Statistical Science
2005, Vol. 20, No. 4, 344–346
DOI 10.1214/088342305000000449

Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics

Michael Sørensen

1. INTRODUCTION

Professor Fan should be congratulated for his review that convincingly demonstrates the usefulness of nonparametric techniques to financial econometric problems. He is mainly concerned with financial models given by stochastic differential equations, that is, diffusion processes. I will therefore complement his selective review by discussing some important problems and useful methods for diffusion models that he has not covered. My concern will mainly, but not solely, be with parametric techniques. A recent comprehensive survey of parametric inference for discretely sampled diffusion models can be found in [19].

2. GAUSSIAN LIKELIHOOD FUNCTIONS

In his brief review of parametric methods, Professor Fan mentions the Gaussian approximate likelihood function based on the Euler scheme and states that this method has some bias when the time between observations Δ is large. This is actually a very serious problem. As an example, consider a model with a linear drift of the form $\mu(x) = -\beta(x-\alpha)$ $(\beta > 0)$. The estimator $\hat{\beta}_n$ of β obtained from the Gaussian approximate likelihood based on the Euler scheme converges to

$$(1 - e^{-\beta_0 \Delta})\Delta^{-1}$$

as the number of observations n tends to infinity. Here β_0 denotes the true parameter value. The limiting value of $\Delta\hat{\beta}_n$ is always smaller than one, and the limit of $\hat{\beta}_n$ is always smaller than Δ^{-1}. Thus the asymptotic bias can be huge if Δ is large. A simulation study in [3] demonstrates that also for finite sample sizes an enormous bias can occur. When $\Delta\beta_0$ is small so that $(1 - e^{-\beta_0\Delta})\Delta^{-1} \approx \beta_0$, the asymptotic bias is negligible. The problem is, however, that if we use the approximate likelihood function based on the Euler scheme, there is no way we can know whether $\Delta\beta_0$ is small or large because $\Delta\hat{\beta}_n$ will always tend to be small. I suspect that the nonparametric methods outlined in Sections 3.2 and 3.5 might suffer from a similar shortcoming as they are based on the same type of approximation as the Euler scheme.

A simple solution to this problem is to use an approximate likelihood function where the transition density is replaced by a normal distribution with mean equal to the exact conditional expectation $F(x,\theta) = E_\theta(X_\Delta | X_0 = x)$ and with the variance equal to the exact conditional variance $\Phi(x;\theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x)$. Here θ is the (typically multivariate) parameter to be estimated. This approach is exactly the same as using quadratic martingale estimating functions; see [3] and [20]. The estimators obtained from quadratic martingale estimating functions have the same nice properties for high frequency observations (small Δ) as the estimators based on the Euler likelihood, but they are consistent for any value of Δ and can thus be used whether or not Δ is small. In most cases there is no explicit expression for the functions $F(x,\theta)$ and $\Phi(x;\theta)$, so often they must be determined by simulation. This requires, however, only a modest amount of computation and is not a problem in practice. If a completely explicit likelihood is preferred, one can approximate $F(x,\theta)$ and $\Phi(x;\theta)$ by expansions of a higher order than those used in the Euler scheme; see [16].

The nonparametric method in Section 3.5 could probably be improved in a similar way by using in (27) and (28) the functions $F(x,\theta)$ and $\Phi(x;\theta)$ (or the higher-order expansions in [16]) instead of the first-order approximation used in the Euler scheme.

3. MARTINGALE ESTIMATING FUNCTIONS

More generally, martingale estimating functions provide a simple and versatile technique for estimation in discretely sampled parametric stochastic differential equation models that works whether or not Δ is small.

Michael Sørensen is Professor and Head, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark (e-mail: michael@math.ku.dk).

An estimator is obtained by solving the equation

$$\sum_{i=1}^{n}\sum_{j=1}^{N} a_j(X_{(i-1)\Delta};\theta) \cdot [f_j(X_{i\Delta}) - H_\Delta^\theta f_j(X_{(i-1)\Delta})] = 0,$$

where H_Δ is the transition operator, and where the function a has the same dimension as the parameter θ. First suitable functions f_j are chosen, and then the weight functions a_j are determined so that an optimal estimating function in the sense of Godambe and Heyde [9] is obtained; see also [10]. Optimal estimating functions are approximations to the non-explicit score function. Usually $H_\Delta^\theta f_j$ must be determined by a modest amount of simulation, but Kessler and Sørensen [17] demonstrated how completely explicit estimating functions can be obtained if the functions f_j are eigenfunctions of the operator L_θ (called the generator); see also [18] for details on how to explicitly find the optimal weight functions. A review of the theory of estimating functions for discretely sampled diffusion-type models can be found in [1].

For martingale estimating functions large sample results concerning estimators can be obtained via martingale limit theory. Under weak conditions, estimators are consistent, and optimal estimating functions tend to work well when the functions f_j are chosen reasonably, that is, such that a good approximation to the score function can be obtained. At low sampling frequencies the estimators are, however, usually not efficient. The behavior of the estimators at high sampling frequencies can be investigated by considering an asymptotic scenario where the time between observations Δ_n is assumed to go to zero, as the sample size n tends to infinity, sufficiently slowly that the time horizon over which observations are made, $n\Delta_n$, tends to infinity. It is well known that in this situation estimators of parameters appearing in the diffusion coefficient may converge at a suboptimal rate, $1/\sqrt{n\Delta_n}$. The reason is that there is a lot of information about the diffusion coefficient in the fine structure of diffusion trajectories, which some estimators do not capture. Recently Sørensen [22] has given conditions ensuring that a martingale estimating function provides estimators that are rate-optimal (rate $1/\sqrt{n}$) and efficient in the high-frequency asymptotic scenario. Optimal martingale estimating functions satisfy these conditions. Quadratic martingale estimating functions are always rate-optimal, and if they are obtained from Gaussian approximate likelihood functions they are efficient too. These results are closely related to the theory of small Δ-optimality developed in [13] and [14].

4. NON-MARKOVIAN OBSERVATIONS

There are several situations in which observations from a diffusion process are non-Markovian. Most prominently this happens if a function of lower dimension of a multivariate diffusion is observed. An example is the stochastic volatility model that plays an important role as a model of financial time series since it is well known that a simple one-dimensional diffusion often cannot capture all the salient features of such data. Another example is given by the sums of diffusions proposed by Bibby, Skovgaard and Sørensen [2] as models of phenomena with more than one time scale. Other situations where diffusion data are non-Markovian are in the presence of measurement error, or when only integrals of the diffusion over time-intervals are observed; see [4]. The latter is, for instance, the case when climate data from ice cores are analyzed by means of a diffusion model. When the data are non-Markovian, it is usually not possible to find a tractable martingale estimating function, but an alternative is provided by the prediction-based estimating functions proposed in [21], which can be interpreted as approximations to martingale estimating functions.

Asymptotic results for estimators based on non-Markovian data are usually based on the assumption that the underlying diffusion process is strongly mixing. The condition ensuring exponential ρ-mixing cited in Section 2.2 is not easy to check for concrete diffusion models. A condition on the drift and diffusion coefficient that is easy to verify and that implies exponential ρ-mixing and α-mixing was given by Genon-Catalot, Jeantheau and Larédo [6].

5. NONPARAMETRIC METHODS

Let me conclude by drawing attention to some relatively early work on nonparametric methods for discretely sampled diffusion models. Wavelet methods for estimating the diffusion coefficient of a time-dependent model were proposed by Genon-Catalot, Larédo and Picard [7]. The first estimator of the diffusion coefficient mentioned in Section 3.2 was first proposed by Florens-Zmirou [5]. She considered a high frequency asymptotic scenario with fixed time span, that is, with $n\Delta_n$ constant, and proved that the asymptotic distribution of her estimator is a mixture of normal distributions where the mixing distribution is the distribution of the local time of the diffusion. If a data-dependent normalization of the estimator is used, an asymptotic normal distribution is obtained. In a series of important papers, Marc Hoffmann has studied optimal rates of convergence of nonparametric estimators of the drift and diffusion coefficient under the three asymptotic scenarios usually considered for diffusion models including optimal estimators; see [8, 11, 12]. Other estimators of the diffusion coefficient were proposed by Soulier [23] and Jacod [15].

REFERENCES

[1] BIBBY, B. M., JACOBSEN, M. and SØRENSEN, M. (2005). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.). North-Holland, Amsterdam. To appear.

[2] BIBBY, B. M., SKOVGAARD, I. M. and SØRENSEN, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* **11** 191–220.

[3] BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39.

[4] DITLEVSEN, S. and SØRENSEN, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist.* **31** 417–429.

[5] FLORENS-ZMIROU, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804.

[6] GENON-CATALOT, V., JEANTHEAU, T. and LARÉDO, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079.

[7] GENON-CATALOT, V., LARÉDO, C. and PICARD, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelet methods. *Scand. J. Statist.* **19** 317–335.

[8] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253.

[9] GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

[10] HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application*. Springer, New York.

[11] HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** 135–163.

[12] HOFFMANN, M. (1999). L_p estimation of the diffusion coefficient. *Bernoulli* **5** 447–481.

[13] JACOBSEN, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Δ-optimality. *Scand. J. Statist.* **28** 123–149.

[14] JACOBSEN, M. (2002). Optimality and small Δ-optimality of martingale estimating functions. *Bernoulli* **8** 643–668.

[15] JACOD, J. (2000). Nonparametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* **27** 83–96.

[16] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229.

[17] KESSLER, M. and SØRENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314.

[18] LARSEN, K. S. and SØRENSEN, M. (2005). A diffusion model for exchange rates in a target zone. *Math. Finance*. To appear.

[19] SØRENSEN, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *Internat. Statist. Rev.* **72** 337–354.

[20] SØRENSEN, M. (1997). Estimating functions for discretely observed diffusions: A review. In *Selected Proceedings of the Symposium on Estimating Functions* (I. V. Basawa, V. P. Godambe and R. L. Taylor, eds.) 305–325. IMS, Hayward, CA.

[21] SØRENSEN, M. (2000). Prediction-based estimating functions. *Econom. J.* **3** 123–147.

[22] SØRENSEN, M. (2005). Efficient martingale estimating functions for discretely sampled ergodic diffusions. Preprint, Dept. Appl. Math. and Statistics, Univ. Copenhagen.

[23] SOULIER, P. (1998). Nonparametric estimation of the diffusion coefficient of a diffusion process. *Stochastic Anal. Appl.* **16** 185–200.