

NBER WORKING PAPER SERIES

**FORECASTING TRANSACTION
RATES: THE AUTOREGRESSIVE
CONDITIONAL DURATION MODEL**

**Robert F. Engle
Jeffrey R. Russell**

Working Paper No. 4966

**NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1994**

The authors would like to thank Sir David Cox, David Hendry, Glenn Sueyoshi, Clive Granger, Bruce Lehman, Alex Kane, Niel Sheppherd, and George Borjas for their suggestions. Useful comments were also received from participants at the Conference on Multivariate Time Series and Financial Econometrics sponsored by the University of California San Diego and the 1994 USC/UCLA/NYSE Conference on Market Microstructure. The research was supported in part by National Science Foundation Grant SES-9122056. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1994 by Robert F. Engle and Jeffrey R. Russell. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

FORECASTING TRANSACTION
RATES: THE AUTOREGRESSIVE
CONDITIONAL DURATION MODEL

ABSTRACT

This paper will propose a new statistical model for the analysis of data that does not arrive in equal time intervals such as financial transactions data, telephone calls, or sales data on commodities that are tracked electronically. In contrast to fixed interval analysis, the model treats the time between observation arrivals as a stochastic time varying process and therefore is in the spirit of the models of time deformation initially proposed by Tauchen and Pitts (1983), Clark (1973) and more recently discussed by Stock (1988), Lamoureux and Lastrapes (1992), Muller et al. (1990) and Ghysels and Jasiak (1994) but does not require auxiliary data or assumptions on the causes of time flow. Strong evidence is provided for duration clustering beyond a deterministic component for the financial transactions data analyzed. We will show that a very simple version of the model can successfully account for the significant autocorrelations in the observed durations between trades of IBM stock on the consolidated market. A simple transformation of the duration data allows us to include volume in the model.

Robert F. Engle
Department of Economics, 0508
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER

Jeffrey R. Russell
Department of Economics, 0508
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508

1. Introduction

With the increasing power and memory of computers it is natural to collect and analyze data at ever higher frequencies. The ultimate in high frequency data collection records every transaction. This limit has now been reached for many financial data sets as well as for various microeconomic transactions such as telephone calls, purchases of items on credit cards and other types of transactions which are automatically recorded by computers. Generally the objective is to model and forecast the frequency and distribution of transactions possibly as a function of other determinants.

The analysis of such data sets presents new and interesting econometric challenges. In particular, most transactions data arrive at irregular intervals. Because our standard methods of analysis are based upon equally spaced data, there is a natural inclination by the econometrician to choose some time interval and aggregate the transactions occurring within that interval. This is a potentially important part of the analysis since the appropriate interval will depend upon the transaction frequency. For purchases of consumer durables by an individual, a natural interval might be months or even years. On the other extreme, stock transactions for frequently traded stocks will have transactions every few seconds, hence a much shorter interval is appropriate. If too short an interval is selected, most cells will be zero, but if too long an interval is chosen, then there will be a dramatic loss of information which mitigates the advantages of moving to transactions data in the first place.

The problem is even more important when it is recognized that the frequencies of transactions may vary over time. For stocks, transactions are more frequent near the open and the close than in the middle of the day. For currencies which are traded twenty four hours per day there are definite periods of slow activity and weekends are typically slow. Even more intriguing is the case of transactions which are generally infrequent but which may suddenly see very high rates of activity. This may be due to some unobservable event such as a news release or to an unobservable event which may best be thought of as a stochastic process. In these cases,

the choice of a fixed interval for data analysis is very perilous as it may leave the investigator with many uninformative data points or disguise the periods of most interest.

This paper will propose an alternative to fixed interval analysis. The proposed method is called Autoregressive Conditional Duration. This procedure models the time intervals directly and therefore is in the spirit of the models of time deformation initially proposed by Tauchen and Pitts (1983), Clark (1973) and more recently discussed by Stock (1988), Muller et al. (1990) and Ghysels and Jasiak (1994) but does not require auxiliary data or assumptions on the causes of time flow.

Instead of selecting a fixed interval for analyzing the data, it is proposed to let the interval between transactions be a random variable to be analyzed. Thus the data set becomes a list of durations and characteristics of each transaction. That is, there is some probability of a transaction in each instant of time and this probability could vary with the type of transaction, the length of time since the previous transaction and any other outside influences. As each interval is naturally thought of as a duration, it is clear that the vast body of research on duration models is immediately potentially relevant.

To illustrate the model and provide an example, transactions data from financial markets will be analyzed. Financial transactions data is particularly interesting to examine because of the recent rekindling of interest in the micro theory surrounding financial transactions data. As proposed by Kyle (1985) Admati and Pfleiderer (1988) and more recently by Easley and O'Hara (1992) the durations between transactions may carry information about the state of the market. These models suggest clustering of trading. Clustering of trading is precisely what we find when we examine IBM transactions data. Even after the time of day effects are partialled out, large autocorrelations in no-trade durations exist.

In the following section the autoregressive duration (ACD) model will be developed. Section 3 will discuss properties of the ACD model. Section 4 will present the likelihood function associated with the ACD model. Section 5 will briefly discuss specification testing with the ACD model. Section 6 will discuss estimation and present examples using the IBM time series discussed above. Section 7 will present some LM tests for the IBM transactions data. Section 8 will present an alternative model to include quantity effects in the ACD model. Section 9 will discuss forecasting using an ACD model. Section 10 will draw conclusions and discuss future research directions.

2. Model Formulation

Consider the stochastic process which is simply a sequence of times $\{t_1, t_2, \dots, t_n, \dots\}$. As these are points distributed in time, this is called a "point process" and the times are called "arrival times" of the point process. Corresponding to these arrival times is a counting process, $N(t)$ which is the number of events which have occurred by the time t . Obviously, $N(t)$ is a non-decreasing function of time that increases in unit steps. Define the unconditional expectation of $N(t)$ as:

$$(1) \quad \Lambda(t) = E(N(t))$$

The counting function representation of the point process is illustrated in figure 1. In the top of the figure we have illustrated the number of events that have occurred by time t . Each vertical jump on this step function represents an event.

Before describing this stochastic process it is useful to define some terms as in for example Todorovic (1992).

This process is said to be "stochastically continuous" if at any point t_0 , and for any $\epsilon > 0$,

$$(2) \quad \lim_{t \rightarrow t_0} P\{N(t) - N(t_0) > \epsilon\} = 0$$

The process is defined to have "independent increments" if *for any* finite number of times $t_0 < t_1 < t_2 < \dots < t_n$ the random variables,

$$N(t_0), N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1})$$

are independent. A process that has both of these characteristics is called a "Poisson Process." A particularly convenient type of Poisson Process is the "time homogeneous" process in which $\Lambda(t) = \lambda t$. In fact, any Poisson Process can be transformed into a time homogenous process by what might be called a deterministic time deformation.

But, is the class of Poisson Processes appropriate for the types of transaction data to be examined? The assumption of stochastic continuity seems innocuous, but the assumption of independent increments is clearly questionable for most transactions processes. In particular, it

implies that the distribution of the number of transactions in a future time period is independent of the transaction rate in the past. This is a very unattractive assumption for a process which seeks to model the frequency of transactions. Essentially, it says there is nothing to model except perhaps a deterministic function of time. The bottom of figure 1 illustrates the implications of this assumption. In particular, if $N_1(t)$ and $N_2(t)$ are two realizations of the counting process, the expected value of the counting function at time t might be the line labeled $\Lambda(t)$. The slope is the expected rate of arrival of events which is constant under the assumption of independent increments.

In particular, the assumption implies that intervals or durations should not be autocorrelated, at least after any deterministic components are removed. For the IBM stock transactions to be discussed below, there is clear evidence of autocorrelation in the durations, with or without elimination of deterministic functions of time.

In order to introduce generalizations of the Poisson Process, it is useful to write the probability density function of the durations. Let the i^{th} duration be $x_i = t_i - t_{i-1}$ and define $f_i(x_i)$ as the density of x_i . The cumulative density function of x for a time homogeneous Poisson Process has the form:

$$(3) \quad P(x_i < t) = F_{i(t)} = 1 - \exp(-\lambda t)$$

where all x_i are independent and identically distributed. Note that:

$$(4) \quad x \sim \lambda \exp(-\lambda x), \text{ an exponential random variable}$$

$$E(x) = \lambda^{-1}$$

$$V(x) = \lambda^{-2}$$

$$L(x, \lambda) = N \log(\lambda) - \lambda \sum_{i=1}^N x_i$$

In the last expression L is the log likelihood of the sample of N durations under the assumption

that they are a homogeneous Poisson Process. The MLE of λ is $N / \sum_{i=1}^N x_i$.

It is natural to extend this model to allow λ to vary over spells. The most popular approach suggested by Cox (1972) is to introduce covariates which are treated as exogenous (or "weakly" exogenous as in Engle Hendry and Richard (1983)). In labor econometric studies variables are introduced for observable heterogeneity. See for example Kiefer (1988) or Lancaster (1990) for surveys. The model is called the Proportional hazard model and is written as

$$(5) \quad \lambda_i = \lambda \exp \beta' z_i$$

where z is a vector of attributes appropriate to the i^{th} spell. This correction does not break the assumption that durations are independent at least conditional on the z 's, and will therefore not be successful for these transactions data.

There are at least two suggestions in the literature for introducing time dependence into λ without simply making it a deterministic function of time. The first is the doubly stochastic Poisson Process of Cox (1972a) as studied by Grandell(1976). In this case λ_i is replaced by an *independent* Poisson Process. While this seems a reasonable model, it is very difficult to deal with and is complex to estimate and forecast. Furthermore, the independence assumption is unmotivated. A second suggestion in Cox (1972a) is to incorporate lagged durations into the z vector of the proportional hazard model. This allows recent experience to influence the current hazard. Models of this type were introduced into econometrics by Heckman and Borjas (1980) and Heckman (1981) among others, to examine the impact of past unemployment on current durations. In these models, the data are typically short time series on many individuals so that the question of whether this truly reflects state dependence or merely unmeasured heterogeneity becomes very important.

These models and many others can of course be expressed in terms of the joint likelihood function

$$(6) \quad L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n l_i = \prod_{i=1}^n \log f_i(x_i | x_{i-1}, \dots, x_1; \theta)$$

It is here proposed to parameterize these conditional likelihoods directly. Clearly, in the Poisson case, each conditional likelihood is an exponential as in (5). A rather weak assumption is that the time dependence can be summarized by a function ψ which maps the past information set

into a scalar with the property that x_i/ψ_i are independent and identically distributed with a mean of unity. That is, the density of these "standardized durations" satisfy:

$$(7) \quad g(x_i / \psi_i | x_{i-1}, \dots, x_1; \theta) = g(x_i / \psi_i; \theta)$$

$$(8) \quad E[x_i | x_{i-1}, \dots, x_1; \theta] = \psi_i = \psi(x_{i-1}, x_{i-2}, \dots, x_{i-p}, \psi_{i-1}, \psi_{i-2}, \dots, \psi_{i-q}; \theta)$$

and ψ can be called the conditional duration and since it depends upon past durations, it is natural to call the model Autoregressive Conditional Duration or ACD.

A simple version of this with exponential durations and parameterized in terms of $\psi = \lambda^{-1}$, is:

$$(9) \quad L(x, \psi) = \sum_{i=1}^N \ln \psi_i + \sum_{i=1}^N \frac{x_i}{\psi_i}$$

$$(10) \quad \psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1} \text{ for } \alpha, \beta \geq 0, \omega > 0, \forall i, i = 1 \dots N$$

This model is convenient because it allows various moments to be calculated by expectation. For example, the conditional mean of x_i is ψ_i , the conditional duration, but the unconditional mean is $\omega/(1-\alpha-\beta) = \mu$. This is most easily seen by taking expectations of both sides of equation (10) although the dynamic analysis in the proof of lemma 1 is needed to insure that it exist when $\alpha+\beta < 1$. Similarly, the conditional variance of x is ψ_i^2 but the unconditional variance is given by σ^2 where

$$(11) \quad \sigma^2 = \mu^2(1 - \beta^2 - 2\alpha\beta) / (1 - \beta^2 - 2\alpha\beta - 2\alpha^2)$$

Thus whenever $\alpha > 0$ the unconditional standard deviation will exceed the mean exhibiting "excess dispersion" as so often noticed in duration data sets. From taking repeated expectations multistep forecasts of durations can be computed directly. That is, the expected duration of the n^{th} transaction can be computed directly from (9) and (10). This model has only one lag of x and one lag of ψ and is therefore conveniently labeled ACD(1,1).

Readers who are familiar with the ARCH class of models will immediately recognize the relationship to models of conditional variance. This ACD(1,1) is analogous to the GARCH(1,1) and will have many of the same properties. Just as the GARCH(1,1) is often a good starting

point, the ACD(1,1) seems like a natural starting point. However, as there are very many alternative models for volatilities, there are lots of interesting possibilities here. For recent surveys on ARCH models and lists of different classes, see Bollerslev, Engle and Nelson(1994), Bollerslev Chou and Kroner(1992) and Bera and Higgins (1992). The ARCH model was originally introduced by Engle(1982).

The specifications in (9) and (10) can be generalized in many ways. The conditional density in (9) is assumed to be exponential but there are countless ways to relax this restriction. The most popular is to assume that the conditional distribution is Weibull. In this case, a useful way to describe this assumption is that $(x_i/\phi_i)^\gamma$ is exponential for ϕ_i a measurable function of the past. Thus the conditional density of x_i is

$$(12) \quad g(x_i) = \frac{\gamma}{x_i} \left(\frac{x_i}{\phi_i} \right)^\gamma \exp \left\{ - \left(\frac{x_i}{\phi_i} \right)^\gamma \right\} \text{ for } x_i > 0$$

The conditional density is now a two parameter family which can exhibit either increasing or decreasing hazard functions. This makes especially long durations more or less likely than for the exponential depending on whether γ is greater or less than unity respectively.

For this model

$$(13) \quad E(x_i | \mathcal{F}_{i-1}) = \Gamma \left(1 + \frac{1}{\gamma} \right) \phi_i \equiv \psi_i$$

where $\Gamma(\cdot)$ is the gamma function. If the function ϕ is parameterized analogously to (10), then

$$(14) \quad \phi_i = \bar{\omega} + \tilde{\alpha}x_{i-1} + \beta\phi_{i-1}$$

and substituting

$$(15) \quad \phi_i = \frac{\psi_i}{\Gamma \left(1 + \frac{1}{\gamma} \right)} \text{ gives}$$

$$(16) \quad \psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}$$

where

$$\omega = \tilde{\omega} \Gamma \left(1 + \frac{1}{\gamma} \right)$$

$$\alpha = \tilde{\alpha} \Gamma \left(1 + \frac{1}{\gamma} \right)$$

which is exactly the same as (10).

Many other parametric specifications are possible. An attractive alternative is to estimate a non-parametric hazard. While this has been proposed and used in many studies, there are no examples known to these authors in which the hazard varies over time stochastically. Nevertheless, it would not be difficult to extend the ACD model to this case. The needed assumption is that when the correct specification for ψ_i is found, then x_i/ψ_i will be independent and the distribution of such random variables can be estimated by some form of histogram estimator such as a spline or kernel. This estimated histogram can then be used as the likelihood function to re-estimate the parameters in ψ . The procedure is closely related to the semi-parametric ARCH model estimated by Engle and Gonzalez-Rivera(1991).

The dynamic specification of the conditional duration can also be easily generalized. A (p,q) linear specification would be:

$$(17) \quad \psi_i = w + \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \dots + \alpha_p x_{i-p} + \beta_1 \psi_{i-1} + \beta_2 \psi_{i-2} + \dots + \beta_q \psi_{i-q}$$

More generally, this could be a non-linear function and it could include other variables as in

$$(18) \quad \psi_i = \psi(x_{i-1}, x_{i-2}, \dots, x_{i-p}, \psi_{i-1}, \psi_{i-2}, \dots, \psi_{i-q}, z_i, z_{i-1}, \dots, z_{i-1}, \theta)$$

Such a specification includes models analogous to the EGARCH, AGARCH, NGARCH, TGARCH and many others as possibilities. More interestingly, it allows economic variables to enter the equation which determines the frequency of transactions. From this version of the model one can test hypothesis on economic determinants of the rates of transactions. This is probably the main reason why the analysis is being undertaken and thus, it is potentially very important to consider how such variables should be specified and estimated.

3. Distribution of the ACD Models

In this section we develop properties of the unconditional moments of an ACD process. These will be convenient for data analysis.

Lemma 1: If $E(x_i | \mathcal{F}_{i-1})$ exists and equals ψ_i ,

where $\psi_i = \omega + \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \dots + \alpha_p x_{i-p} + \beta_1 \psi_{i-1} + \beta_2 \psi_{i-2} + \dots + \beta_q \psi_{i-q}$
and if all the roots of the equations

$$(19) \quad 1 - \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) z^i = 0$$

lie outside the unit circle, then x has a constant unconditional mean given by

$$(20) \quad \mu = \frac{\omega}{1 - \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)}$$

A necessary condition that all the roots lie outside the unit circle is

$$(21) \quad \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) < 1$$

Proof : See Appendix

Similar results can be obtained for higher moments although we establish these only for the first order case in the following lemma.

Lemma 2: Under the conditions of lemma 1 and when $V(x_i | \mathcal{F}_{i-1})$ exists and equals $\kappa \psi_i^2$ and when $p=q=1$, and $(\alpha + \beta)^2 + \alpha^2 \kappa < 1$, then

$$(22) \quad V(x_i) = \mu^2 \kappa \left\{ \frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2 - \alpha^2 \kappa} \right\}$$

Proof: See Appendix

This result shows that the unconditional standard deviation exceeds the mean by more than does the conditional standard deviation. This model therefore is consistent with the excess dispersion that is typically observed in duration data. For the exponential model, $\kappa=1$ so there is no excess dispersion in the standard deviations even though there is in the unconditional durations. For the Weibull with parameter γ , it can be shown that

$$(23) \quad \kappa = \left\{ \frac{\Gamma\left(1 + \frac{2}{\gamma}\right)}{\Gamma\left(1 + \frac{1}{\gamma}\right)^2} - 1 \right\}$$

Similar but increasingly complicated expressions can be obtained for the higher moments and for higher order models.

4. The Likelihood Function of an ACD Model

This section will develop the likelihood functions for the ACD models described above. In general, x_i will not be independently distributed. The joint likelihood function can always be written as the product of the conditional likelihood.

Hence the joint log likelihood function for the exponential version can be written as

$$(24) \quad L(\theta) = - \sum_{i=1}^N \left[\ln(\psi_i) + \frac{x_i}{\psi_i} \right]$$

Differentiating with respect to θ

$$(25) \quad \frac{\partial l_i}{\partial \theta} = \left(\frac{x_i}{\psi_i} - 1 \right) \frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \theta}$$

$$(26) \quad \frac{\partial^2 l_i}{\partial \theta \delta \theta'} = \left[-\frac{x_i}{\psi_i} \right] \frac{1}{\psi_i^2} \frac{\partial \psi_i}{\partial \theta'} \frac{\partial \psi_i}{\partial \theta} + \left[1 - \frac{x_i}{\psi_i} \right] \frac{1}{\psi_i^2} \frac{\partial \psi_i}{\partial \theta'} \frac{\partial \psi_i}{\partial \theta}$$

where

$$(27) \quad \frac{\partial \psi_i}{\partial \theta} = z_i + \sum_{i=1}^p \beta_i \frac{\partial \psi_{i-1}}{\partial \theta}$$

and

$$(28) \quad z_i = \left[1, x_{i-1}, x_{i-2}, \dots, x_{i-p}, \psi_{i-1}, \psi_{i-2}, \dots, \beta_q \psi_{i-q} \right]$$

This recursive score function is analogous to that of Bollerslev's (1986) GARCH model.

Under correct specification, the standardized series $\left(\frac{x_i}{\psi_i} \right)$ is iid with an exponential distribution and expectation of one. It is easy to see that the score will have expectation of zero and the information matrix can be estimated by.

$$(29) \quad \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{\psi_i^2} \frac{\partial \psi_i}{\partial \theta'} \frac{\partial \psi_i}{\partial \theta} \right]$$

For the WACD the log likelihood function is

$$(30) \quad l_i(\theta, \gamma) = \ln \left(\frac{\gamma}{x_i} \right) + \gamma \ln \left(\frac{x_i}{\phi_i} \right) - \left(\frac{x_i}{\phi_i} \right)^\gamma$$

Differentiating with respect to γ and θ

$$(31) \quad \frac{\delta l_i}{\delta \gamma} = \frac{1}{\gamma} - \left(\left(\frac{x_i}{\phi_i} \right)^\gamma - 1 \right) \ln \left(\frac{x_i}{\phi_i} \right)$$

$$(32) \quad \frac{\delta l_i}{\delta \theta} = \left(\left(\frac{x_i}{\phi_i} \right)^\gamma - 1 \right) \frac{\gamma}{\phi_i} \frac{\delta \phi_i}{\delta \theta}$$

$$(33) \quad \left(\frac{\delta^2 l_i}{\delta \gamma^2} \right) = - \left(\frac{1}{\gamma^2} + \left(\frac{x_i}{\phi_i} \right)^\gamma \left(\ln \left(\frac{x_i}{\phi_i} \right) \right)^2 \right)$$

$$(34) \quad \left(\frac{\delta^2 l_i}{\delta \gamma \delta \theta'} \right) = \left[\frac{\gamma}{\phi_i} \left(\frac{x_i}{\phi_i} \right)^\gamma \ln \left(\frac{x_i}{\phi_i} \right) + \left(1 - \left(\frac{x_i}{\phi_i} \right)^\gamma \right) \frac{1}{\phi_i} \right] \frac{\delta \phi_i}{\delta \theta}$$

and

$$(35) \quad \frac{\delta^2 l_i}{\delta \theta \delta \theta'} = \left(\frac{\gamma}{\phi_i} \right)^2 \left(\frac{x_i}{\phi_i} \right)^\gamma \frac{\delta \phi_i}{\delta \theta} \frac{\delta \phi_i}{\delta \theta'} + \left(\left(\frac{x_i}{\phi_i} \right)^\gamma - 1 \right) \frac{\delta}{\delta \theta} \left[\frac{\gamma}{\phi_i} \frac{\delta \phi_i}{\delta \theta} \right]$$

where

$$(36) \quad \frac{\partial \phi_i}{\partial \theta} = z_i + \sum_{i=1}^p \beta_i \frac{\partial \phi_{i-1}}{\partial \theta}$$

and θ is defined as above.

Noting that the conditional expectation of $\left(\frac{X_{i+1}}{\phi_{i+1}} \right)^\gamma$ is one, it can be shown that the expectation of the score is zero. The Hessian can be estimated by

$$(37) \quad \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} -\frac{1}{\gamma^2} - \left(\frac{x_i}{\phi_i} \right)^\gamma \left(\ln \left(\frac{x_i}{\phi_i} \right) \right)^2 & \frac{\gamma}{\phi_i} \left(\frac{x_i}{\phi_i} \right)^\gamma \ln \left(\frac{x_i}{\phi_i} \right) \frac{\delta \phi_i}{\delta \theta} \\ \frac{\gamma}{\phi_i} \left(\frac{x_i}{\phi_i} \right)^\gamma \ln \left(\frac{x_i}{\phi_i} \right) \frac{\delta \phi_i}{\delta \theta} & \left(\frac{\gamma}{\phi_i} \right)^2 \frac{\delta \phi_i}{\delta \theta} \frac{\delta \phi_i}{\delta \theta'} \end{bmatrix}$$

The expectation of the Hessian is

$$(38) \quad \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} (\zeta + 1) \frac{1}{\gamma^2} & v \frac{1}{\phi_i} \frac{\delta \phi_i}{\delta \theta} \\ v \frac{1}{\phi_i} \frac{\delta \phi_i}{\delta \theta} & \left(\frac{\gamma}{\psi_i} \right)^2 \frac{\delta \phi_i}{\delta \theta} \frac{\delta \phi_i}{\delta \theta'} \end{bmatrix}$$

where¹ $\zeta \equiv E[\varepsilon \ln(\varepsilon_i)^2] = \int_0^{\infty} y \ln(y)^2 \exp(-y) dy \approx .8237$

and $v \equiv E[\varepsilon_i \ln(\varepsilon_i)] = \int_0^{\infty} y \ln(y) \exp(-y) dy = 1 - e \approx .4227$.

where e is Euler's constant.

Because the off diagonal element will not be zero, efficient estimation requires the joint estimation of the parameters θ and γ . Furthermore, if the model were truly Weibull and one estimated an Exponential, the parameters θ will be biased.

5. Specification Testing Using Lagrange Multiplier Tests.

The Lagrange Multiplier test has proven to be a useful diagnostic tool. See for example Breusch and Pagan(1978, 1980), Godfrey(1978, 1985), and Engle(1979). Many possible tests come to mind. A test of primary interest is a test for misspecification in the form of omitted variables. In the context of financial data sets, we might like to test for the omission of weakly exogenous variables such as volume, price, or spread in the context of financial data. A second test for a more general form of misspecification will look at deviations from moment conditions implied by the model specification. The latter tests are known as tests for unexplained heterogeneity, or excess dispersion in the duration literature.

For the Exponential model, the LM test can be reduced to examining NR^2 where R^2 is the R-squared obtained from regressing a vector of generalized errors defined as

¹ Calculated numerically using Mathematica.

$$(39) \left[\left(1 - \frac{x_1}{\psi_1} \right), \dots, \left(1 - \frac{x_N}{\psi_N} \right) \right]'$$

on the vector of partial derivatives

$$(40) \left[\frac{1}{\psi_1} \frac{\delta \psi_1}{\delta \theta}, \dots, \frac{1}{\psi_N} \frac{\delta \psi_N}{\delta \theta} \right]$$

where both are evaluated at the null.

For both the Weibull and the Exponential, it is well known that the LM statistic can be calculated by $N \cdot R^2$ where R^2 is the R-squared from regressing a vector of ones on the scores evaluated at the null. The R-squared can be obtained from the first iteration of the BHHH algorithm taken from the maximum likelihood estimates under the null. The test statistic will have the usual chi-squared distribution.

It is easily shown that just as for GARCH models and ARMA models, we can not derive a general LM test for the alternative of an ACD(p,q) process.

Under correct specification, the mean of the generalized errors should be equal to the variance which should be equal to one. In the duration literature, tests of this condition are known as tests for excess dispersion or unexplained heterogeneity. There is a simple form of this test based on the LM test initially proposed by Lancaster(1985). The test looks for multiplicative excess heterogeneity in the hazard specification.

The generalized likelihood function can be approximated by

$$(41) P(x_i; \theta, \gamma, \sigma^2) = \log(g(x_i)) + \log\left(1 + \frac{\sigma^2}{2}(\epsilon_i^2 - 2\epsilon_i)\right)$$

Under the null $\sigma^2 = 0$ the Hessian is block diagonal with respect to the additional parameter.

6. Application to Duration Between Successive Trades of IBM Stock.

The data examined are the trading times, measured in seconds after midnight, of IBM stock through the consolidated transaction system. These data were abstracted from the Trades, Orders, Reports, and Quotes (TORQ) data set compiled by Joel Hasbrouck and the NYSE. There are over 50,000 transactions of IBM stock over the 3 monthly period of November 1990 through January 1991. We examine the durations between successive transactions. These durations are measured in seconds. The duration between the close of the market one day and opening the following day is deleted. The execution of a limit order was not treated any differently from market orders. Estimation was performed on two sub samples of the data both containing 15,000 observations. The first is approximately the month of November, and the second sample is approximately the month of December. Estimating parameters for two data sets will allow us to compare results. Furthermore, the first data set contains days surrounding Thanksgiving day. The Friday following Thanksgiving day exhibited a lull in activity. In particular, the 3 most extreme durations were observed on this day, the longest of which is almost 200 times the mean. These extreme values could potentially be modeled by including other explanatory variables.

The first sample has an average time interval between trades of 26.2 seconds, the minimum interval is 0 seconds, and the maximum interval is 4592 seconds or just over 1 hour and 15 minutes. The standard deviation is 53.9 seconds and the skewness is 42.3. The top portion of Figure 2 presents a histogram of the durations. The second sample has an average of 27.07 seconds with a minimum of 0 seconds and a maximum of 426 seconds or just over 7 minutes. The standard deviation is 37.69 seconds. A histogram for these durations is on the bottom of figure 2. For a Poisson process, the mean should be equal to the standard deviation. Both of these data sets clearly exhibit excess variance in durations.

The autocorrelations and partial autocorrelations of the duration between trades are presented in figure 3. Both of the Ljung-Box statistics are very large. The first data set has a Lung-Box statistic of 1058 and the second of 2413. The null hypothesis of white noise is easily rejected for both data sets based on the critical value of 24.99 at the 5% level. These long sets of positive autocorrelations are precisely what one finds for autocorrelations of squared returns. The cause is volatility clustering in that case and interval or duration clustering in this case.

These samples exhibit the well known property of high activity in the morning and just prior to close². This can be seen by the very different means for different hours of the day presented in figure 4. The mean duration is twice as high at mid day than that of the morning. These deterministic components were partialled out using a linear spline as a function of time of day. For both data sets the spline exhibits the expected inverted "U" shape. The graphs for each sub sample are presented in figure 5. Durations are longest in the middle of the day and shorter just after opening, and just prior to close. The durations are 4 times longer on average in the middle of the day than in the morning and the end of the trading day has trading averages twice as long as in the morning. In particular, the average duration ranges from 10 seconds in the morning to almost 40 seconds around noon.

We shall consider modeling the transformed series

$$(42) \quad \bar{x}_i = \frac{x_i}{\Phi(t_{i-1})}$$

where $\Phi(t)$ is an estimate of $E[x_i | t_{i-1}]$. Both the new series \bar{x}_i have a mean of approximately 1. As expected, the first data set (containing the outliers) has a higher standard deviation of 1.98 compared to 1.37 in the second. The new Ljung-Box statistic associated with the series is 1271 for the first data set and 1169 for the second data set. The Lung-Box statistic actually increased for the first data set after partialing out the time of day effects. In both cases, the Lung-Box statistic remains very high relative to the critical value.

Several models were estimated using maximum likelihood. Due to the non linearity of the model, the BHHH algorithm was used with numerical derivatives. The algorithm has no trouble converging for these samples and the results appear robust to initial values imposed. To allow for inter day effects, a dummy variable was included taking the value one if it is the first observation of the day and 0 otherwise.

² For example, see Jain and Joh (1986)

6.1 Estimation of the Exponential ACD (EACD) Model.

The simplest model examined was the EACD(1,1). Estimates for both sample periods and the results are presented in figure 6. There is a significant negative inter day effect. For a constant unconditional mean to exist, lemma 1 requires the condition $(\alpha + \beta) < 1$. Both models are close to being integrated with the sum of α and β equal to .9960 and .9895 for the first and second samples respectively. The implied unconditional means for the two samples, calculated using results of lemma 1, are 1.80 and 1.11 respectively. Both of these values are much larger than the expected value of 1 under correct specification. To further examine the fit of the model, we examine the standardized series

$$(43) \quad \varepsilon_i = \frac{\bar{x}_i}{\phi_i}$$

Because the distribution of ε_i does not depend on past observations, the ε_i should be independent. Furthermore, because ε_i is distributed as a unit exponential, the mean and variance³ of ε_i should be 1. To examine the intertemporal dependence of the new series, the Ljung-Box statistic for 15 lags was examined. The first sample has a Ljung-Box statistic of 51.56 (remember the critical value is 24.99 at 5%) which is double the critical value at 5%. The second sample has a Ljung-Box statistic of 17.32. This is easily less than the critical value at 5% and the null hypothesis of white noise is not rejected. It is likely that the outliers previously discussed in the first sample have posed a more difficult modeling problem.

First order conditions imply that the mean of ε_i is 1. The standard deviation is 1.35 for the first sample and 1.26 for the second sample. Tests for excess dispersion easily reject the null. It is therefore not likely that this model will provide a good representation of the data generating process.

Two conclusions may be drawn at this point. First, a very simple version of the ACD model is able to capture observed autocorrelations in the data. For the second sample, the Ljung-Box statistic is reduced from levels far in excess of the critical value to levels easily below the critical value at 5%. Second, the large implied unconditional means and large standard deviations in the standardized series suggest that the EACD(1,1) is not able to account for the large

³ Higher order restrictions could easily be examined as well.

observed dispersion. It is possible that a more extensive version of the EACD will. To examine this possibility, an EACD(2,2) was estimated.

Results for the EACD(2,2) are presented in figure 7. All of the included variables are significant at the 5% level for the first sample and all but β_2 are significant for the second sample. The greatly reduced t-ratios are probably due to multicollinearity among the explanatory variables. The sum of the α_i and β_i is 0.9985 for the first data set and .9926 for the second data set. Again, both models are close to being integrated. The implied unconditional mean is 2.1 and 1.11 for the first and second samples respectively. The implied unconditional means remain very large.

The Ljung-Box statistic associated with the two transformed series are 27.32 and 14.42 respectively. The first is very close to the critical value at 5% and the second is easily less than the critical value. The standard deviation of ϵ_i is 1.34 for the first data set and 1.26 for the second. The two added terms have helped account for the intertemporal correlations observed in the first data set. However, the remaining excess dispersion and the large implied unconditional mean indicate poor model specification. We could expand the model further for the first data set in hope of attaining a better fit or we could try an alternative model specification.

While the EACD models are capable of capturing the intertemporal correlations, this model does not look like a good candidate for the data generating process. As mentioned before, the exponential is a very tractable model specification but it is also very inflexible. The single parameter in the distribution yields the well known property of a constant proportional hazard. One might expect that the probability of a trade at time t conditioned on no trade up to time t may not be constant. This is precisely what we find when the WACD model is estimated.

6.2 Estimation of the Weibull ACD (WACD) Model.

The results of the WACD(1,1) for both samples are presented in figure 8. All t-ratios are significant at the 5% level. γ is estimated to be near .8 for both data sets and associated standard errors of 0.00560 and 0.00593. Hence the null hypothesis of a conditional exponential distribution ($\gamma=1$) is easily rejected for both data sets. $\gamma < 1$ implies that the hazard is decreasing in t . Equivalently, the longer the observed duration of no trade, the less likely a trade will occur at that time.

This sum $\alpha+\beta$ is 0.9955 and 0.9895 for the first and second data set respectively. The implied unconditional mean is 1.48 for the first data set, and 1.03 for the second. The effects of outliers remain a modeling problem for the first data set. The unconditional mean for the first data set is almost 1.5 times what it should be if the model were correctly specified while the second sample has an unconditional mean very close to the sample value of 1. We next examine the generalized errors associated with the Weibull model. Examining the standardized series for the Weibull

$$(44) \quad \varepsilon_i = \left(\frac{\bar{x}_i}{\phi_i} \right)^\gamma$$

we find the standard deviations are 1.060 and 1.024. This is a great improvement over the EACD models. In both samples, we still reject the null of no excess dispersion. The associated Ljung-Box statistics are 46.85 and 16.91 . The Ljung-Box statistic for the second sample is far below the critical value at the 5% level.

The Ljung-Box statistic is very small, the implied unconditional mean is acceptably close to 1, and the standard deviations are close to the expected value under the null of a conditional Weibull distribution. This is an improvement over the EACD model specification and this model appears to be a good choice for the second sample. A final model estimated was the WACD(2,2).

The estimates for the WACD(2,2) models are in figure 9. As for the EACD models, the t-ratios have been reduced due to the multicollinearity. The sum of α_i 's and β_j 's is 0.9968 and 0.9809 for the first and second data set respectively. The implied unconditional mean is 1.59 for the first data set, and 1.05 for the second.

The resulting standard deviation of the new ε_i series are 1.06 and 1.02 for the first and second data sets respectively. Although reduced, we still find evidence of excess dispersion for the both samples. The new Ljung-Box statistics are 24.62 and 16.07 . With the exception of the reduced Ljung-Box statistic for the first sample, the WACD(2,2) model offers little improvement over the WACD(1,1).

The EACD is able to capture the intertemporal correlations, but the restrictive constant proportional hazard is not appropriate for this data. A WACD(1,1) is a good choice for the second sample. A high enough order WACD is capable of accounting for the observed autocorrelations for either data set. Although the models reduce the excess dispersion, we still

reject the null of no excess dispersion. A more detailed model is probably necessary to fully account for the dispersion.

7. LM Tests and the Autoregressive Conditional Duration Models.

Two immediate uses of the ACD model are forecasting transactions rates and gaining insight into how markets work and evolve. The former will be examined in the next section. This section will provide some simple examples of the Lagrange Multiplier test for the ACD. The LM test provides a simple framework for testing not only the autoregressive structure of the model but also possible omission of covariates. The LM statistic is easily calculated as $T \cdot R^2$ of the regression of a vector of ones on the numerical evaluations of the score vector evaluated at the null.

From the growing literature surrounding transactions data, one might expect to find correlations between price changes, volume, and no-trade durations. This idea has been examined by, among others, Kyle(1985), Admati and Pfleiderer(1988) and more recently, Easley and O'Hara (1992). In particular these papers theorize that no-trade durations should be correlated with price changes and volume. The nature of the correlation is ambiguous depending on the model examined. These theoretical models, therefore, suggest formulating a test for omission of these variables from our model specification. In addition to these variables, some recent empirical studies in foreign exchange markets have suggested that the difference between the bid and the ask (the spread) would also be an interesting variable to examine. See for example, Madhavan(1992) or Bollerslev and Domowitz(1993). Another interesting test might be of a more general autoregressive specification of ψ_{jt} . A simple alternative would be to introduce squares of lagged durations to the model.

To formulate these tests we first created summary statistics for the state of the market with respect to each of these variables. This was done by taking a moving average of 10 equally weighted lags for each series. The idea is that a single transaction does not carry much information, rather, an average provides a better measure of the current state of volume, volatility, and the percent spread. Ten transactions corresponds to less than 5 minutes of trading on average. Just as the durations exhibited a deterministic component, it is a stylized fact that volume and price changes exhibit a similar deterministic component. We therefore partialled these effects out just as we did for the durations. The tests were performed on the second data set using a WACD(1,1) as the null hypothesis.

The following table contains the LM statistics for the omission of each of the variables individually. Current values of the variables were not used due to the potential endogeneity. $\bar{\sigma}_{10,i}^2$ is the normalized moving average of volatility, $\bar{V}_{10,i}$ is the normalized moving average of volume, $\bar{S}_{10,i}$ is the normalized moving average of the percent difference in ask and bid, and \bar{x}_i^2 is the square of lagged normalized durations. The critical value for a chi-squared statistic with 1 degree of freedom at the 5% level is 3.84. Lagged squared durations and the measures of lagged volatility and volume easily reject the null hypothesis that the coefficients are zero. The lagged spread measure is marginally not rejected..

Variable	$\bar{\sigma}_{10,i-1}^2$	$\bar{V}_{10,i-1}$	$\bar{S}_{10,i-1}$	\bar{x}_{i-1}^2
LM -Statistic	8.61	16.73	3.77	415.35
$\chi_{.05}^2(1) = 3.84$				

The LM tests suggest that shorter durations follow high volatility, large volume, and large spreads. The lagged squared duration is very significant suggesting that the linear specification of ϕ_i might be an oversimplification. These tests provide evidence for correlation between durations and several other measures of the state of the market. This might be indicative of comovement between these market variables. The simple extensions of the ACD model specifications examined here bring to mind many other possible extensions and tests.

8. An Alternative Model Specification to Introduce Quantity Effects

For most of the examples of processes that arrive at irregular intervals there is some notion of a quantity associated with the observation. For credit card purchases, it is the dollars spent in the transaction, for telephone calls it might be the minutes associated with the call. For the stock market transactions data, it is the quantity of shares traded. This section of the paper will propose an alternative model specification that incorporates the idea of volume into the measure of duration.

A straight forward way to introduce volume affects into the ACD model is to redefine our durations. Previously we defined a duration as the time between two successive transactions. We

could alternatively examine the time it takes for a given volume to be traded. The result will be a model of the flow of shares rather than the rate of arrival of traders.

Although the notion of liquidity is not easily defined see Black(1971), it would seem likely that the notion of the rate at which shares are traded is a more meaningful measure than the rate of arrival of traders. Therefore, in equity markets, it is particularly interesting to examine volume based durations.

Another reason we might want to allow for volume affects when examining the financial data sets is to implicitly account for large transactions that have been broken up into several smaller ones executed over a short time. More information might be contained in the flow of shares rather than the arrival of traders.

The average volume traded over both samples is 1775 with a standard deviation of 4118. The minimum number of shares traded was 100 and the maximum was 125,000. We chose 20,000 shares to define a duration. This implied an average of just over 10 trades per duration. Given the maximum volume observed, zero second durations were still present. From the original 30,000 observations, 2764 observations are generated.

The deterministic time of day effects were still present so these effects were partialled out as before. The partial autocorrelation coefficients are presented in figure 10. The intertemporal correlations are still very large. The Ljung-Box statistic has been slightly decreased to 931 but is still very significant. As one might expect, the standard deviation has been reduced substantially by this aggregation. The standard deviation is actually less than one now (.95). The EACD model is marginally rejected in favor of the WACD. For both models, a (2,2) is necessary to account for the large intertemporal autocorrelations. The results for the WACD(2,2) volume based duration model are presented in figure 11. The exponent on the Weibull is now greater than one. This was expected due to the small observed variance. The Ljung-Box statistic of 15.44 has, again, been greatly reduced to well below the critical value.

Large autocorrelations are observed in the volume duration data. The WACD is capable of capturing the intertemporal correlations. The large autocorrelations in the volume data indicates that the observed correlations are not likely due to broken up trades.

9. Forecasting Using the EACD and WACD Models.

This section will discuss forecasting trading rates using the EACD and the WACD model. The ACD models provide a framework for forecasting the time between future events, in our example, it is the time between IBM stock transactions. This section will demonstrate one step and multi step forecasts for the ACD model. As noted in section 1, the one step forecast for the ACD(p,q) takes the form of

$$(45) \quad E(x_{i+1} | \mathcal{F}_{t_i}) = \psi_{i+1} = w + \sum_{j=1}^q \alpha_j x_{i-j} + \sum_{k=1}^p \beta_k \psi_{i-k}$$

Hence ψ_{i+1} is the one step duration forecast at time t_i and it will be denoted by x_i^1 . The k-step forecast for an ACD(p,q) can be evaluated recursively as follows

$$(46) \quad E(x_{i+k} | \mathcal{F}_{t_i}) = w + \sum_{j=1}^q \alpha_j E(x_{i-j} | \mathcal{F}_{t_i}) + \sum_{m=1}^p \beta_m E(\psi_{i-m} | \mathcal{F}_{t_i})$$

In particular, the k step forecast for the ACD(1,1) takes the simple form of

$$(47) \quad x_i^k \equiv E(x_{i+k} | \mathcal{F}_{t_i}) = w \left(\frac{1 - (\alpha + \beta)^{k-1}}{1 - (\alpha + \beta)} \right) + (\alpha + \beta)^{k-1} \psi_{i+1}$$

Substituting in μ simplifies the expression to

$$(48) \quad x_i^k = (1 - (\alpha + \beta)^{k-1}) \mu + (\alpha + \beta)^{k-1} \psi_{i+1}$$

This representation provides some insight to the properties of the k-step forecast. In particular, for the ACD(1,1), the forecast of the expected duration is simply a weighted average of the unconditional implied mean and the one step forecast. When $\alpha + \beta$ is close to one, which

means the process is close to being integrated, the effects of the one step forecast will dominate the forecast longer. When $\alpha+\beta$ is small, the forecasts will quickly return to the unconditional mean. The estimates performed in this paper sum to close to one and therefore k-step forecasts of the durations between trades for this particular asset will be affected by the current conditional mean for many steps ahead.

The implied means are much larger than the observed means for the EACD models. From equation (48) we see that, for the data analyzed, the EACD models will have undesirable forecasting properties. In particular, the forecast will not converge to the unconditional mean as the forecast horizon increases. Forecasts will therefore be presented for the WACD(1,1) using the second data set.

Forecasts of x_{i+k} , require evaluation of the deterministic component. That is,

$$(49) \quad E(x_{i+k} | \mathcal{F}_i) = E(\tilde{x}_{i+k} \Phi(t_{i+k-1}) | \mathcal{F}_i)$$

which, by independence, is

$$(50) \quad E(\tilde{x}_{i+k} | \mathcal{F}_i) E(\Phi(t_{i+k-1}) | \mathcal{F}_i)$$

Hence the one step forecast x_{i+1} is

$$(51) \quad E(\tilde{x}_{i+1} | \mathcal{F}_i) \Phi(t_i).$$

Figure 12 presents the one step forecast of the WACD(1,1) model for an arbitrary day in the second sample. The dotted line is observed durations. The two solid lines are the deterministic spline and the one-step forecast. It is apparent from the graph that the one step forecast of x_{i+k} is substantially more variable than the deterministic part. The lowest forecasts are about half the expected value conditioned on time of day and the highest are near double.

The top of figure 13 presents the multi-step forecasts of \tilde{x}_{i+k} for the WACD model under two scenarios, high and low arrival rates. We will say the market has high activity if the rate of arrival is high (durations are short) and low activity if the rate of arrival is low (durations are

long). The horizontal axis is the k^{th} future transaction and the vertical axis is the expected duration between trades. The upper convex curve represents the low activity multi-step forecast when current durations begin at 1.5 times the unconditional average duration. The lower concave curve represents the high activity multi-step forecast when current durations begin at half the unconditional average duration. They are symmetric about the unconditional mean of 1. Because the sum of the α 's and β 's is near one, the forecasts converge slowly to the unconditional mean.

An alternative forecast is of the expected time until k transactions take place. This is easily calculated by

$$(52) \quad E(t_{i+k}) - t_i = E\left(\sum_{j=1}^k x_{i+j} \middle| \mathcal{F}_{t_i}\right) = \sum_{j=1}^k E(x_{i+j} \middle| \mathcal{F}_{t_i})$$

The bottom of figure 13 presents these forecasts for the same two "high" "low" scenarios. These forecasts are analogous to the example in the bottom of figure 1. We see that the forecasts generated by the ACD are, as expected, non-linear reflecting the dependence in durations. These forecasts will become parallel as the forecast horizon increases. The slope will converge to the unconditional mean as observed in equation (48). The bottom of figure 13 forecasts the expected time for a set number of events n . The top of figure 13 measures the rate of transactions.

10. Conclusions/Discussion

This paper has introduced a model for data that does not arrive at regular time intervals. Rather than relying on an exogenous variables to drive time deformation or assumptions on the causes of time deformation, time flow is modeled directly via an autoregressive process. The model is easily generalized to include other exogenous variables and for the financial transactions data used in this paper. LM tests suggest the inclusion of other measures of the state of the market.

For the financial transaction data used in this paper, it was shown that very simple model formulations were successful in reducing excessively large Ljung-Box statistics of well over 1000 in the observed no-trade durations to values less than the critical value of 24.99 for the standardized series. While the exponential version of the ACD model is capable of accounting for the intertemporal correlations, the large implied means and the large standardized standard

deviations imply that the conditional exponential assumption may not be a good one. A simple extension from the conditional exponential distribution of durations to the Weibull distribution, further improves the Ljung-Box statistic for the standardized durations and reduces the standard deviations to values very close to those implied under correct model specification. The additional parameter of the Weibull is very significant and changes the constant hazard implied by the exponential assumption to a decreasing hazard. The Weibull distribution also reduces the implied unconditional mean to values much closer to the observed mean. The WACD appears to be a good model choice for this data. In particular a relatively simple specification of WACD(1,1) does a good job of modeling for the second sample.

LM tests proved to be simple and valuable for the data analyzed. The LM tests for omitted measures of lagged volume and volatility easily rejected the null hypothesis. The percent difference between bid and ask was marginally not rejected. The rejection of the null hypothesis for lagged squared durations suggested that more complex models might be considered. The tests indicated that shorter durations follow periods of high volatility, large volume per transaction, and larger spreads. These tests suggest the possibility of comovement between durations, volatility, volume, and spread.

An alternative use of the data allowed us to associate a measure of quantity with the durations. For the financial data set we used the time for a set number of shares to be traded. This provided the interpretation of modeling the rate that shares are traded rather than the rate of arrival of traders. For financial transaction data, this might provide a better way of measuring and forecasting liquidity. The amount of time necessary for 20,000 shares to take place still exhibits large intertemporal autocorrelations that were successfully modeled by a WACD(2,2).

Forecasts were developed for expected durations. The one-step forecasts of the conditional mean exhibited significant volatility. The multi-step forecasts of duration did not return to the unconditional mean very rapidly. A period of sustained high (low) activity will, in expectation, be followed by sustained high (low) activity.

Clearly, there are numerous generalizations of the model that come to mind. The LM tests indicated that there are several other variables that could have entered our model. If spreads, prices, volume, and durations are moving together then a natural generalization of the model would jointly model all these variables. Perhaps this is the most interesting extension of the model.

The linearity hypothesis in the specification of $\psi(\cdot)$ is also suspect. While linearity is a convenient property, the models could easily be misspecified. We will in general have abundant data when working with these transactions data hence a natural extension of the model would be to consider a semi-parametric estimation procedure. While these methods are very common in the duration literature, we do not know of any applications to stochastically time varying hazards.

Figure 1. Counting Process

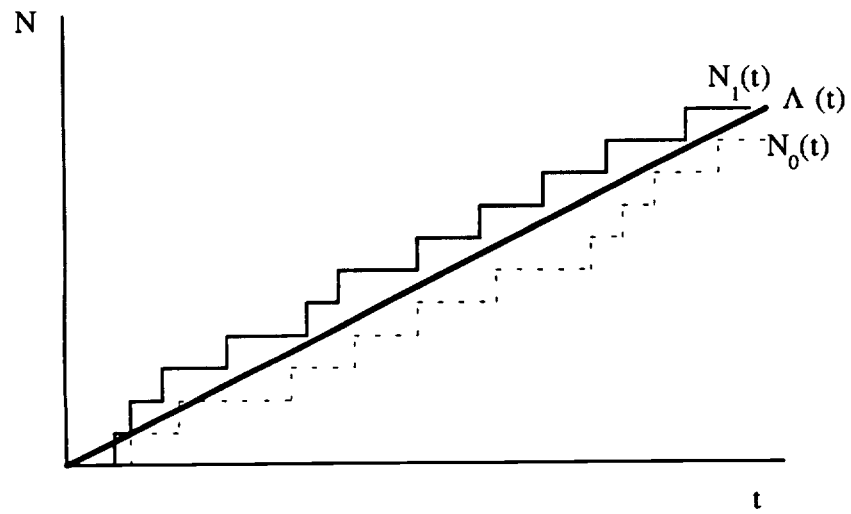
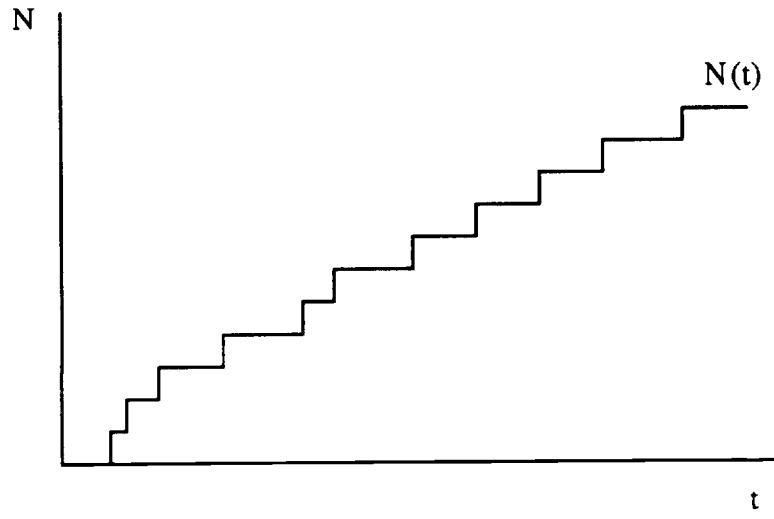


Figure 2 Duration Histograms for Each Sample

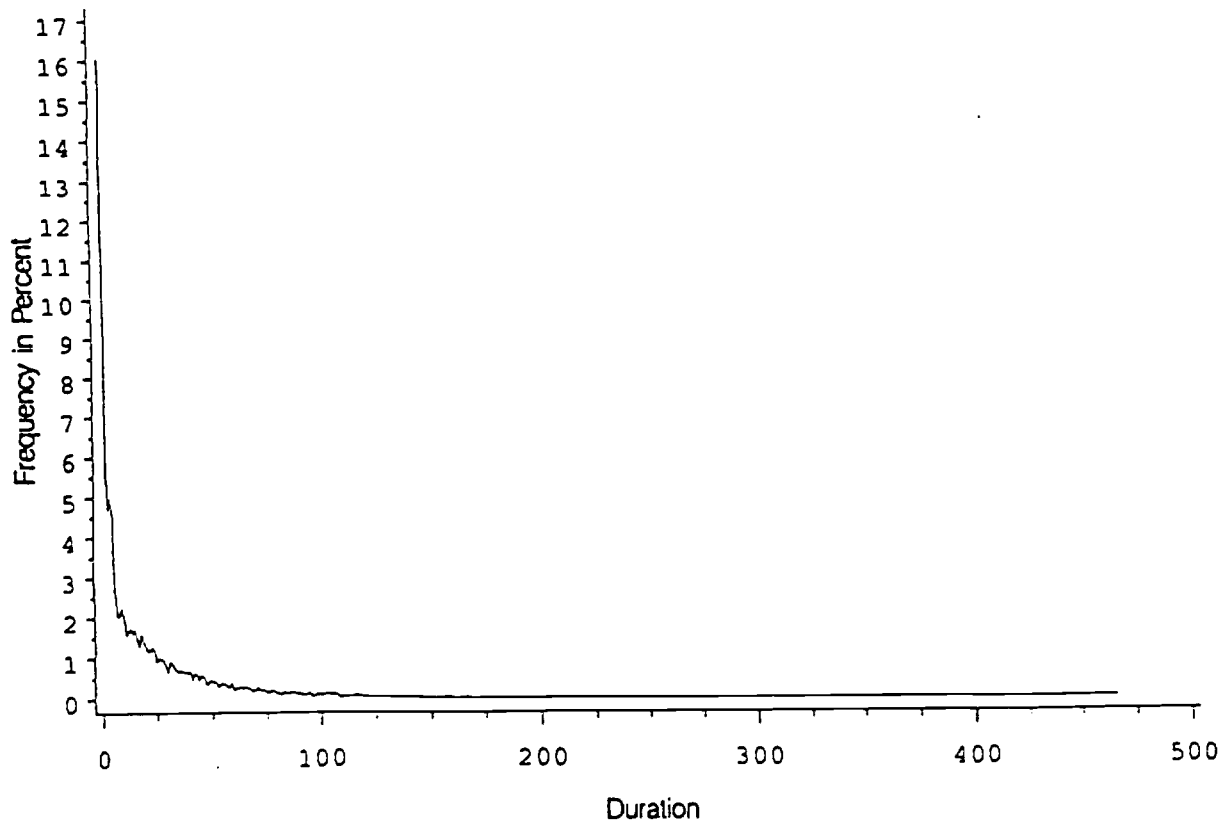
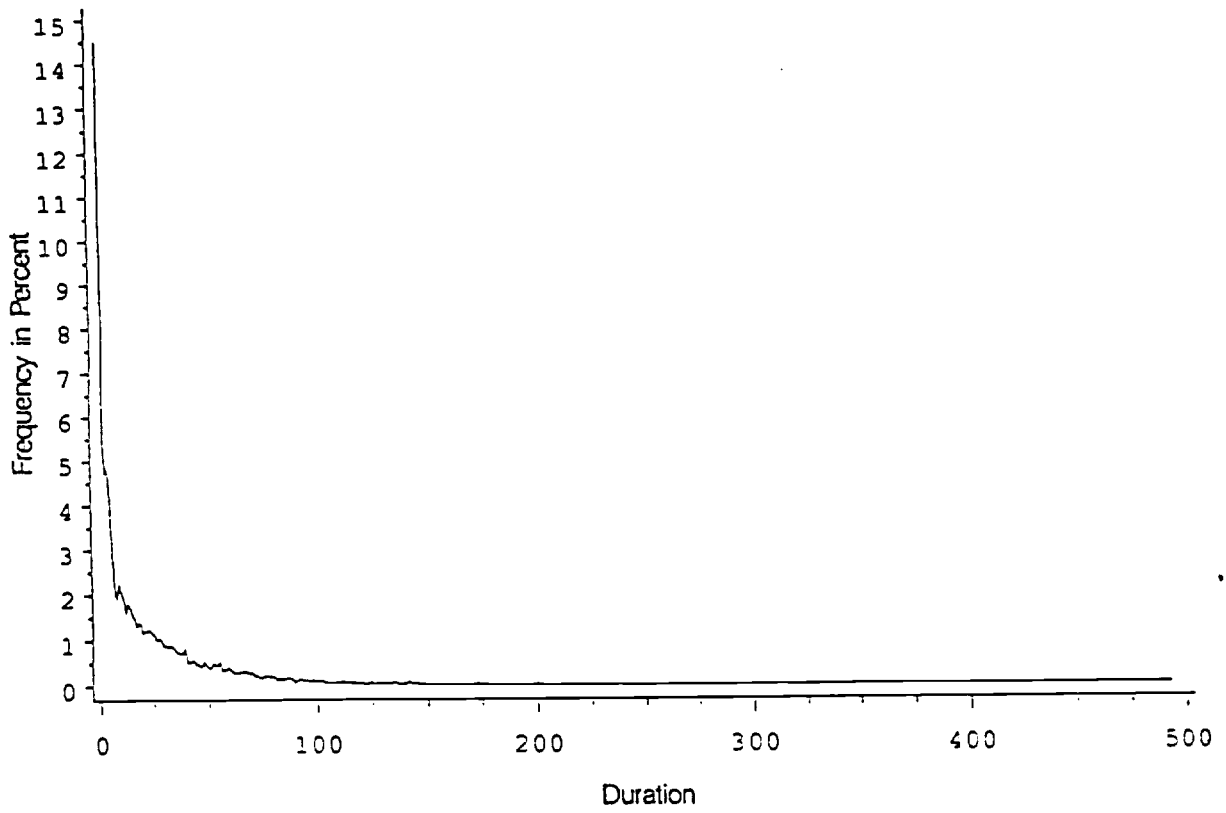


Figure 3. Autocorrelations and Partial Autocorrelations of Trading Intervals

	Sample 1		Sample 2	
	acf	pacf	acf	pacf
lag 1	.168	.168	.129	.129
lag 2	.090	.064	.120	.106
lag 3	.068	.044	.106	.081
lag 4	.074	.053	.119	.089
lag 5	.059	.034	.107	.068
lag 6	.069	.046	.096	.053
lag 7	.051	.023	.100	.055
lag 8	.046	.021	.099	.051
lag 9	.035	.011	.123	.074
lag 10	.042	.021	.085	.029
lag 11	.043	.022	.105	.051
lag 12	.045	.022	.087	.029
lag 13	.047	.029	.089	.030
lag 14	.037	.013	.089	.031
lag 15	.025	.003	.083	.024

Ljung-Box(15) = 1271.62

Ljung-Box(15) = 2423.12

Figure 4. Mean Duration by Hour

Sample 1

Hour	Freq.	Mean	Std.
9:30-10	1905	15.57	20.39
10-11	2769	22.02	32.09
11-12	2360	29.03	101.0
12-1	1928	33.48	44.79
1-2	1962	33.06	52.16
2-3	2350	27.62	37.76
3-4	2612	24.91	34.12

Sample 2

Hour	Freq.	Mean	Std.
9:30-10	1957	15.32	21.50
10-11	2365	27.22	34.66
11-12	2238	27.38	37.33
12-1	1895	32.47	42.53
1-2	1640	37.46	52.67
2-3	2127	29.12	38.09
3-4	2778	23.58	32.01

Figure 5. Mean Conditioned on Time of Day

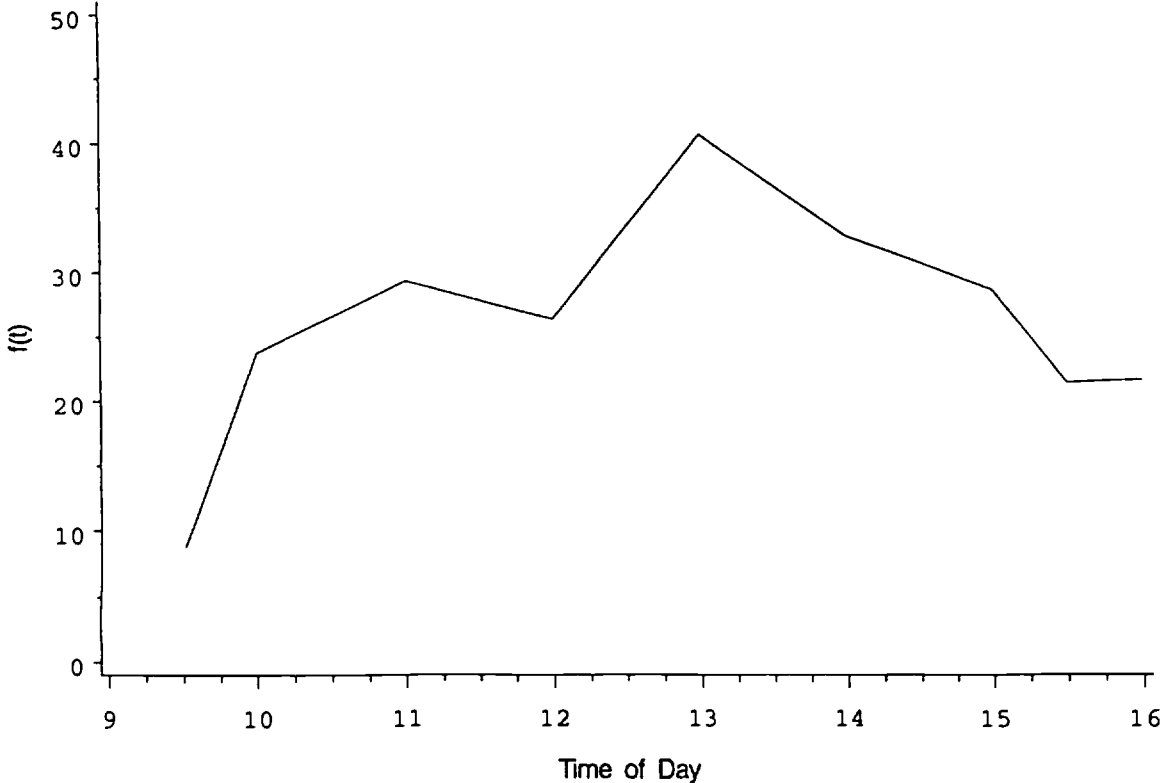
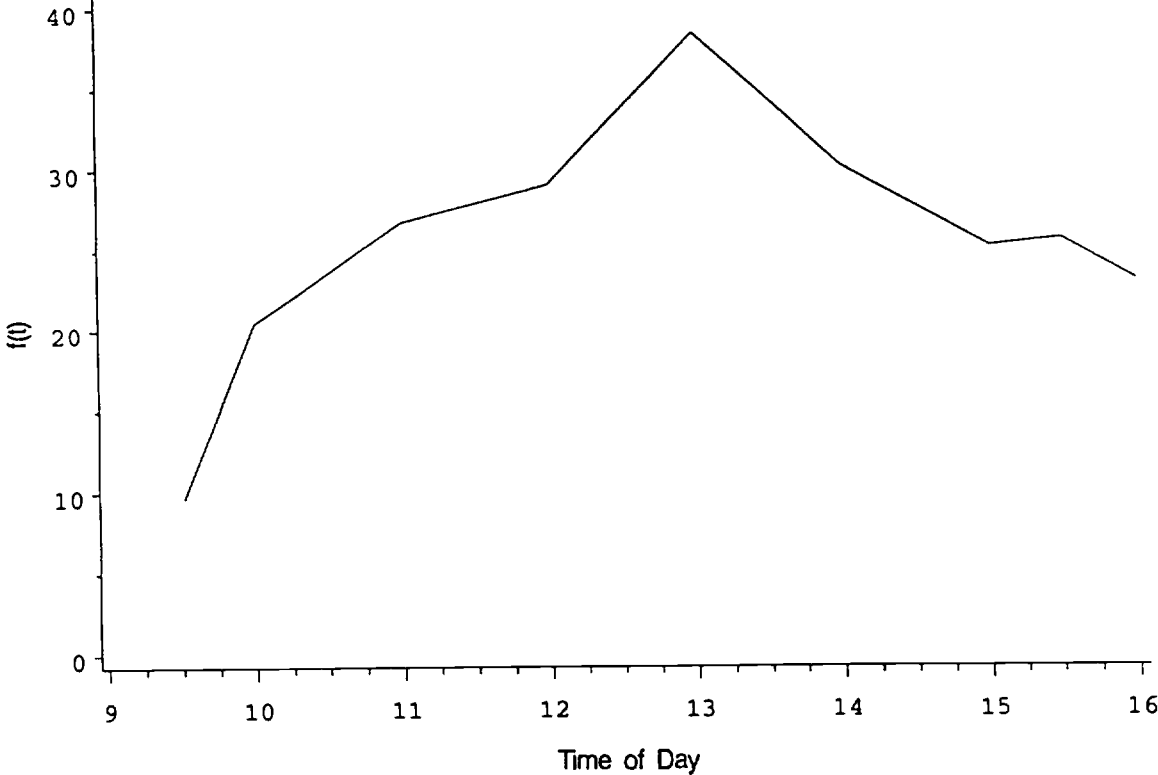


Figure 6. Maximum Likelihood Estimates of EACD(1,1) Model after Removing Time of Day Effects

Sample 1

	Coefficient	Std. Error	t-Ratio
ω	.00713	.000687	10.36
α_1	.0816	.00219	37.159
β_1	.9143	.00230	396.48
new day dummy	-.1140	.0239	-4.75

Sample 2

	Coefficient	Std. Error	t-Ratio
ω	.0116	.001175	9.91
α_1	.0584	.002399	24.38
β_1	.9311	.002934	317.34
new day dummy	-.2860	.04599	-6.21

Figure 7. Maximum Likelihood Estimates of EACD(2,2) Model After Removing Time of Day Effects.

Sample 1

	Coefficient	Std. Error	t-Ratio
ω	.00315	.000601	5.22
α_1	.130	.00613	21.32
α_2	-.0854	.00694	-12.30
β_1	1.218	.0847	14.38
β_2	-.265	.0784	-3.38
new day dummy	-.0453	.0174	-2.60

Sample 2

	Coefficient	Std. Error	t-Ratio
ω	.00827	.00258	3.20
α_1	.0779	.00634	12.28
α_2	-.0329	.0137	-2.40
β_1	1.06	.233	4.55
β_2	-.115	.218	-.52
new day dummy	-.207	.070	-2.93

Figure 8. Maximum Likelihood Estimates of WACD(1,1) Model After Removing Time of Day Effects.

Sample 1

	Coefficient	Std. Error	t-Ratio
ω	.00693	.00096	6.34
α_1	.08129	.00317	22.60
β_1	.9143	.00373	244.92
γ	.8019	.00560	-35.36*
new day dummy	-.1942	.00374	-4.58

Sample 2

	Coefficient	Std. Error	t-Ratio
ω	.0113	.00156	6.40
α_1	.0578	.00317	16.20
β_1	.9317	.00438	212.62
γ	.8110	.00593	-31.86*
new day dummy	-.3031	.06099	-4.57

* $H_0: \gamma=1$

Figure 9. Maximum Likelihood Estimates of WACD(2,2) Model After Removing Time of Day Effects.

Sample 1			
	Coefficient	Std. Error	t-Ratio
ω	.005406	.001176	4.59
α_1	.13006	.008753	14.85
α_2	-.05839	.01287	-4.53
β_1	.90987	.13769	6.61
β_2	-.01536	.12609	-.12
γ	.8030	.005672	-34.73*
new day dummy	-.2265	.04796	-4.72

Sample 2			
	Coefficient	Std. Error	t-Ratio
ω	.02057	.003112	5.89
α_1	.05889	.005302	9.90
α_2	-.04498	.008566	4.68
β_1	.11072	.12506	.88
β_2	.76632	.116636	6.57
γ	.8113	.005931	-31.81*
new day dummy	-.3728	.112123	-3.32

* $H_0: \gamma=1$

Figure 10. Autocorrelations and Partial Autocorrelations of Volume Durations

	Both Samples	
	acf	pacf
lag 1	.282	.282
lag 2	.192	.123
lag 3	.182	.110
lag 4	.151	.066
lag 5	.133	.051
lag 6	.164	.090
lag 7	.121	.026
lag 8	.137	.058
lag 9	.127	.037
lag 10	.115	.029
lag 11	.107	.023
lag 12	.091	.007
lag 13	.143	.079
lag 14	.113	.018
lag 15	.103	.018

Ljung-Box(15) = 931.29

Figure 11. Estimated Coefficients for Volume Based Durations

	Coefficient	Std. Error	t-Ratio
ω	.01070	.00542	2.001
α_1	.2913	.03537	8.350
α_2	-.2291	.03423	-6.784
β_1	1.1195	.13056	8.574
β_2	-.1938	.109918	-1.763
γ	1.025	.01173	2.131*
new day dummy	.1476	.07172	2.088

* $H_0: \gamma=1$

Figure 12. Observed Durations, the Deterministic Spline, and the One-Step Forecasts for the WACD(1,1) Model.

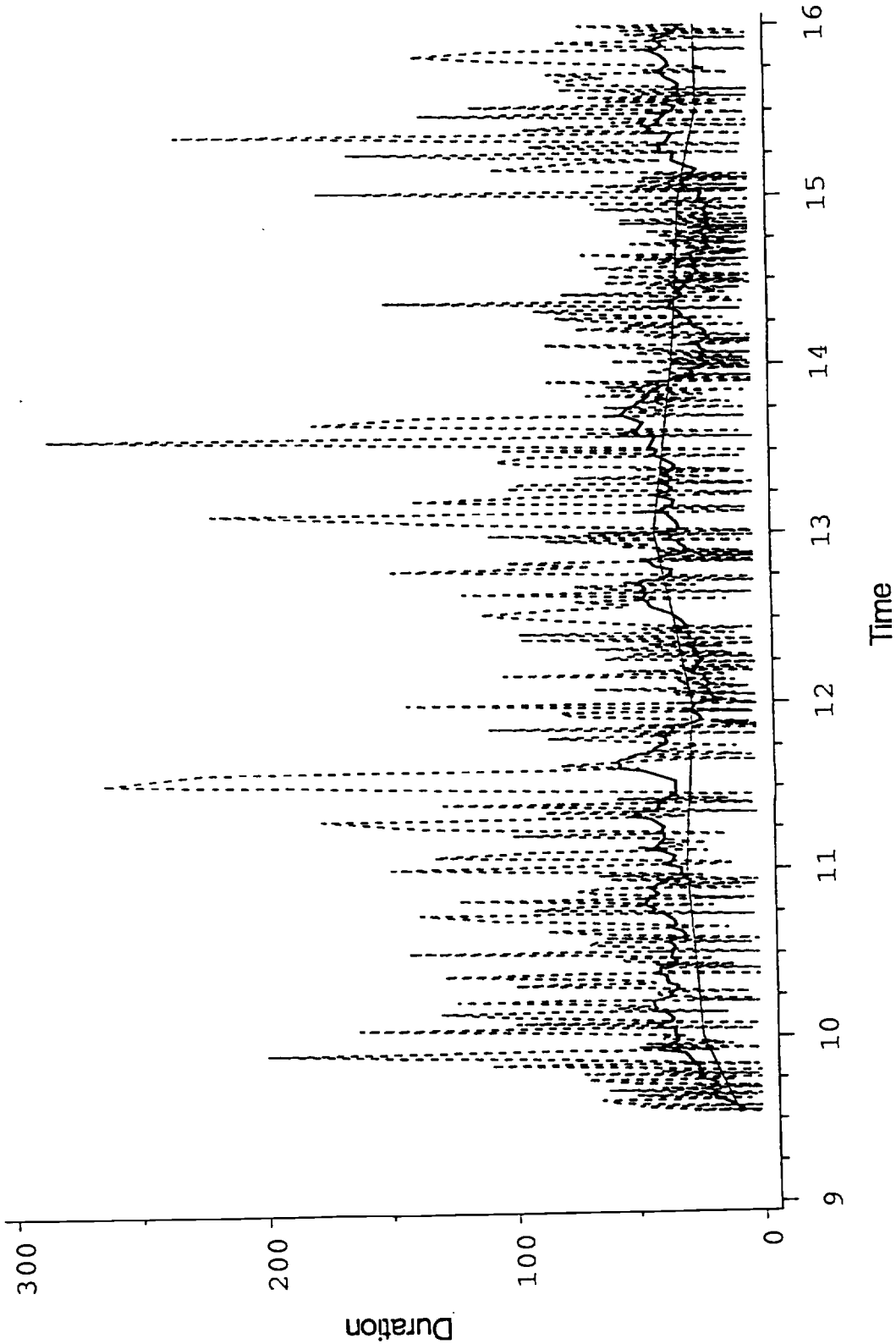
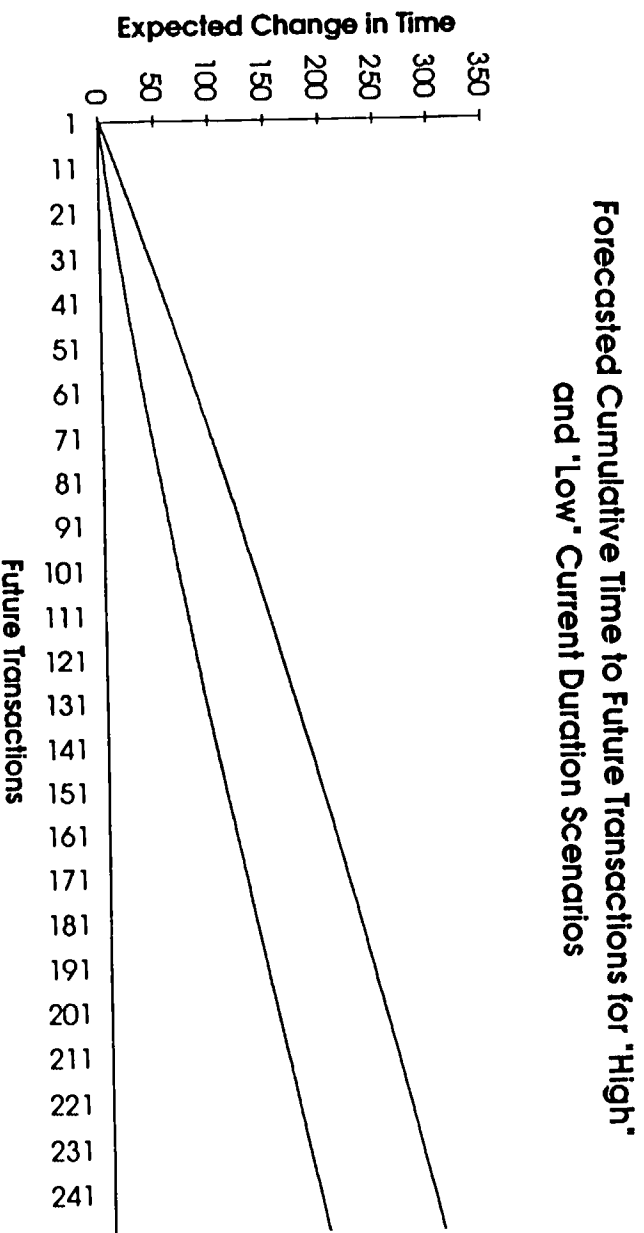
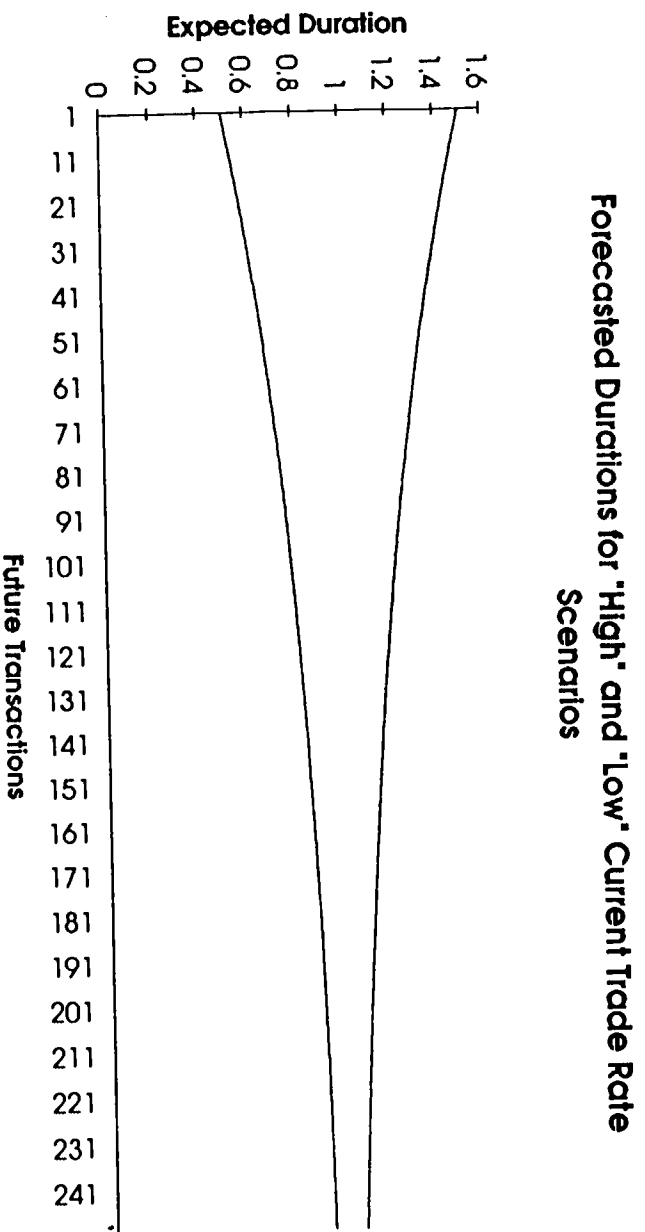


Figure 13. Multi-Step Forecasting Using the WACD(1,1) Model.



Appendix

Proof of Lemma 1

Without loss of generality let $p=q$.

Define $z_j = E(x_{i+j} | \mathcal{F}_i)$. Then taking expectations of z_i w.r.t. \mathcal{F}_{i-p} gives

$$E(x_i | \mathcal{F}_{i-p-1}) = \omega + \alpha_1 E(x_{i-1} | \mathcal{F}_{i-p-1}) + \dots + \alpha_p E(x_{i-p} | \mathcal{F}_{i-p-1}) + \beta_1 E(\psi_{i-1} | \mathcal{F}_{i-p-1}) + \dots + \beta_p E(\psi_{i-p} | \mathcal{F}_{i-p-1})$$

$$z_{p+1} = \omega + \alpha_1 z_p + \dots + \alpha_p z_1 + \beta_1 z_p + \dots + \beta_p z_1$$

$$= \omega + \sum_{i=1}^p (\alpha_i + \beta_i) z_{p-i+1}$$

So, for $j > p$,

$$z_j = \omega + \sum_{i=1}^p (\alpha_i + \beta_i) z_{j-i}$$

If all the roots of this difference equation lie outside the unit circle then

$$\lim_{j \rightarrow \infty} z_j = \frac{\omega}{1 - \sum_{i=1}^p (\alpha_i + \beta_i)}$$

QED

The condition that $\sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) < 1$ is necessary for all the roots to be outside the unit circle.

Proof of Lemma 2

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

$$\psi_i - \mu = \alpha(x_{i-1} - \mu) + \beta(\psi_{i-1} - \mu)$$

$$= \alpha(x_{i-1} - \psi_{i-1}) + (\alpha + \beta)(\psi_{i-1} - \mu)$$

$$E(\psi_i - \mu)^2 = \alpha^2 E(x_{i-1} - \psi_{i-1})^2 + (\alpha + \beta)^2 E(\psi_{i-1} - \mu)^2$$

$$\begin{aligned}
&= \alpha^2 \kappa E(\psi_{i-1}^2) + (\alpha + \beta)^2 E(\psi_{i-1} - \mu)^2 \\
&= \alpha^2 \kappa E(\psi_{i-1}^2 - \mu) + \alpha^2 \kappa \mu + (\alpha + \beta)^2 E(\psi_{i-1} - \mu)^2 \\
&= \alpha^2 \kappa \mu^2 + ((\alpha + \beta)^2 + \alpha^2 \kappa) E(\psi_{i-1} - \mu)^2
\end{aligned}$$

$$E(\psi_i - \mu)^2 = \frac{\alpha^2 \kappa \mu^2}{1 - (\alpha + \beta)^2 - \alpha^2 \kappa}$$

Now,

$$\begin{aligned}
E(x_i - \mu)^2 &= E(x_i - \psi_i)^2 + E(\psi_i - \mu)^2 \\
&= \kappa E\psi_i^2 + E(\psi_i - \mu)^2 \\
&= \kappa E[(\psi_i - \mu)^2 + \mu^2] + E(\psi_i - \mu)^2 \\
&= \kappa \mu^2 + (1 + \kappa) E(\psi_i - \mu)^2 \\
&= \kappa \mu^2 + (1 + \kappa) \frac{\alpha^2 \kappa \mu^2}{1 - (\alpha + \beta)^2 - \alpha^2 \kappa} \\
&= \kappa \mu^2 \left(1 + \frac{(1 + \kappa) \alpha^2}{1 - (\alpha + \beta)^2 - \alpha^2 \kappa} \right) \\
&= \kappa \mu^2 \left(\frac{1 - (2\alpha\beta + \beta^2)}{1 - (\alpha + \beta)^2 - \alpha^2 \kappa} \right)
\end{aligned}$$

QED

Bibliography

1. Admati, Anat R. and Paul Pfleiderer, 1988, A theory of Intraday Patterns: Volume and Price Variability, *The Review of Financial Studies* 1, 3-40.
2. Bera, Anil and Mathew Higgins, 1992, Survey of ARCH Models: Properties, Estimation and Testing, unpublished manuscript, Department of Economics, University of Illinois
3. Black, Fischer, 1971, Towards a Fully Automated Exchange, Part I, *Financial Analysts Journal*, 27, 29-34
4. Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327
5. Bollerslev, Tim and Ian Domowitz, 1993, Trading Patterns and Prices in the Interbank Foreign Exchange Market, *The Journal of Finance*, 48, 1421-1443
6. Bollerslev, Tim, Robert Engle, and Dan Nelson, 1994, ARCH models, in *Handbook of Econometrics IV*, (Robert Engle and Dan McFadden eds.), North Holland
7. Breusch, T.S. and A.R., Pagan 1978, A simple test for heteroskedasticity and random coefficient variation, *Econometrica* 47, 1287-1294
8. Cox, D. R., 1972, Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B*, 34, 187-220
9. Cox, D. R., 1972a, The Statistical Analysis of Dependencies in Point Processes, in *Symposium on Point Processes*, (P.A.W. Lewis, ed.), New York, John Wiley
10. Clark, Peter, 1973, A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica*, 41, 135-156
11. Easley and O'Hara, 1992, Time and the Process of Security Price Adjustment. *The Journal of Finance* 19, 69-90
12. Engle, Robert, 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, 50, 987-1008
13. Engle, Robert and Gloria Gonzalez-Rivera, 1991, Semiparametric ARCH Models, *Journal of Business and Economic Statistics*, 9, 345-359
14. Ghysels, Eric and J. Jasiak 1994, Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects, C.R.D.E., Universite de Montreal, unpublished manuscript

15. Glosten, Lawrence R., Milgrom P., 1985, Bid Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Agents, *Journal of Financial Economics* 14 71-100
16. Godfrey, L.J., 1978, Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables, *Econometrica* 46, 1293-1302
17. Godfrey, L.J., 1988, Misspecification Tests in Econometrics The Lagrange Multiplier Principle and Other Approaches, Cambridge University Press, Cambridge
18. Grandell, J., 1976, Doubly Stochastic Poisson Processes (Lecture Notes Math529), Springer Verlag, New York
19. Hasbrouck, J. Measuring the Information Content of Stock Trades, *The Journal of Finance* 66,1, 179-207
20. Hausman, Lo, and MacKinlay, An Ordered Probit Analysis of Transaction Stock Prices, 1992, An Ordered Probit Analysis of Transaction Stock Prices, *Journal of Financial Economics* 31, 319-379
21. Heckman, J. J., and Borjas, B. J., 1980, Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence. *Economica* 47: 247-283
22. Jain, P. J., and Joh, G., 1986, The Dependence Between Hourly Prices and Trading Volume, working paper, University of Pennsylvania, Wharton School.
23. Kyle, Albert, 1985, Continuous Time Auctions and Insider Trading, *Econometrica* 53, 1315-1336
24. Lancaster, Tony, 1990, *The Econometric Analysis of Transition Data*, Cambridge University Press.
25. Lamoureux, C. G., and W. D. Lastrapes, 1990, Heteroskedasticity in Stock Return Data: Volume Versus GARCH effects, *Journal of Finance* 45, 221-229.
26. Muller U.A., M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, C. Morgenegg, 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
27. Stock James, 1988, Estimating Continuous time Processes Subject to Time Deformation, *Journal of the American Statistical Association*, 83, 77-85

28. Tauchen, George and Mark Pitts, 1993, The Price Variability-Volume Relationship on Speculative Markets, *Econometrica*, 51, 485-505
29. Todorovic, Petar, 1991, *An Introduction to Stochastic Processes and Their Applications*, Springer-Verlag, New York