# 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 reserach 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:

$$(1) \qquad dX_t = \mu_t \, dt + \sigma_t \, dB_t,$$

where $B_t$ is a standard Brownian motion. Typically, $\mu_t$, the drift coefficient, and $\sigma_t^2$, the instantaneous variance

*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).*

(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 $i$. $\Xi_i$ 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

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

Then stochastic calculus tells us that

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

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* $\Xi_i$. 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 $2 \frac{T_i^+ - T_i^-}{n}$ .

$\int_{T_i^-}^{T_i^+} \sigma_t^4 \, dH(t)$, where $H$ is the *quadratic variation of time*. $H(t) = t$ in the case where the $t_i$ are equidistant.

Further econometric literature includes, in particular, [3, 4, 8, 9, 14, 18, 32]. Problems that are attached to the estimation of covariations between two processes are discussed in [22]. Estimating $\sigma_t^2$ at each point $t$ goes back to [13]; see also [30], but this has not caught on quite as much in econometric applications.

## 2. THE PRESENCE OF MEASUREMENT ERROR

The theory described above runs into a problem with real data. For illustration, consider how the realized volatility depends on sampling frequency for the stock (and day) considered in Figure 1. The estimator does not converge as the observation points $t_i$ become dense in the interval of this one day, but rather seems to take off to infinity. This phenomenon was originally documented in [2]. For transaction data, this picture is repeated for most liquid securities [19, 37].

In other words, the model (1) is wrong. What can one do about this? A lot of people immediately think that

the problem is due to jumps, but that is not the case. The limit in (3) exists even when there are jumps. The requirement for (3) to exist is that the process $X$ be a semimartingale (we again cite Theorem I.4.47 of [26]), which includes both Itô processes and jumps.

The inconsistency between the empirical results where the realized volatility diverges with finer sampling, and the semimartingale theory which dictates the convergence of the realized volatility, poses a problem, since financial processes are usually assumed to be semimartingales. Otherwise, somewhat loosely speaking, there would be arbitrage opportunities in the financial markets. For rigorous statements, see, in particular, [11]. The semimartingaleness of financial processes, therefore, is almost a matter of theology in most of finance, and yet, because of Figure 1 and similar graphs for other stocks, we have to abandon it.

Our alternative model is that there is measurement error in the observation. At transaction number $i$, instead of seeing $X_{t_i}$ from model (1) or, more generally, from a semimartingale, one observes

$$(4) \qquad\qquad Y_{t_i} = X_{t_i} + \varepsilon_i.$$

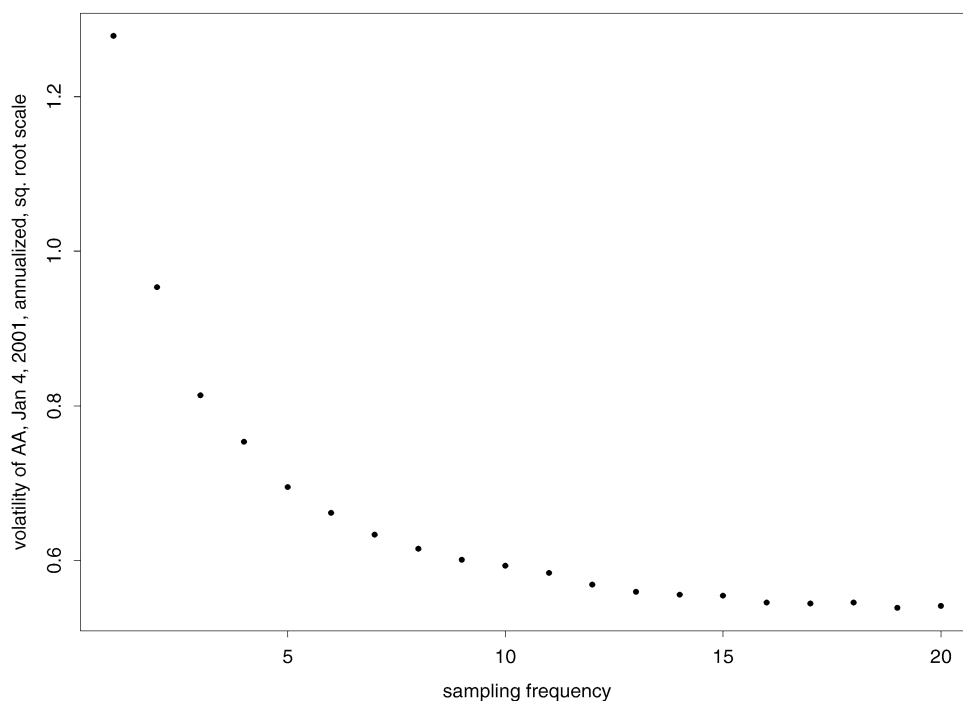### dependence of estimated volatility on sampling frequency



FIG. 1. *Plot of realized volatility for Alcoa Aluminum for January 4, 2001. The data is from the TAQ database. There are* 2011 *transactions on that day, on average one every* 13.365 *seconds. The most frequently sampled volatility uses all the data, and this is denoted as "frequency = 1." "Frequency = 2" corresponds to taking every second sampling point. Because this gives rise to two estimators of volatility, we have averaged the two. And so on for "frequency = k" up to* 20. *The plot corresponds to the average realized volatility discussed in* [37]. *Volatilities are given on an annualized and square root scale.*

We call this the *hidden semimartingale model*. The rationale is (depending on your subject matter) either that a transaction is a measurement of the underlying price $X_{t_i}$, and of course there is error, or that it is due to *market microstructure*, as documented by, among others, Roll [34], Glosten [15], Glosten and Harris [16], Brown [7], Harris [20] and Hasbrouck [21]. See [1] for a discussion of this.

A natural model for the error is that it is either i.i.d. or a stationary process, as considered by Zhou [38], Gloter and Jacod [17], Zhang, Mykland and Aït-Sahalia [37], Bandi and Russell [5], Zhang [36], Aït-Sahalia, Mykland and Zhang [1] and Hansen and Lunde [19].

Under quite loose conditions, this alternative model is consistent with the plot in Figure 1. Instead of (3), one gets that the realized volatility becomes $nE(\varepsilon_1 - \varepsilon_0)^2 + O_p(n^{-1/2})$. In the early literature (as cited in the previous section), the problem is usually taken care of by (*sic*) reducing $n$. A variety of approaches that improve on this are documented in [37], to which we refer for an in depth discussion. As demonstrated by Zhang [36], the true volatility $\Xi_i$ can be consistently estimated at rate $O_p(n^{-1/4})$, as opposed to $O_p(n^{-1/2})$ when there is no error. This is not as slow as it seems, since $n$ is quite large for liquid securities.

An alternative description of the error is that it arises due to rounding (financial instruments are, after all, traded on a price grid). Research in this direction has been done by Delattre and Jacod [10] and by Zeng [35]. To first order, the rounding and additive error models are similar, as documented by Delattre and Jacod [10]; see also [28].

It is awkward that these models imply the existence of arbitrage. The size of the error, however, is so small that it is hard to take economic advantage of them, and this, presumably, is why such deviations can persist.

## 3. IMPLICATIONS FOR MARKOV MODELS

We now return to the subject to Jianqing Fan's overview, namely the Markov case. It is clear that the model without observation error is not consistent with the data. This may not be a problem when working with, say, daily data, but would pose problems when using high-frequency (intraday) observations. It is presumably quite straightforward to extend the methods discussed in the paper to the case of observation error, and it would be interesting to see the results. The same applies to similar studies on Markov models by the "French school," such as Hoffmann [23] and Jacod [24].

## REFERENCES

[1] AÏT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

[2] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2000). Great realizations. *Risk* **13** 105–108.

[3] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55.

[4] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625.

[5] BANDI, F. M. and RUSSELL, J. R. (2005). Microstructure noise, realized volatility and optimal sampling. Technical report, Graduate School of Business, Univ. Chicago.

[6] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 253–280.

[7] BROWN, S. J. (1990). Estimating volatility. In *Financial Options: From Theory to Practice* (S. Figlewski, W. Silber and M. Subrahmanyam, eds.) 516–537. Business One-Irwin, Homewood, IL.

[8] CHERNOV, M. and GHYSELS, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *J. Financial Economics* **56** 407–458.

[9] DACOROGNA, M. M., GENÇAY, R., MÜLLER, U., OLSEN, R. B. and PICTET, O. V. (2001). *An Introduction to High-Frequency Finance*. Academic Press, San Diego.

[10] DELATTRE, S. and JACOD, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3** 1–28.

[11] DELBAEN, F. and SCHACHERMAYER, W. (1995). The existence of absolutely continuous local martingale measures. *Ann. Appl. Probab.* **5** 926–945.

[12] FLEMING, J., KIRBY, C. and OSTDIEK, B. (2001). The economic value of volatility timing. *J. Finance* **56** 329–352.

[13] FOSTER, D. and NELSON, D. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64** 139–174.

[14] GALLANT, A. R., HSU, C.-T. and TAUCHEN, G. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics* **81** 617–631.

[15] GLOSTEN, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *J. Finance* **42** 1293–1307.

[16] GLOSTEN, L. R. and HARRIS, L. E. (1988). Estimating the components of the bid/ask spread. *J. Financial Economics* **21** 123–142.

[17] GLOTER, A. and JACOD, J. (2000). Diffusions with measurement errors: I—Local asymptotic normality and II—Optimal estimators. Technical report, Univ. de Paris VI.

[18] GONCALVES, S. and MEDDAHI, N. (2005). Bootstrapping realized volatility. Technical report, Univ. de Montréal.

[19] HANSEN, P. R. and LUNDE, A. (2006). Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* To appear.

[20] HARRIS, L. (1990). Statistical properties of the Roll serial covariance bid/ask spread estimator. *J. Finance* **45** 579–590.

[21] HASBROUCK, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies* **6** 191–212.

[22] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379.

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

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

[25] JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307.

[26] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. Springer, New York.

[27] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.

[28] KOLASSA, J. and MCCULLAGH, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18** 981–985.

[29] MYKLAND, P. A. (2003). Financial options and statistical prediction initervals. *Ann. Statist.* **31** 1413–1438.

[30] MYKLAND, P. A. and ZHANG, L. (2001). Inference for volatility type objects and implications for hedging. Technical report, Dept. Statistics, Carnegie Mellon Univ.

[31] MYKLAND, P. A. and ZHANG, L. (2002). ANOVA for diffusions. Technical report, Dept. Statistics, Univ. Chicago.

[32] OOMEN, R. (2004). Properties of realized variance for a pure jump process: Calendar time sampling versus business time sampling. Technical report, Warwick Business School, Univ. Warwick.

[33] PROTTER, P. (2004). *Stochastic Integration and Differential Equations*: *A New Approach*, 2nd ed. Springer, New York.

[34] ROLL, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finance* **39** 1127–1139.

[35] ZENG, Y. (2003). A partially-observed model for micromovement of asset process with Bayes estimation via filtering. *Math. Finance* **13** 411–444.

[36] ZHANG, L. (2004). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Technical report, Dept. Statistics, Carnegie Mellon Univ.

[37] ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411.

[38] ZHOU, B. (1996). High-frequency data and volatility in foreign-exchange rates. *J. Bus. Econom. Statist.* **14** 45–52.