Statistical Science
2005, Vol. 20, No. 4, 347–350
DOI 10.1214/088342305000000458

Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics

Per A. Mykland and Lan Zhang

We would like to congratulate Jianqing Fan for an excellent and well-written survey of some of the literature in this area. We will here focus on some of the issues which are at the research frontiers in financial econometrics but are not covered in the survey. Most importantly, we consider the estimation of actual volatility. Related to this is the realization that financial data is actually observed with error (typically called *market microstructure*), and that one needs to consider a *hidden semimartingale model*. This has implications for the Markov models discussed above.

For reasons of space, we have not included references to all the relevant work by the authors that are cited, but we have tried to include at least one reference to each of the main contributors to the realized volatility area.

1. THE ESTIMATION OF ACTUAL VOLATILITY: THE IDEAL CASE

The paper discusses the estimation of Markovian systems, models where the drift and volatility coefficients are functions of time t or state x. There is, however, scope for considering more complicated systems. An important tool in this respect is the direct estimation of volatility based on high-frequency data. One considers a system of, say, log securities prices, which follows:

$$dX_t = \mu_t\,dt + \sigma_t\,dB_t, \tag{1}$$

where B_t is a standard Brownian motion. Typically, μ_t, the drift coefficient, and σ_t^2, the instantaneous variance (or volatility) of the returns process X_t, will be stochastic processes, but these processes can depend on the past in ways that need not be specified, and can be substantially more complex than a Markov model. This is known as an *Itô process*.

A main quantity of econometric interest is to obtain time series of the form $\Xi_i = \int_{T_i^-}^{T_i^+} \sigma_t^2\,dt$, $i = 1, 2, \ldots$. Here T_i^- and T_i^+ can, for example, be the beginning and the end of day number i. Ξ_i is variously known as the *integrated variance* (*or volatility*) or *quadratic variation* of the process X. The reason why one can hope to obtain this series is as follows. If $T_i^- = t_0 < t_1 < \cdots < t_n = T_i^+$ spans day number i, define the *realized volatility* by

$$\hat{\Xi}_i = \sum_{j=0}^{n-1} (X_{t_{j+1}} - X_{t_j})^2. \tag{2}$$

Then stochastic calculus tells us that

$$\Xi_i = \lim_{\max|t_{j+1}-t_j|\to 0} \hat{\Xi}_i. \tag{3}$$

In the presence of high frequency financial data, in many cases with transactions as often as every few seconds, one can, therefore, hope to almost *observe* Ξ_i. One can then either fit a model to the series of $\hat{\Xi}_i$, or one can use it directly for portfolio management (as in [12]), options hedging (as in [29]), or to test goodness of fit [31].

There are too many references to the relationship (3) to name them all, but some excellent treatments can be found in [27], Section 1.5; [26], Theorem I.4.47 on page 52; and [33], Theorem II-22 on page 66. An early econometric discussion of this relationship can be found in [2].

To make it even more intriguing, recent work both from the probabilistic and econometric sides gives the mixed normal distribution of the error in the approximation in (3). References include [6, 25, 31]. The random variance of the normal error is $2\frac{T_i^+ - T_i^-}{n}$ ·

Per A. Mykland is Professor, Department of Statistics, The University of Chicago, Chicago, Illinois 60637, USA (e-mail: mykland@galton.uchicago.edu). Lan Zhang is Assistant Professor, Department of Finance, University of Illinois at Chicago, Chicago, Illinois 60607, and Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: lzhang@stat.cmu.edu).