition density defined by the importance sampler cannot decay much more rapidly than the tails of the transition density. Assumption 4 states that the ratio of the $M-1$ dimensional target and importance densities needs to have finite expectation,

$$E\left[\frac{\prod_{m=1}^{M} p^{(1)}(U_m, \tau_m; U_{m-1}, \tau_{m-1}, \theta)}{q(U_1, .., U_{M-1})}\right] < \infty.$$

The expectation is defined with respect to joint density $p(U_M, .., U_1 \mid U_0, \theta) := \prod_{m=1}^{M} p^{(1)}(U_m, \tau_m; U_{m-1}, \tau_{m-1}, \theta)$. Notice that while each term in the product sum is conditionally Gaussian, the joint distribution $p(U_M, .., U_1 \mid U_0, \theta)$ is not Gaussian except for the special case of an O-U process. Indeed, the class of target densities, $p(U_M, .., U_1 \mid U_0, \theta)$ generated by diffusions is very large, and therefore it is clear that some tailoring of the importance sampler to each problem is necessary. In particular, it should be pointed out that some fixed Gaussian density **will not** generally satisfy assumption 4.

With this in mind, I now discuss the three samplers considered by DG:

1. The Pedersen sampler trivially satisfy assumption 4 since $p(U_M, .., U_1 \mid U_0, \theta) = q(U_M, .., U_1 \mid U_0, \theta)$.

2. The Elerian, Chib, and Shephard (ECS) algorithm draws from a multivariate normal density based on a second order taylor expansion of the target density. The importance density defined by the ECS sampler will have first and second moments matching the target, but its tails will decay more quickly than those of the target density. Therefore, the ECS sampler generally violates assumption 4.

   ECS also propose to use a multivariate t density, presumably to deal with the coverage problem. Their paper offers no proof as to whether this t density has sufficient regularity.

3. The Brownian Bridge sampler advocated by DG will, in general, not imply

a gaussian importance density. While this is good, it still needs to be proven whether or not the BB sampler satisfies assumption 4. Note however that the BB sampler is a multivariate generalization of the proposal density $\phi(u_m; \frac{1}{2}(u_{m-1} + u_{m+1}), \frac{1}{2}\tilde{\sigma}^2)$ used in Eraker (2001). In this paper, however, I proposed an MCMC sampler that updates U_m one at the time - a simpler problem than drawing $U_m, m = 1, .., M-1$ simultaneously. This one-at-a-time Gibbs sampler has the fortunate property that the target density converges to this normal density as M increases. This adds credibility to the choice of proposal density, at least for M large. Unfortunately, the result does not carry over the the multivariate setting, and as I have argued, the target density is likely to be highly non-gaussian in this case.

In light of the above discussion, it is first of all clear that it is not difficult to construct sampling algorithms based on importance or MCMC sampling that satisfy the necessary regularity condition. The Pedersen sampler does. It can be shown that the MCMC sampler proposed in Jones (1998) also trivially satisfies the condition. The difficulty lies in constructing samplers that satisfy the condition **and** are fast to sample from. In the setting of Durham and Gallant, the problem calls for drawing a $M-1$ random vector that ideally bridges between successive observation pairs. Such bridging processes are designed with the purpose of restricting the support of the sampled paths. While such restrictions apparently improve upon the efficiency of the algorithm, restricting the support too much can lead to violations of the necessary regularity condition. Therefore, one might argue, there is a tradeoff between computational efficiency and robustness of the algorithm. I expect future work in this area to address this tradeoff.

3 The filtering problem in SV models

I now turn to the issue of estimating continuous time stochastic volatility (SV) models. The difficulty facing estimation for these models is the unobservability of the volatility process, $H_{t_i} =: H_i$. To focus on the issue of the latent volatility, assume (without loss of generality) that the observations $X_i, i = 0, .., n$ are sampled sufficiently closely to eliminate discretization bias. The likelihood function corresponding to the discretized diffusion, is

$$l_{X,H}(X, H; \theta) = \prod_{i=1}^{n} p(X_i, H_i \mid X_{i-1}, H_{i-1}, \theta).$$

Whether Bayesian or classical likelihood maximization is the objective, the latent volatility needs to be integrated out of the joint density/ likelihood function

$$l_X(X; \theta) = \int l_{X,H}(X, H; \theta) dH$$

By Bayes theorem, it can be seen that the marginal likelihood of X can be obtained as

$$l_X(X;\theta) = E\prod_{i=1}^{n} p(X_i \mid H_i, \mathcal{F}_i, \theta) \tag{1}$$

where the expectation is taken w.r.t. the conditional distribution of latent volatility,

$$p(H \mid X, \theta) \propto \prod_{i=1}^{n} p(H_i \mid \mathcal{F}_i, \theta). \tag{2}$$

and $\mathcal{F}_i$ is the information set generated by $(X_0, H_0), .., (X_i, H_i)$.

The expectation in (1) can be approximated by a Monte Carlo estimator which requires samples from $p(H \mid X, \theta)$. To sample the ith element of H from this density we can eliminate the terms where H_i does not enter directly to obtain

$$p(H_i \mid H_{i-1}, H_{i+1}, X_i, \theta) \propto p(H_i \mid H_{i-1}, \theta)p(H_{i+1} \mid H_i, \theta)p(X_i \mid H_i, \theta). \tag{3}$$

Simulations from this density are sometimes referred to as *smoothed* or posterior simulations. Notice that both Bayes and ML estimation requires simulation from the density in (3).

Durham and Gallant propose to sample the latent volatilities by an algorithm known as *particle filtering*. Particle filtering requires H_i to be drawn from a density proportional to

$$p(H_i \mid X_i, H_{i-1}) \propto p(H_i \mid H_{i-1}, \theta)p(X_i \mid H_i, \theta) \tag{4}$$

There are several noteworthy differences between the smoothing filter obtained by (3) and particle filters (4):

Consider a real time filtration. At the arrival of a new observation, the particle filter calls for estimating H_{i+1} conditional upon the past values of H_i and the new observation X_{i+1}. The past filtered values, $H_i, H_{i-1}, ...$ are not updated at the arrival of the new information. Conversely, smoothing filters seek to update both the current and past state values $H_j, j = 0, .., i+1$ conditional upon the arrival of new information X_{i+1}.

While in practice the filters perform similar tasks, they work very differently: Consider for instance the event of a large market move. The smoothing filter calls for a revision of volatility estimates prior to the move, essentially providing a smoothed volatility trajectory consistent with the large move. The particle filter does not call for such a revision, leaving the volatility estimate to jump in response to the new information. Particle filters therefore produce more erratically behaving volatility paths.

In light of this discussion, it is interesting to examine the simulation evidence presented in DG's table 4 where the performance of the estimator is compared to

Table 1: Monte Carlo study for SV model: n=2,000 observations

	θ_1	θ_2	θ_3	θ_4	θ_5
true	0.00020	-0.00200	-0.30000	-0.03000	0.30000
			Durham & Gallant		
mean	0.00029	-0.00294	-0.34060	-0.03432	0.31425
RMSE	0.00013	0.00142	0.09355	0.00943	0.02849
			Eraker (2001)		
mean	0.00031	-0.00316	-0.32770	-0.03262	0.29906
RMSE	0.00017	0.00176	0.07807	0.00765	0.02880

results in Eraker (2001). Remember here that the parameters of the volatility process are identified only from estimates of the volatility process, and as such, they reflect the different properties of the particle and smoothing filters. θ_4 and θ_5, which measure the persistence and volatility of H, are both estimated to be higher by DG than in my paper. This reflects the fact that filtered volatility estimates are more erratic when estimated using the particle filter, than what is the case when using the posterior smoothing with MCMC.

Table 1 extends the results in table 4 in DG by presenting simulation evidence obtained using a larger sample-size ($n = 2,000$). Results for the DG estimator were kindly provided to me by Garland Durham. As is shown in table 1, the estimators both perform reasonably well when 2000 data points are used for estimation. The more erratic nature of the particle filter is still evident in the results based on 2,000 observations as the downward bias in θ_4 is larger for the DG estimator.

In a final comment on the choice of filtering algorithms, it is worthwhile to point out that the particle filtering algorithm greatly simplifies the estimation problem. Simulations via the particle filter is straightforward and can be implemented through the forward recursive scheme used in DG. Posterior simulations on the other hand, is much more difficult because the ith volatility depends on its future realizations. In practice, it is difficult to handle this type of dependence without the iterative simulation schemes such as those offered by MCMC methods. The iterative nature of MCMC does, on the other hand, typically imply a considerably heavier computational burden than particle filters. Therefore, the use of particle filters partially explains why DG are able to obtain the attractive running times reported in their paper.

4 Conclusion

The paper by Durham and Gallant is an important step towards designing computationally efficient simulation based estimators of diffusion processes observed at discrete times. The generality of this problem makes the construction of generally applicable methods a daunting task. I have argued that the desire to design sampling schemes that restrict the support of the importance densities in order to improve upon computational efficiency, can be in conflict with the basic regularity needed to ensure convergence. Further research on algorithms that balance the two goals is therefore desirable. For specific applications, it is off course possible to verify that a particular ”acceleration sampler” performs adequately by experimentation: Since the Pedersen estimator always satisfy the requirement, this method can serve as a point of reference in simulation experiments. If, for a given model, an acceleration sampler produces results that are very different from the Pedersen approach, it signals a potential violation of assumption 4.

My second comment concerns the use of particle filtering rather than posterior smoothing for the latent volatility process in the SV model. I welcome the suggestion to use this algorithm, and it is apparent from the results in DG that it attains attractive computing speeds. MLE and Bayesian inference, however, require simulation from the full posterior distribution of the latent volatility. While the particle filtering algorithm very well may provide a very good approximation to the posterior filtering algorithm, it is unclear what its approximating properties are. The simulation experiments reported, however, suggest that the algorithm may be very promising, and I therefore welcome further analysis of particle filtering algorithms for estimation purposes.