Econometrica, Vol. 68, No. 1 (January, 2000), 1–22

THE ECONOMETRICS OF ULTRA-HIGH-FREQUENCY DATA[1]

BY ROBERT F. ENGLE[2]

Ultra-high-frequency data is defined to be a full record of transactions and their associated characteristics. The transaction arrival times and accompanying measures can be analyzed as marked point processes. The ACD point process developed by Engle and Russell (1998) is applied to IBM transactions arrival times to develop semiparametric hazard estimates and conditional intensities. Combining these intensities with a GARCH model of prices produces ultra-high-frequency measures of volatility. Both returns and variances are found to be negatively influenced by long durations as suggested by asymmetric information models of market micro-structure.

KEYWORDS: Transactions data, point processes, hazard functions, survival models, ACD, volatility, ARCH, GARCH, market micro-structure.

1. INTRODUCTION

ONE MEASURE OF PROGRESS in empirical econometrics is the frequency of data used. Upon entering Graduate School, I learned that T. C. Liu had just estimated the first quarterly model of the U.S. economy in Liu (1963). Shortly thereafter, he surpassed this threshold by publishing a monthly macroeconomic model, Liu (1969). I've not yet seen a full weekly model of the macroeconomy but I suspect there may be some in both government and private sector research groups. In finance, a similar succession has lead from the analysis of annual data to monthly data, to weekly data, to daily data, and now there is great interest in intradaily models.

In each case, much of the movement to higher frequency econometrics was a consequence of the availability of higher frequency measurements of the econ-

[1] This paper was presented as the Fisher Schultz Lecture at the Econometric Society Meeting in Istanbul, Turkey, August, 1996.

[2] The author is very grateful to Jeff Russell whose collaboration on two previous papers forms the basis for this paper. He carried out preliminary computer analysis and continually helped to extend the models presented here. Simone Manganelli suggested the economic variables in the last section. Financial support is gratefully acknowledged from NSF Grant SBR-94-22575 and SBR-97-30062. Many offered insights into this work including David Brillinger, Sir David Cox, David Easley, Clive Granger, Jim Hamilton, Maureen O'Hara, Ken Wallis, Mark Watson, Hal White, and participants at UCSD, Princeton, UBC, and Washington University Econometric Workshops, London Business School Conference on Macro and Finance, Tilburg Conference on Financial Econometrics '96, the Malinvaud Seminar at INSEE '97, and the Western Finance Association Meetings '97. This paper was presented as the Frank Paisch Lecture, Royal Economic Society, Swansea, Wales '96, and the A. W. Phillips Lecture at the Australasian Econometric Society Meeting, Melbourne, Australia, '97. The data and programs used in this paper are posted on the author's website *http://weber.ucsd.edu/~mbacci/engle*. These run on EViews 3.1.

This version of the paper reestimates the models with a larger data set and commonly available software. The results are quite similar although some paradoxes have disappeared.

omy. It is natural to suppose that this will continue and that we will have ever increasing frequencies of observations. However, a moment's reflection will reveal that this is not the case. The limit in nearly all cases, is achieved when *all transactions are recorded*. These transactions may occur in the supermarket, on the internet or in financial markets. It is difficult to think of economic variables that really are measurable at arbitrarily high frequencies.

Let us call this limiting frequency "*ultra-high frequency*" and spend the time in this paper discussing econometric methods for the analysis of ultra-high-frequency data. The salient feature of such ultra-high-frequency data is that they are fundamentally irregularly spaced. Of course, one can aggregate this data up to fixed intervals of time, but one might then argue that it is no longer ultra-high-frequency data. There is naturally a loss of information in such aggregates. This loss occurs partly because the large number of zeros makes econometric analysis very complex if the intervals are small.

The thrust of this paper will be to develop methods that are directly tailored to the irregular spacing of the data, rather than to adapt fixed interval econometrics to this new situation. The statistics literature is replete with models for data of this form. These models treat events as arriving in time according to some probability law. Famous stochastic processes such as the Poisson process and its doubly stochastic, cluster and self-exciting forms, birth and death processes, and many other continuous time discrete state processes, have been developed to solve problems in science and engineering. Many of these processes can be used or extended to address economic problems. The basic model to be presented and extended in this paper is the autoregressive conditional duration model developed by Engle and Russell (1997, 1998), which is a type of dependent Poisson process.

Section 2 of the paper will pose the economic questions the statistical model is to answer. Sections 3 and 4 develop the econometric models to be applied. Section 5 gives results from IBM stock transactions, while Section 6 presents the IBM price model. Section 7 concludes.

2. FORMULATING THE ECONOMIC QUESTIONS STATISTICALLY

Transactions data can be described by two types of random variables. The first is the time of the transaction, and the second is a vector observed at the time of the transaction. In the literature of point processes, these latter variables are called marks as they identify or further describe the event that occurred. In the type of financial data to be examined here, the point of time is the time at which a contract to trade some number of shares of IBM stock is agreed upon. The marks are the volume of the contract, the price of the contract, and the posted bid and ask prices at the time. Additional marks that could be used or observed would be the counter-parties to the trade, the posted bid and ask prices for other stocks, the order mechanism, and many other features of a trade that are of interest in studying market microstructure.

Let t_i be the time at which the ith trade occurs and let $x_i = t_i - t_{i-1}$ be the duration between trades. At the ith event the marks are observed and let these be denoted y_i, which is a $k \times 1$ vector from a sample space Ξ. The data therefore can be viewed as

$$(1) \qquad \{(x_i, y_i),\ i = 1, \ldots, N\}$$

where the ith observation has joint density conditional on the past filtration of (x, y) given by

$$(2) \qquad (x_i, y_i)|\mathsf{F}_{i-1} \sim f\left(x_i, y_i|\breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right)$$

where $\breve{z}_i = \{z_i, z_{i-1}, \ldots, z_1\}$ denotes the past of z and θ's are parameters that are potentially different from observation to observation.

Economic hypotheses or measures of interest can now be expressed in terms of this density function. The analysis of quantity data in traditional fixed interval econometrics typically involves estimating the expected number of transactions or volume of transactions in a particular time interval as a function of explanatory variables. For transactions data the realized transactions are zero at almost every point in time; hence the natural measure of quantity is the probability of an event at each point of time. Let $N(t)$ be the number of transactions occurring by time t and define the conditional intensity λ as

$$(3) \qquad \lambda_i(t, \breve{x}_{i-1}, \breve{y}_{i-1}) = \lim_{\Delta t \to 0} \frac{\Pr\left(N(t + \Delta t) > N(t)|\breve{x}_{i-1}, \breve{y}_{i-1}\right)}{\Delta t}, \quad \text{for} \quad t_{i-1} \le t < t_i.$$

The conditional intensity summarizes future expected arrival rates and how these depend upon observables.

The conditional intensity is easily derived from (2). For any $t > t_{i-1}$, the probability of an event must be conditioned not only on all past events but also on the fact that there has not been an event since t_{i-1}. This is simply expressed as the density of $t - t_i$ (or x) divided by the survival probability, which is the probability that the next event will be at a time greater than t. Since y is irrelevant to this calculation it must be integrated out, giving

$$(4) \qquad \lambda_i(t, \breve{x}_{i-1}, \breve{y}_{i-1}) = \frac{\int_{u \in \Xi} f\left(t - t_{i-1}, u|\breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right) du}{\int\int_{s \ge t,\, u \in \Xi} f\left(s - t_{i-1}, u|\breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right) du\, ds}, \quad \text{for} \quad t_{i-1} \le t \le t_i.$$

A simpler expression can easily be obtained. Without loss of generality, the joint density can be written as the product of the marginal density of the duration times the conditional density of the marks given the duration, all conditioned upon the past transactions:

$$(5) \qquad f\left(x_i, y_i|\breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right) = g\left(x_i|\breve{x}_{i-1}, \breve{y}_{i-1}; \theta_{1i}\right) q\left(y_i|x_i, \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_{2i}\right).$$

Substituting (5) into (4) gives the standard formula for the conditional intensity, which now allows for past influences of both durations and marks:

$$\lambda_i(t, \breve{x}_{i-1}, \breve{y}_{i-1}) = \frac{g\left(t - t_{i-1} | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_{1i}\right)}{\int_{s \geq t} g\left(s - t_{i-1} | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_{1i}\right) ds}, \quad \text{for} \quad t_{i-1} \leq t \leq t_i. \tag{6}$$

In the analysis of survival data, this is called the hazard since it refers to the exit from the state defined by $N(t) = i - 1$. See, for example, Halbfleisch and Prentice (1980) and Lancaster (1990).

Economic hypotheses are often associated with the distribution of the marks. Two such questions are the distribution of the next mark, regardless of when it occurs, or the distribution of the next mark if it occurs at time t. To calculate the distribution of the next mark conditional on t, one simply needs the function q defined in equation (5). To calculate the distribution of the next mark, regardless of when it occurs, requires calculating the marginal density of the mark:

$$r\left(y_i | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right) = \int_{s \geq 0} f\left(s, y_i | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_i\right) ds. \tag{7}$$

Corresponding to each of these versions are prediction questions. What is the intensity expected to be at some specified time in the future or after a certain number of trades? What is the distribution of the marks at some fixed time in the future or after a certain number of transactions have occurred? Each of these questions can be answered by manipulation of the densities in (5), although in most cases, closed form solutions cannot be obtained. Instead, simulations can be used to generate answers. These simulations are precisely defined by the joint density functions in (5) conditional on past observations.

In the examples to be discussed below, the prices and times are modeled jointly. This allows measurement not only of the transaction rate but its interaction with volatility. The model can be thought of as a microscopic view of the relation between volume and volatility, a widely studied phenomenon.

A popular approach to this analysis is through models of time deformation where the relevant time scale is "economic time" rather than "calendar time." Intuitively, economic time measures the arrival rate of new information that influences both volume and volatility. The joint analysis of transaction times and prices implements the standard time deformation models by obtaining a direct measure of the arrival rate of transactions and then measuring exactly how this influences the distribution of the other observables in the market.

3. ECONOMETRIC ISSUES

The econometric issues in applying these techniques are specifying and testing the parameterizations of the functions g and q in equation (5) since the relevant economic questions can all be determined from these functions.

The approach developed here is maximum likelihood with all the associated parametric inference and testing procedures. In the subsequent section, a semiparametric approach to hazard estimation will be presented.

In order to estimate parameters by maximum likelihood, it is necessary to formulate the process so that there are a finite number of parameters θ and that these are invariant over events. The log likelihood function is simply the sum of the logs of all the N individual joint densities conditional on the past and can therefore be written as

$$(8) \qquad \mathscr{L}(X,Y;\theta) = \sum_{i=1}^{N} \log f\left(x_i, y_i | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta\right)$$

where X and Y are all the data and $\theta \in \Theta$ is the set of parameters. The log likelihood can be expressed in terms of the parameterization of (5) as

$$(9) \qquad L(X,Y;\theta) = \sum_{i=1}^{N} \left[\log g\left(x_i | \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_1\right) + \log q\left(y_i | x_i, \breve{x}_{i-1}, \breve{y}_{i-1}; \theta_2\right)\right],$$

which can be maximized with respect to the unknown parameters (θ_1, θ_2). If the parameters (θ_1, θ_2) are variation free as in Engle, Hendry, and Richard (1983), and if the parameters of interest are contained in θ_1, then the MLE can be obtained simply from maximizing the first term. Alternatively, if the parameters of interest are in θ_2 and they are variation free, then x is weakly exogenous and the MLE will be obtained from a maximization only of the second term. However, if the parameters are not variation free, then the MLE can be obtained only by maximizing the sum of the two terms. In such a case, the separate maxima will be inefficient relative to the MLE. This is precisely the case for many of the models estimated in this paper. Parameters in the two parts of the model are related. Consistent but inefficient estimates are obtained by taking estimates of θ_1 from the first term, as given in the second. Because samples are very large, this loss of efficiency should be unimportant.

In principle, such a likelihood can be derived from a model where there are unobserved latent processes. For example, some parameters may actually be unobserved stochastic processes. Initially suppose that there is a series $\{\phi\}$ that is unobserved by the econometrician but that is assumed to follow a probability law with conditional density given by p with unknown parameters that are included in θ. The conditional density of the observables can be expressed in terms of the different information sets as follows:

$$\begin{aligned} &\phi_i | \mathsf{F}_{i-1}^{(x,y)} \cup \mathsf{F}_{i-1}^{\phi} \sim p\left(\phi_i | \breve{x}_{i-1}, \breve{y}_{i-1}, \breve{\phi}_{i-1}; \theta\right), \\ (10) \qquad &(x_i, y_i, \phi_i) | \mathsf{F}_{i-1}^{(x,y)} \cup \mathsf{F}_{i-1}^{\phi} \sim f^*\left(x_i, y_i | \breve{x}_{i-1}, \breve{y}_{i-1}, \breve{\phi}_i; \theta\right) \\ &\qquad \times p\left(\phi_i | \breve{x}_{i-1}, \breve{y}_{i-1}, \breve{\phi}_{i-1}; \theta\right), \end{aligned}$$

where f^* is the density of the observables (x, y) given current and past values of ϕ. To obtain the density of (x, y) conditional only on observables, the density in the second line of (9) must be integrated with respect to $\{\phi_i,\ i=1,\dots,N\}$ leaving a likelihood as in (8) with only the fixed parameters θ. In practice, the multidimensional integral is very difficult to evaluate and requires sophisticated Monte Carlo methods. See, for example, Jacquier, Polson, and Rossi (1994) for a volatility process and Shephard (1993) for more general problems. Furthermore, it is not clear that it is easier to specify the densities p and f^* in (9) than the processes in (2) from a priori considerations.

The specification of the conditional density of the durations given covariates is a familiar problem in statistics and biostatistics. Much of this literature is focussed on the treatment of censored durations, which are endemic in survival analysis. In the transactions analyses here, there are no censored durations, so the range of specifications that can be considered is more generous. However, since the focus is on the temporal dependence of the durations, the covariates are typically going to be lagged dependent variables and functions of lagged dependent variables.

Engle and Russell (1997, 1998), propose a specification of the conditional density that requires only a mean function. They define ψ as the conditional duration given by

$$(11) \qquad \psi_i \equiv \psi(\breve{x}_{i-1},\breve{y}_{i-1};\theta) = E_{i-1}\big(x_i|\breve{x}_{i-1},\breve{y}_{i-1};\theta\big) = \int s g\big(s|\breve{x}_{i-1},\breve{y}_{i-1};\theta\big)\,ds$$

and then assume that

$$(12) \qquad x_i = \psi_i \varepsilon_i, \quad \text{where} \quad \varepsilon_i \sim i.i.d.$$

Because durations and expected durations are positive, the multiplicative disturbance naturally will have positive probability only for positive values and it must have a mean of unity. This assumption requires that all the temporal dependence in the durations be captured by the mean function. Such an assumption is testable in the sense that the standardized durations can be checked for various forms of deviation from independence or identical distribution. Notice of course that the squares of x_i will have autocorrelation; only after dividing the durations by their expected values should the conditional moments be constant. The assumption that the disturbances are i.i.d. can be relaxed as is shown below. Under (12),

$$(13) \qquad g\big(x_i|\breve{x}_{i-1},\breve{y}_{i-1};\theta\big) = g(x_i|\psi_i;\theta),$$

where possibly g is completely known as a function of ψ. Because the $\{\varepsilon_i\}$ are independent, the likelihood is easily evaluated and the intensity can be computed as a function of the density of ε.

If the density and survivor function of the errors are given by

$$(14) \qquad \varepsilon \sim p_0(\cdot;\theta_{11}) \quad \text{and} \quad S_0(t;\theta_{11}) = \int_{s>t} p_0(s;\theta_{11})\,ds$$

with unknown parameters θ_{11}, then the likelihood is given by

$$(15)\qquad L(X,Y,\theta)=\sum_i \log\left(p_0\left(\frac{x_i}{\psi_i};\theta_{11}\right)\right)+\sum_i \log\left(q\left(y_i|\breve{x}_i,\breve{y}_{i-1};\theta_2\right)\right).$$

The baseline intensity or baseline hazard is given by

$$(16)\qquad \lambda_0(t;\theta_{11})=\frac{p_0(t;\theta_{11})}{S_0(t;\theta_{11})}$$

and the intensity for an arrival at time t is

$$(17)\qquad \lambda_i(t,\breve{x}_{i-1},\breve{y}_{i-1})=\lambda_0\left(\frac{t-t_{i-1}}{\psi_i}\right)\frac{1}{\psi_i},\quad \text{for}\quad t_{i-1}\le t<t_i.$$

Because ψ enters into the baseline hazard, it is called an "accelerated failure time" model. The name "accelerated failure time" has intuitive appeal in medical applications where the patient is viewed as progressing through the disease at a faster or slower pace depending upon the covariates. A similar interpretation has often been used in finance where covariates might determine whether "economic time" moves faster or slower than calendar time.

The assumption in (12) is powerful and is incorporated in some but not all familiar specifications for hazard models. For example, the familiar log linear regression models can be specified in terms of covariates z as:

$$(18)\qquad \log x_i=z_i\beta+w_i,$$

where the error density is independent and identically distributed over observations and does not depend upon β. Thus

$$(19)\qquad x_i=v_i\exp z_i\beta$$

where v_i is the exponential of w and is therefore positive. The expected value of x conditional on the covariates will be proportional to $\exp(z\beta)$ so that x/ψ will be independent and identically distributed as assumed in (12).

An alternative popular specification is the proportional hazard model (Cox (1972)) in which the baseline hazard is simply multiplied by the function of the covariates giving rise to the specification

$$(20)\qquad \lambda(t,z)=\lambda_0(t)e^{-z\beta}.$$

This model does not in general satisfy assumption (12) although it does when the baseline hazard is constant as in the exponential or when it is proportional to t^{ρ} as in the Weibull as shown by Kalbfleisch and Prentice (1980, p. 34).

4. SEMIPARAMETRIC HAZARD ESTIMATION

With simply the assumption (12) and correct specification of (11), it is desirable to find an estimate of the hazard function for x. This is called a semiparametric estimate of the hazard because it does not require parameterizing the density of x but does require specifying the mean of x. It is proposed to maximize the quasi-likelihood function

$$(21) \qquad Q(X,\theta) = -\sum_{i=1}^{N}[\log \psi_i + x_i/\psi_i] = -\sum_{i=1}^{N}\ell_i,$$

which would be the true log likelihood function if g were the exponential density. Define the score, hessian and expected hessian of ℓ as s, $\tilde{a}$, and a respectively:

$$s_i = \frac{\partial \ell_i}{\partial \theta} = -\frac{(\psi_i - x_i)}{\psi_i^2}\frac{\partial \psi_i}{\partial \theta},$$

$$(22) \qquad \tilde{a}_i = \frac{\partial^2 \ell_i}{\partial \theta \, \partial \theta'},$$

$$a_i = E_{i-1}(\tilde{a}_i) = -\frac{1}{\psi_i^2}\frac{\partial \psi_i}{\partial \theta}\frac{\partial \psi_i}{\partial \theta'}.$$

A QMLE estimator can then be established that is consistent for the parameters and has a well defined asymptotic covariance matrix. This result was first proven by Gourieroux, Monfort, and Trognon (1984) for the exponential family with independent observations. The theorem here focuses on the dependence of the data and follows the QMLE results for ARCH models given in Bollerslev and Wooldridge (1992).

THEOREM 1: *Under the following conditions*:

(i) *for some* $\theta_0 \in int\,\Theta$ *a compact parameter space*,

$$E(x_i|\mathsf{F}_{i-1};\theta_0) = \psi_i(\theta_0);$$

(ii) (a) $\psi_i(\cdot,\theta)$ *is measurable and positive with probability* 1 *for all* $\theta \in \Theta$; (b) $\psi_i(\cdot,\theta)$ *is twice continuously differentiable in int* Θ;

(iii) (a) $\{\ell_i(\theta) - \ell_i(\theta_0): i = 1,2,\ldots\}$ *satisfies the UWLLN, and* (b) θ_0 *is the identifiably unique maximizer of*

$$\frac{1}{N}\sum_{i=1}^{N}E[\ell_i(\theta) - \ell_i(\theta_0)];$$

then if $\hat{\theta}_N$ *maximizes* (21) *over* Θ,

$$\hat{\theta}_N \xrightarrow{p} \theta_0.$$

If in addition:

(iv) (a) $\{\tilde{a}_i(\theta_0)\}$ *and* $\{a_i(\theta_0)\}$ *satisfy the WLLN*; (b) $\{\tilde{a}_i(\theta) - \tilde{a}_i(\theta_0)\}$ *satisfies the UWLLN*; (c) $A_N^0 \equiv N^{-1}\Sigma_{i=1}^N E[a_i(\theta_0)]$ *is uniformly positive definite*;

(v) (a) $\{s_i(\theta_0)'s_i(\theta_0)\}$ *satisfies the WLLN*; (b) $B_N^0 \equiv N^{-1}\Sigma\, E[s_i(\theta_0)'s_i(\theta_0)]$ *is uniformly positive definite*; (c) $B_N^{0^{-1/2}}N^{-1/2}\Sigma_{i=1}^N s_i(\theta_0)' \to^d N(0, I_P)$;

(vi) (a) $\{a_i(\theta) - a_i(\theta_0)\}$ *satisfies the UWLLN*; (b) $\{s_i(\theta)'s_i(\theta) - s_i(\theta_0)'s_i(\theta_0)\}$ *satisfies the UWLLN*;
then

$$(23) \qquad \left[A_N^{o-1} B_N^o A_N^{o-1}\right]^{-1/2} \sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \xrightarrow{d} N(0, I).$$

Furthermore,

$$(24) \qquad \hat{A}_N - A_N^o \xrightarrow{p} 0, \quad \text{and} \quad \hat{B}_N - B_N^o \xrightarrow{p} 0$$

where

$$(25) \qquad \hat{A}_N = N^{-1} \sum_{i=1}^{N} \left[a_i\left(\hat{\theta}_N\right)\right], \quad \hat{B}_N = N^{-1} \sum_{i=1}^{N} \left[s_i\left(\hat{\theta}_N\right) s_i\left(\hat{\theta}_N\right)'\right].$$

The proof of this theorem follows the Bollerslev Wooldridge proof directly and will not be presented here. This theorem supports estimation and inference by QMLE without specifying the density of the disturbances. It does not require that the disturbances be i.i.d. as in (12). Verification of these conditions for particular models may not be trivial, but the existence of unit roots in the autoregressive process of ψ would generally make it especially difficult to establish these properties. For results that allow unit roots in the ACD(1, 1) model, see Engle and Russell (1998).

The result of such an estimate is a set of conditional durations for each observation in the sample. By strengthening the assumptions to include (12), the ratio of the realized duration to its expectation is i.i.d. at the true parameter values. The residuals evaluated at the estimated parameters can be used to estimate the baseline hazard function $\hat{\lambda}_0(t)$. A particular nearest neighbor version is proposed below. Since the parameters are $N^{1/2}$ consistent, and the nonparametric hazard estimator will generally have a slower rate, the hazard estimator will be a consistent estimate of the true baseline hazard.

Once the baseline hazard is estimated, the hazard for any particular arrival can be computed from (17).

$$(26) \qquad \hat{\lambda}_i(t) = \hat{\lambda}_0\left(t - t_{i-1}/\hat{\psi}_i\right)/\hat{\psi}_i \quad \text{for} \quad t_{i-1} \le t < t_i.$$

There are many ways to empirically estimate the hazard for the standardized durations. As there is no censoring or truncation in these data sets, various approaches are available. One can estimate the density nonparametrically, compute the survivor function from it, and then take a ratio, or one can calculate the sample hazard function and smooth it. Here we employ a version of the latter estimator that is essentially a k-nearest neighbor estimator. When k is chosen as one on one side and zero on the other and there are no ties, this is a Kaplan Meier estimate. Since the durations are continuous and have few mass points, it is preferable to smooth this estimator by choosing a wider bandwidth.

The hazard for a collection of individuals, is the failure rate per unit time, and the failure rate is measured as the number of failures divided by the number of individuals that could have failed, called the number at risk. Consider the failure rate of the smallest $2k$ standardized durations, which is simply $2k$ divided by the number at risk. The time interval for this set of failures is $(0, t_{2k})$ and the estimate of the hazard is simply $2k$ divided by the time interval times the number at risk. In general, let n_i be the number of individuals surviving at time t_i; then the $2k$ nearest neighbor estimate of the hazard rate is computed from

$$\hat{\lambda}(t_i) = \frac{2k}{n_i(t_{i+k} - t_{i-k})}. \tag{27}$$

The nearest neighbor estimator gives a variable bandwidth as the number of individuals at risk varies over the time axis. If the true density is exponential, then the hazard is constant and the hazard estimates should be unbiased.

5. ESTIMATING THE HAZARD FOR IBM TRADES

Data on all trades for a random collection of stocks traded on the NYSE is available from the NYSE as the TORQ database, which stands for Trades, Orders and Quotes. The sample period runs from November, 1990 through January, 1991. For IBM there are approximately 60,000 trades during this period. All trades before 9:30 AM or after 4:00 PM are discarded as well as all transactions without a reported set of quotes. Transactions on Thanksgiving Friday and the day before Christmas and New Years are deleted as well as the overnight durations. This leaves 52146 unique transaction times. By considering only the unique times, all zero durations are removed. This is consistent with interpreting a trade as a transfer of ownership from one or more sellers to one or more buyers at a point in time. The volume of shares transferred can be considered a mark.

Following Engle and Russell, the data are first "diurnally adjusted" to take out the typical time of day effect. This is accomplished by regressing the durations on the time of day using a piecewise linear spline specification and then taking ratios to get "diurnally adjusted" durations that are expressed as fractions above or below normal. The spline has knots at 9:30, 10:00, 11:00, 12:00, 1:00, 2:00, 3:00, 3:30, 4:00. While this could be done in one step as in Engle and Russell (1998), there is little to be gained in such a large data set. The range of the adjusted data is from .027 to 23.8 or $1/40$ of normal to 20 times normal. The standard deviation is 1.32.

The adjusted duration data show striking evidence of autocorrelation with a Ljung-Box statistic with 15 degrees of freedom of 8040, which dramatically exceeds the 5% point of 25. The first fifteen are all between .15 and .075 and the first 200 are all positive. In fact the first one less than .05 is at lag 129. The data show a small but very persistent signal. The models estimated are the ACD(1, 1), and the component model (see Engle and Lee (1999)), which is an ACD(2, 2) parameterized in terms of its roots so that they are constrained to be real. The component model can be expressed directly as in (29) or in terms of a "permanent" component q as in (30). These can be expressed as:

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}, \tag{28}$$

$$\begin{aligned} \psi_i = {} & \omega(1-\rho) + (\alpha+\phi)x_{i-1} + (-\phi\alpha - \phi\beta - \alpha\rho)x_{i-2} \\ & + (\beta + \rho - \phi)\psi_{i-1} + (\phi\alpha + \phi\beta - \beta\rho)\psi_{i-2}, \end{aligned} \tag{29}$$

$$\begin{aligned} \psi_i &= q_i + \alpha(x_{i-1} - q_{i-1}) + \beta(\psi_{i-1} - q_{i-1}), \qquad \text{where} \\ q_i &= \rho q_{i-1} + \omega(1-\rho) + \phi(x_{i-1} - \psi_{i-1}). \end{aligned} \tag{30}$$

In each case, the density is assumed to be exponential. The estimates are therefore interpreted as QMLE estimates as they are consistent regardless of the true density. These are estimated using GARCH software with the dependent variable as $\sqrt{x}$ and no intercept.[3] Because no marks are introduced into the model, the durations are assumed to be "*not Granger Caused*" by prices. In other models and datasets, spreads, volume, and quote arrivals are found to be significant variables. See, for example, Russell and Engle (1998) and Engle and Lunde (1999).

The estimates for the ACD(1, 1) are $\alpha = .0631$ and $\beta = .9309$ with robust t statistics of 22 and 296 respectively and a log likelihood of -71986. It is immediately clear that the persistence of this model is very high and that the coefficients are estimated very precisely. A somewhat better model is the component model presented in Table I with robust standard errors computed as in Theorem 1. The additional variables are significant and the likelihood is about 50 points higher. The persistence as measured by the largest root is .99915, which is very close to a unit root. In these models, just as in GARCH models however, this is not a point of discontinuity in the asymptotic distribution and consequently one can see that one is not included in the 95% confidence interval. The half-life of this process is 815 transactions, which, on average, would take less than 6 hours. Some of this persistence could well be due to changing regimes, which could be observed with economic determinants; however the dynamics of these variables would then also affect the persistence analysis.

The LB(15) test for autocorrelation in standardized residuals, defined by

$$\tilde{x}_i = x_i / \psi_i, \tag{31}$$

is 32.8 which exceeds the 5% point of 25 but is quite reasonable considering the large sample size. The same test for autocorrelation in squared standardized residuals is 31.2 with a similar interpretation.

The standardized durations are used to compute a density and a semiparametric hazard following the approach of Section 4. The empirical density with 2000 nearest neighbors on each side is plotted in Figure 1 with the hazard in Figure 2.

[3] EViews 3.1 is used for all calculations.

TABLE I

ACD ESTIMATES WITH COMPONENT STRUCTURE

Dependent Variable: SQR (DURS)
Method: ML-ARCH
Date: 01/28/99 Time: 12:44
Sample(adjusted): 2 52146
Included observations: 52145 after adjusting endpoints
Convergence achieved after 14 iterations
Bollerslev-Wooldridge robust standard errors and covariance

	Variance Equation			
	Coefficient	Std. Error	z Statistic	Prob.
ω	1.074	0.151	7.098	0.000
ρ	0.99915	0.00032	3073.364	0.000
φ	0.019	0.002	8.260	0.000
α	0.052	0.003	17.817	0.000
β	0.911	0.005	180.032	0.000
R-squared	−2.559	Mean dependent var		0.848
Adjusted R-squared	−2.559	S.D. dependent var		0.530
S.E. of regression	1.000	Akaike info criterion		2.759
Sum squared resid	52115.362	Schwarz criterion		2.760
Log likelihood	−71937.113	Durbin-Watson stat		0.473

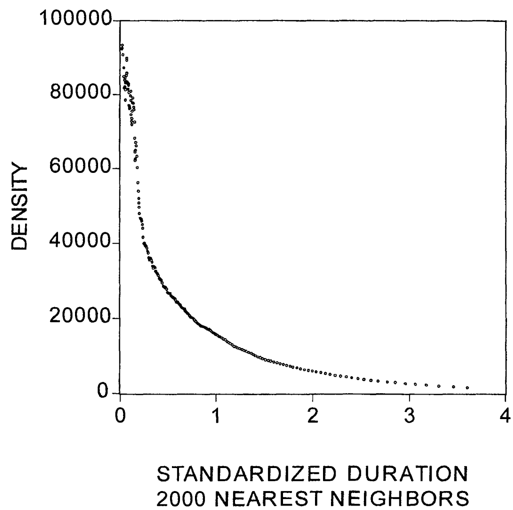

FIGURE 1

There is strong evidence of a sharp drop in the hazard for very small durations after which it only gradually declines. Such a picture is broadly consistent with a Weilbull density with parameter less than unity as was found in Engle and Russell. The actual shape of the Weibull hazard when $\gamma = .8$ is, however, not as abrupt as in the figures.

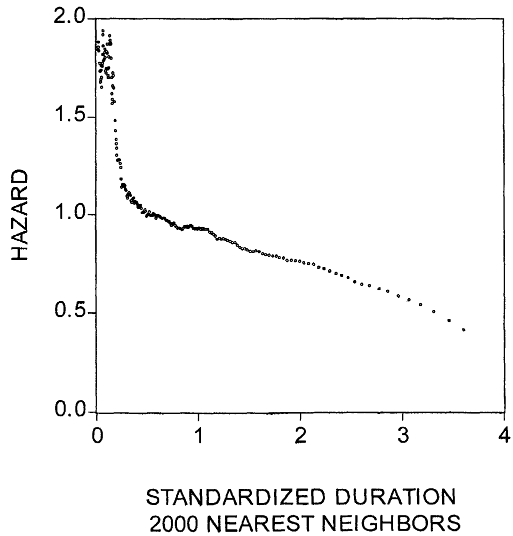

FIGURE 2

6. ESTIMATING PRICE VOLATILITY WITH TRANSACTION DATA

A. Theory

The most important mark available with each trade is the price at which it occurred. These prices convey information about the volatility of the market and about a variety of market microstructure hypotheses. In this section, a preliminary analysis of price data corresponding to the IBM trades will be undertaken. The analysis focuses on the relation between the timing of trades and the volatility of prices.

Since the time between trades is the reciprocal of the transaction rate, which is itself a proxy for volume, this study can draw on both the vast theoretical and empirical literature on the relation between volatility and volume. Much of the empirical literature is based on aggregated data, which shows that there is a strong contemporaneous relation between volume and volatility; e.g., see Lamoureux and Lastrapes (1990), Gallant, Rossi, and Tauchen (1992), and a survey by Karpoff (1987), although the predictive information in volume is much less clear.

A theoretical construct that is often used in modeling both volume and volatility is time deformation. Following the original ideas of Clark (1973) and Tauchen and Pitts (1983), the relation between economic time and calendar time is specified either as a latent process or as a function of observables. For example, Ghysels and Jasiak (1995) propose having time pass as a function of quote arrival rates while Müller et al. (1990) use absolute quote changes and geographical information on market closings.

With transactions data, models are typically estimated in transaction time without explicit account of the calendar time. For example, see Hasbrouck (1988, 1991), and Harris (1986). Hausman, Lo, and MacKinlay (1992) introduce the duration of the last trade as an exogenous explanatory variable but do not discuss its implications except to express concern that it may be endogenous. Pai and Polasek (1995) treat time as exogenous but allow the parameters of the process to depend upon durations in simple ways.

The market microstructure approach to models of time and volume are particularly useful as starting points. O'Hara (1995) points out that "if market participants can learn from watching the timing of trades, then the adjustment of prices to information will also depend on time." It is useful to formulate the price adjustment process in terms of the asymmetric information models introduced originally by Glosten and Milgrom (1985). In these models, the specialist sets bid and ask prices to use when trading with individuals who may be better informed than he is. In the original structure, there is new information revealed to a known fraction of the agents; the rest will buy or sell with a fixed probability. The information could be either good news or bad news. A randomly selected trader is offered a chance to trade. If he is a buyer, the specialist asks a price that reflects the probability that this trader could be informed. This inference is by Bayesian updating so the asking price is the posterior expected value given that the trader wants to buy. The more informed traders there are, the higher the asking price, given the observed trades. Eventually, the specialist infers the private information and sets bids and asks at the new true value. If the informed traders act distinctively, such as buying large volumes, then the specialist will rapidly discover the information and converge to the new price.

Several extensions of this model are relevant for this analysis. Diamond and Verrecchia (1987) note that sellers who do not own the stock must therefore short sell. If some fraction of the informed traders are prohibited from short selling, then they cannot profit from their information and when offered a chance to trade will not trade at the existing prices. Thus nontrades are evidence that the news may be bad; the specialist learns from the durations between trades and lowers his prices. This model can be summarized as *no trade means bad news*.

An alternative model is in Easley and O'Hara (1992) where there is also uncertainty as to whether there is information. Again, a fraction of the agents are informed and thus know whether there is news or not. When it is their turn to trade, they will buy if the news is good, sell if the news is bad, and not trade if there is no news. Long intervals between trades are consequently interpreted by the specialist as evidence that there is no news. The specialist therefore keeps prices relatively stable if the trading intervals are long and reduces the bid ask spread. This model can be interpreted as *no trade means no news*.

A third possibility is that frequent trading or short durations are a result of bunching of liquidity traders as proposed by Admati and Pfleiderer (1988). In this case, the trading is not associated with an information event and there is consequently no asymmetric information effect, hence volatility and spreads would be low just when the market is active. Conversely, slow trading would mean that the liquidity traders are staying away leaving a high proportion of informed traders. Consequently, spreads should be high and trades should have a big effect on prices because the trades are highly revealing. Therefore, *slow trading means informed trading and high volatility.*

B. Empirical Results

The goal of the following analysis is to determine a measure of price volatility using transaction data and discover how the timing of trades influences this volatility. Such models might therefore be called ultra-high-frequency GARCH or UHF–GARCH models. Letting the return from transaction $i-1$ to i be r_i, the conditional variance per transaction is defined as

$$(31) \qquad V_{i-1}(r_i | x_i) = h_i,$$

where the variance is conditioned on the current duration as well as the past returns and durations. Volatility is always measured over a fixed time interval and is frequently quoted in annualized terms. The conditional volatility per unit of time is thus the quantity to be estimated. This is naturally defined as

$$(32) \qquad V_{i-1}\left(\frac{r_i}{\sqrt{x_i}} \middle| x_i \right) = \sigma_i^2.$$

Consequently, these two conditional variances are related by

$$(33) \qquad h_i = x_i \sigma_i^2.$$

The predicted transaction variance conditional only on past prices and durations is then given by

$$(34) \qquad E_{i-1}(h_i) = E_{i-1}(x_i \sigma_i^2).$$

Both prices and durations have typical patterns over the trading day. Markets are more active in the morning and late afternoon than in the middle of the day and this shows up in both durations and volatilities. For the empirical analysis, durations are diurnally adjusted as described above. Returns must also be diurnally adjusted since volatility is known to have a daily shape. Because the natural measure of volatility is σ_i^2 in (32), the return data is divided by the square root of duration and then adjusted diurnally. The absolute value of $r/\sqrt{x}$ is regressed on the same splines as used for durations, and the adjusted series are obtained by dividing each return series by the spline prediction.

The daily splines for durations and returns are shown in Figures 3 and 4. The durations show very substantial differences over the day and have the typical pattern of high activity at the beginning and end of the day. At the open the average time between trades is about 10 seconds while at lunch it rises to almost 40 seconds. The volatility effect is much smaller but shows the same pattern. At

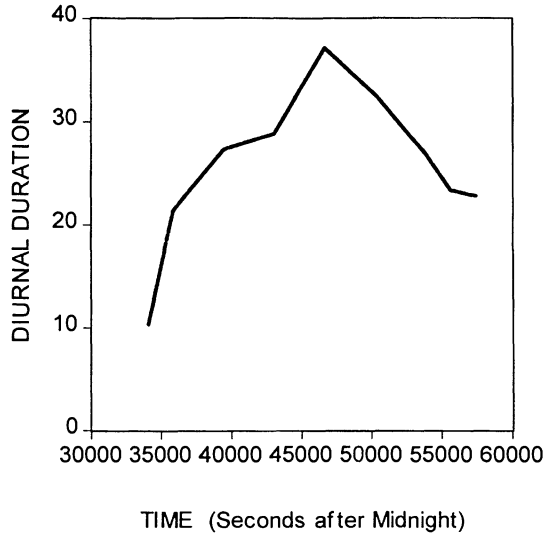

FIGURE 3

lunch volatility is about 40% of the morning peak. The volatility per trade, which is not shown, is lowest at the open and highest at the end of the day.

The prices are assumed to be best measured by the midquote which is the average of the bid and ask price at the time of the transaction. This choice of price measure reduces the econometric issues of bid ask bounce and price discreteness but it does not eliminate these problems. Both bid and ask prices

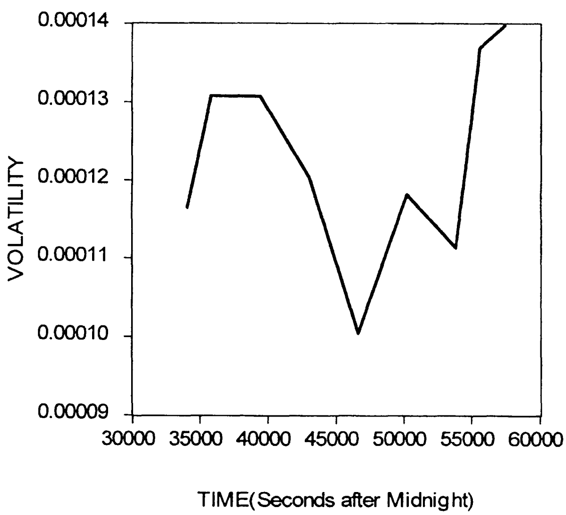

FIGURE 4

are quoted in eighths, and therefore the midquote is in sixteenths. Letting p_i be the log of the discretely measured midquote, an element of y_i, and letting m_i be the unobservable log price in continuously measured units,

$$p_i = m_i + \zeta_i, \tag{35}$$

where the errors result from the truncation. Assume the latent price to be a Martingale with respect to public information with innovation that can be written, without loss of generality, proportional to the square root of the time. The measurement errors may be autocorrelated since the truncation at one point in time is likely to be the same as the truncation several seconds later. As a rough approximation, assume the first difference over the square root of time is an ARMA(1, 1):

$$m_i = m_{i-1} + \sqrt{x_i}\, v_i, \qquad \sqrt{x_i}^{\,-1}\Delta\zeta_i \equiv \eta_i = \rho\eta_{i-1} + \xi_i + \chi\xi_{i-1}, \tag{36}$$

implying that measured returns per square root of time follow an ARMA(1, 1) process with innovations e:

$$\begin{aligned} r_i/\sqrt{x_i} &\equiv \Delta p_i/\sqrt{x_i} = v_i + \eta_i = \rho r_{i-1}/\sqrt{x_{i-1}} + v_i - \rho v_{i-1} + \xi_i + \chi\xi_{i-1}, \\ r_i/\sqrt{x_i} &= \rho r_{i-1}/\sqrt{x_{i-1}} + e_i + \phi e_{i-1}. \end{aligned} \tag{37}$$

The variance of r per unit of time is the expected value of the square of e.

Prices are modeled conditionally on the current duration as well as on past prices and durations as in (5). Thus equation (37) may also include the current duration. Economically, this reflects two effects. The longer the interval over which the return is measured, the higher the expected return since both the risky and riskless rates are measured per unit of time. However, if *no news is bad news* as in Diamond and Verrecchia, then long durations would imply declining prices.

A simple GARCH specification, where current durations are not informative, is:

$$\sigma_i^2 = \omega + \alpha e_{i-1}^2 + \beta\sigma_{i-1}^2, \tag{38}$$

which is a GARCH(1, 1) modified to take account of the irregular trading intervals and is the simplest UHF-GARCH. This model assumes that the news from the last trade is measured as the square of the last price innovation per second, and that the persistence of shocks is unaffected by the durations so that the coefficients can be taken as fixed parameters. This model can be estimated as a conventional GARCH(1, 1) with the dependent variable defined as returns divided by the square root of the time and an ARMA expression in the mean. Each observation is a transaction so the model is estimated in transaction time.

It is clear that this model does not yet admit the possibility that variations in x and variations in σ could be related to the same news events. A more realistic and a more interesting specification introduces durations directly into the conditional variance as

$$\sigma_i^2 = \omega + \alpha e_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma x_i^{-1}, \tag{39}$$

which allows the conditional variance to depend upon the reciprocal of duration. In this way a positive sign is expected under the Easley and O'Hara hypothesis since long durations indicate no news and lower volatility. A large duration cannot therefore drive the variance negative. A model with the level of durations was also estimated obtaining a negative coefficient, but this fit better. In this model, the persistence of volatility depends on the persistence of durations as well as the GARCH parameters.

A richer formulation allows both observed and expected durations to enter the model and also introduces a long run volatility variable:

$$\sigma_i^2 = \omega + \alpha e_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma_1 x_i^{-1} + \gamma_2 \frac{x_i}{\psi_i} + \gamma_3 \xi_{i-1} + \gamma_4 \psi_i^{-1} \tag{40}$$

where ξ_i is the long run volatility computed by exponentially smoothing r^2/x with a parameter .995. That is,

$$\xi_i = .005(r_{i-1}^2/x_{i-1}) + .995\xi_{i-1}. \tag{41}$$

Presumably, far better estimates of long run volatility can be developed by optimizing this parameter. The half-life of this smoother is 138 trades.

The impact of durations on volatility is incorporated in three coefficients that measure the effects of surprises in durations, reciprocal durations, and expected trade arrival rates, which is the reciprocal of expected durations. This model requires parameters from the ACD model in order to define expected durations.

Finally, it is possible to introduce economic variables into the model to improve the volatility forecasts. In the Easley and O'Hara framework, the volatility of prices will be greater, for a given sequence of trades, if the proportion of traders with information is known to be higher. If this is so, the bid ask spread will also be widened. Thus wide spreads can be interpreted as indicators that the proportion of traders with information is thought to be high and will consequently be a positive predictor of future volatility. Similarly, high volume trades are likely to be indicators of information and predictors of volatility.

The results are presented in Table II along with robust standard errors as discussed by Bollerslev and Wooldridge (1992). Although this model is estimated in transaction time, the asymptotic properties of GARCH models with exogenous variables should still apply. In Bollerslev and Wooldridge, the functional form of the variables is not explicitly mentioned and covariates are allowed as long as the limiting properties of scores and hessians are preserved.

The first three rows correspond to the variables that appear in the mean while the rest are in the variance. All results are calculated in EViews 3.1. Ljung Box statistics for serial correlation in the standardized residuals and in their squares are presented with p-values. The test for 15 lags is given but it is representative of most shorter or longer lags. Because the samples are so large, experimenta-

TABLE II

ESTIMATES OF UHF-GARCH VOLATILITY MODELS
(BOLLERSLEV-WOOLDRIDGE ROBUST STANDARD ERRORS)

	EQUATION (40)			EQUATION (39)			EQUATION (38)			EQUATION (42)		
Variable	Coef	Std. Err	Z-Stat	Coef	Std. Err	Z-Stat	Coef	Std. Err	Z-Stat	Coef	Std. Err	Z-Stat
MEAN												
DURS	**−0.004**	0.002	−1.892	**−0.006**	0.002	−3.082	**−0.008**	0.004	−1.892	**−0.007**	0.002	−4.027
AR(1)	**0.154**	0.022	6.900	**0.191**	0.021	8.932	**0.279**	0.023	12.29	**0.186**	0.022	8.507
MA(1)	**−0.564**	0.016	−35.13	**−0.590**	0.016	−37.16	**−0.656**	0.019	−33.86	**−0.570**	0.016	−35.70
VARIANCE												
C	**−0.224**	0.079	−2.835	**0.272**	0.039	6.970	**0.988**	0.092	10.74	**−0.111**	0.047	−2.358
ARCH(1)	**0.314**	0.014	22.13	**0.304**	0.014	21.84	**0.245**	0.020	12.33	**0.250**	0.013	18.73
GARCH(1)	**0.147**	0.013	11.00	**0.230**	0.014	16.04	**0.622**	0.025	24.70	**0.158**	0.014	11.71
1/DUR	**0.579**	0.029	19.69	**0.632**	0.026	23.84				**0.587**	0.028	21.27
DUR/EXPDUR	**−0.055**	0.009	−5.807							**−0.040**	0.005	−7.992
LONGVOL(−1)	**0.086**	0.014	6.235							**0.096**	0.011	8.801
1/EXPDUR	**0.484**	0.087	5.554									
SPREAD(−1) ≫										**0.736**	0.065	11.29
SIZE > 10000										**0.193**	0.119	1.624
LOGLIK	−107688.9			−108202.0			−112246.3			−107406.4		
LB(15)	67.156 0.000			33.680 0.001			93.092 0.000			40.810 0.000		
$LB^2(15)$	157.0 0.000			165.97 0.000			30.422 0.004			169.12 0.000		

tion with starting values is often needed to obtain global optima. Tighter convergence criteria and many iterations are used to obtain the best possible results.

These models reveal strong autocorrelation in the mean through the highly significant AR and MA coefficients. This is a familiar result in midquote models such as Hasbrouck (1991). The Ljung Box statistic for the returns data with 15 lags is 4892.6 and for the returns per square root of time diurnally adjusted is 2882.8. The mean is expressed as a function of the duration of the trade. This should be the drift in returns but also provides evidence of the bad news effect of long durations. Interestingly this is negative, supporting the Diamond and Verrecchia model; the robust t statistic varies from just under 2 to 4, suggesting support for the hypothesis.

Further interesting results are found in the variance equation. In all but the simplest model, the coefficients of the conventional GARCH model α and β sum to just under .5 which is very small for such a high frequency data set. This result is also familiar in the papers of Ghose and Kroner (1995), and Andersen and Bollerslev (1997), where it is found that the persistence of a GARCH model drops dramatically with intradaily data. In the volatility models from Table II, the persistence is made up from the duration and other variables. In two of the models, this is partly represented by the LONGVOL variable, which gives persistence directly. This variable is quite significant.

The reciprocal of duration is significantly positive in all specifications supporting the Easley and O'Hara formulation in which no trade is interpreted as no news so that volatility is reduced. Durations divided by expected duration is interpreted as the surprise in durations. This enters negatively in all cases showing a short run impact of durations in the same direction. The trading intensity also enters positively as the reciprocal of the expected duration indicating again that high transaction rates lead to high volatility.

These terms have rather different impacts for forecasting. Taking expectation of equation (40) conditional on the past gives

$$E_{i-1}(\sigma_i^2) = \omega + \alpha e_{i-1}^2 + \beta\sigma_{i-1}^2 + \gamma_1 E_{i-1}(x_i^{-1}) + \gamma_2 + \gamma_3 \xi_{i-1} + \gamma_4 \psi_i^{-1}, \tag{42}$$

revealing that γ_2, which is highly significant, has no persistence. On the other hand, γ_1 and γ_4 have long term impacts on future volatility since durations are themselves persistent. Because x is truncated from below (it can never be smaller than 1 second adjusted for the time of day), $E_{i-1}(x_i^{-1})$ exists, and, is related to ψ_i^{-1}. Thus forecasts of transaction arrival rates are an important component of volatility forecasts and surely contribute to their persistence.

Economic variables are introduced in the final model. Dummy variables for spreads greater than .003 and for trades bigger than 10,000 shares are added. The spread variable is highly significant although the size is only marginally important. The lack of significance of the size variable is surprising as it differs from work on similar data sets and models such as Dufour and Engle (1999) and Engle and Lunde (1999).

All of these models fail the Ljung Box diagnostic tests. However, they show dramatic improvement over the original data. The square of the dependent variable had a $LB(15)$ of 2644.5. The residuals from an ARMA(1,1) still had LB^2 (15) of 2730.6. Thus all the models represent major improvements even if they did not completely clean the residuals. With 50,000 observations, it is not clear how seriously to take these p-values. Interestingly, the models with higher likelihoods have worse diagnostic tests.

7. CONCLUSIONS

This paper has introduced a framework to estimate models for ultra-high-frequency or transactions data. When the data arrive at random times and these times themselves may carry information, the basic procedure is to model the associated variables called marks conditional on the times, and then to separately model the times.

In this example, 52,144 IBM stock transactions are analyzed to find a model of the timing of trades and then to measure the impact of this timing on the price volatility. The ACD model introduced by Engle and Russell (1998) is used to estimate the dependent point process for the arrival rates. A semiparametric approach to estimating the hazard function is introduced and applied.

Finally, the price quotes are examined to obtain models of volatility conditional on transaction times. Evidence is found for both short and long run components of volatility and that longer durations and longer expected durations are associated with lower volatilities as predicted by the Easley and O'Hara model. Economic variables also enter as expected from Easley and O'Hara. Higher bid ask spreads and larger volume both predict rising volatility.

Department of Economics, University of California, San Diego, Ja Jolla, CA 92093-0508, U.S.A.

Manuscript received July, 1997; final revision received March, 1999.

REFERENCES

ADMATI, ANAT R., AND PAUL PFLEIDERER (1988): "A Theory of Intraday Patterns: Volume and Price Variability," *The Review of Financial Studies*, 1, 3–40.

ANDERSEN, TORBEN G., AND TIM BOLLERSLEV (1997): "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 4, 115–158.

BOLLERSLEV, TIM, AND JEFFREY WOOLDRIDGE (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143–172.

CLARK, P. K. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–156.

COX, D. R. (1972): "Regression Models and Life Tables (with discussion)," *Journal of the Royal Statistical Society*, Series B, 34, 187–220.

DIAMOND, D. W., AND R. E. VERRECCHIA (1987): "Constraints on Short-selling and Asset Price Adjustments to Private Information," *Journal of Financial Economics*, 18, 277–311.

DUFOUR, ALFONSO, AND ROBERT ENGLE (1999): "Time and the Price Impact of a Trade," U.C.S.D. Discussion Paper.

EASLEY, D., AND M. O'HARA (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 905–927.

ENGLE, R. F., DAVID F. HENDRY, AND JEAN-FRANCOIS RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.

ENGLE, ROBERT, AND GARY LEE (1999): "A Long Run and Short Run Component Model of Stock Return Volatility," in *Causality, Cointegration and Forecasting, a Festschrift Honoring Clive W. J. Granger*, ed. by R. F. Engle and H. White. Oxford: Oxford University Press, pp. 475–497.

ENGLE, ROBERT, AND ASGER LUNDE (1999): "Trades and Quotes, A Bivariate Point Process," U.C.S.D. Discussion Paper.

ENGLE, R. F., AND J. R. RUSSELL (1997): "Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the ACD Model," *Journal of Empirical Finance*, 4, 187–212.

——— (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

GALLANT, R., P. E. ROSSI, AND G. TAUCHEN (1992): "Stock Prices and Volume," *Review of Financial Studies*, 5, 871–908.

GHOSE, DEVAJYOTI, AND KENNETH F. KRONER (1995): "Components of Volatility in Foreign Exchange Markets: An Empirical Analysis of High Frequency Data," Manuscript.

GHYSELS, E., AND J. JASIAK (1995): "Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets," in *High Frequency Data in Finance, Proceedings*. Zurich, Switzerland: Olsen and Associates.

GLOSTEN, L., AND P. MILGROM (1985): "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 13, 71–100.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681–700.

HARRIS, L. (1986): "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns," *Journal of Financial Economics*, 16, 99–117.

HASBROUCK, J. (1988): "Trades, Quotes, Inventories and Information," *Journal of Financial Economics*, 22, 229–252.

——— (1991): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179–207.

HAUSMAN, J. A., A. W. LO, AND A. C. MACKINLAY (1992): "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319–330.

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–389.

KALBFLEISCH, J. D., AND R. L. PRENTICE (1980): *The Statistical Analysis of Failure Time Data*. New York: Wiley.

KARPOFF, J. (1987): "The Relation Between Price Change and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22, 109–126.

LAMOUREUX, C., AND W. LASTRAPES (1990): "Heteroskedasticity in Stock Return Data: Volume Versus GARCH Effect," *Journal of Finance*, 45, 221–229.

LANCASTER, ANTHONY (1990): *The Econometric Analysis of Transition Data*, Econometric Society Monographs. Cambridge: Cambridge University Press.

LIU, T. C. (1963): "An Exploratory Quarterly Econometric Model of Effective Demand in the Postwar U.S. Economy," *Econometrica*, 31, 301–348.

——— (1969): "A Monthly Recursive Model of the U.S. Economy," *Review of Economics and Statistics*, 51, 1–13.

MÜLLER, U. A., M. M. DACOROGNA, R. B. OLSEN, O. V. PICTET, M. SCHWARZ, AND C. MORGENEGG (1990): "Statistical Study of Foreign Exchange Rates. Empirical Evidence of a Price Change Scaling Law and Intraday Analysis," *Journal of Banking and Finance*, 14, 1189–1208.

O'HARA, M. (1995): *Market Microstructure Theory*. Oxford, England: Basil Blackwell.

PAI, J. S., AND W. POLASEK (1995): "Irregularly Spaced AR and ARCH (ISAR-ARCH) Models," in *High Frequency Data in Finance, Proceedings*. Zurich, Switzerland: Olsen and Associates.

RUSSELL, J. R., AND R. F. ENGLE (1999): "An Econometric Analysis of Discrete-Valued Irregularly Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinomial Model," U.C.S.D. Manuscript.

SHEPHARD, N. E. (1993): "Fitting Nonlinear Time-Series Models with Applications to Stochastic Variance Models," *Journal of Applied Econometrics*, 8, Suppl., 135–152.

TAUCHEN, G. E., AND M. PITTS (1983): "The Price Variability Volume Relationship on Speculative Markets," *Econometrica*, 51, 485–505.