Analysis of High Frequency Financial Data

Robert F. Engle
New York University and University of California, San Diego

Jeffrey R. Russell
University of Chicago, Graduate School of Business

December 21, 2004

Contents

1 Introduction

From a passing airplane one can see the rush hour traffic snaking home far below. For some, it is enough to know that the residents will all get home at some point. Alternatively, from a tall building in the center of the city one can observe individuals in transit from work to home. Why one road is moving more quickly than another can be observed. Roads near the coastal waters might be immersed in a thick blanket of fog forcing the cars to travel slowly due to poor visibility while roads in the highlands, above the fog, move quickly. Traffic slows as it gets funneled through a narrow pass while other roads with alternate routes make good time. If a critical bridge is washed out by rain then some travelers may not make it home at all that night.

Like the view from the airplane above, classic asset pricing research assumes only that prices eventually reach their equilibrium value, the route taken and speed of achieving equilibrium is not specified. How does the price actually adjust from one level to another? How long will it take? Will the equilibrium be reached at all? How do market characteristics such as transparency, the ability of traders to view others actions, or the presence of several markets trading the same asset affect the answers to these questions? Market microstructure studies the mechanism by which prices adjust to reflect new information.

Answers to these questions require studying the details of price adjustment. From the passing plane in the sky, the resolution is insufficient and the view from the market floor, like the view from the street below the building, provides a very good description of the actions of some individuals, but lacks perspective. With high frequency financial data we stand atop the tall building, poised to empirically address such questions.

1.1 Data Characteristics

With these new data sets come new challenges associated with their analysis. Modern data sets may contain tens of thousands of transactions or posted quotes in a single day time stamped to the nearest second. The analysis of these data are complicated by irregular temproal spacing, diurnal patterns, price discretness, and complex often very long lived dependence.

1.1.1 Irregular Temporal Spacing

Perhaps most important is that virtually all transactions data are inherently irregularly spaced in time. Figure 1 plots one two hours of transaction prices for an arbitrary day in March 2001. The stock used is the US stock Airgas which will be the subject of several examples throughout the paper. The horizontal axis is the time of day and the vertical axis is the price. Each diamond denotes a transaction. The irregular spacing of the data is immediately evident as some transactions appear to occur only seconds apart while others, for example between 10:30 and 11:00 may be five or ten minutes apart.

Since most econometric models are specified for fixed intervals this poses an immediate complication. A choice must be made regarding the time intervals over which to analyze the data. If fixed intervals are chosen then some sort of interpolation rule must be used when no transaction occurs exactly at the end of the interval. Alternatively if stochastic intervals are used then the spacing of the data will likely need to be taken into account. The irregular spacing of the data becomes even more complex when dealing with multiple series each with its own transaction rate. Here, interpolation can introduce spurious correlations due to non-syncronous trading.

1.1.2 Discreteness

All economic data is discrete. When viewed over long time horizons the variance of the process is usually quite large relative to the magnitude of the minimum movement. For transaction by transaction data, however, this is not the case and for many data sets the transaction price changes take only a handful of values called ticks. Institutional rules restrict prices to fall on a pre-specified set of values. Price changes must fall on multiples of the smalles allowable price change called a tick. In a market for an actively traded stock it is generally not common for the price to move a large number of ticks from one transaction to another. In open outcry markets the small price changes are indirectly imposed by discouraging the specialist from making radical price changes from one transaction to the next and for other markets, such as the Taiwan stock exchange these price restrictions are directly imposed in the form of price change limits from one

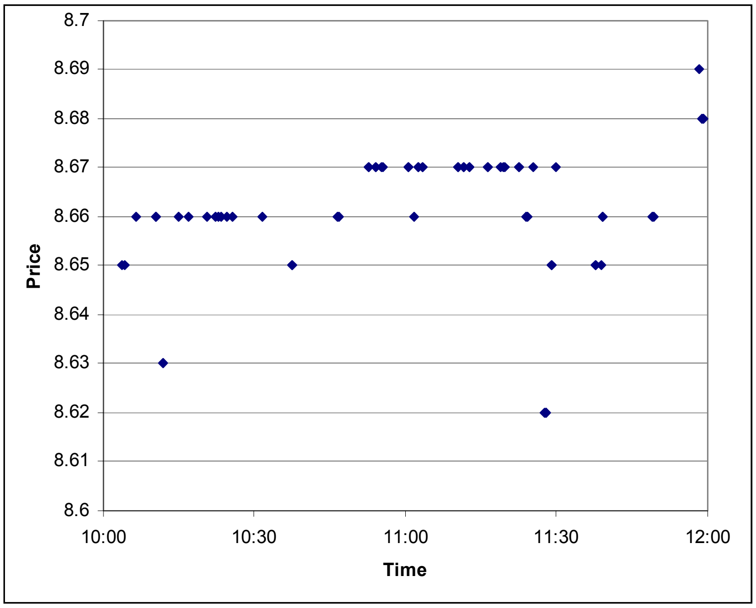

Figure 1: Plot of a small sample of transaction prices for the Airgas stock

transaction to the next (say 2 ticks). The result is that price chages often fall on a very small number of possible outcomes.

US stocks have recently undergone a transition from trading in 1/8ths of a dollar to decimalization. This transition was initially tested for 7 NYSE stocks in August of 2000 and was completed for the NYSE listed stocks on January 29th, 2001. NASDAQ began testing with 14 stocks on March 12, 2001 and completed the transition April 9, 2001. In June of 97 NYSE permitted 1/16th prices.

As an example, Figure 2 presents a histogram of the Airgas data transaction price changes after deleting the overnight and opening transactions. The sample used here contains 10 months of data spanning from March 1, 2001 through December 31, 2001. The horizontal axis is measured in cents. 52% of the transaction prices are unchanged from the previous price. Over 70% of the transaction prices fall on one of three values; no change, up one cent or down one cent. Over 90% of the values lie between -5 and +5 cents. Since the bid and ask prices are also restricted to the same minimum adjustment the bid, ask, and the midpoint of the bid ask prices will exhibit similar

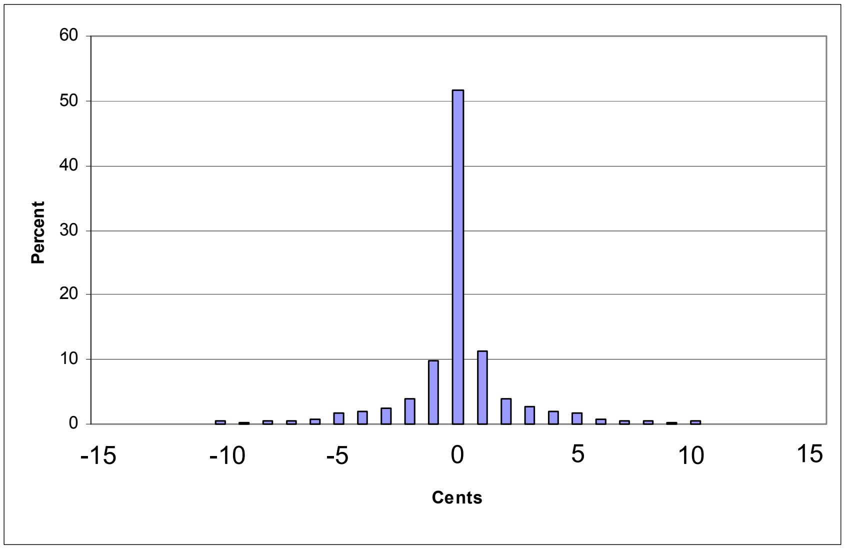

Figure 2: Histogram of price transaction changes for Airgas stock

discreteness. Of course data prior to decimalization is even more extreme. For these data sets it is not uncommon to find over 98% of the data taking just one of 5 values. This discreteness will have an impact on measuring volatility, dependence, or any characteristic of prices that is small relative to the tick size.

This discreteness also induces a high degree of kurtosis in the data. For example, for the Airgas data the sample kurtosis is 66. Such large kurtosis is typical of high frequency data.

1.1.3 Diurnal Patterns

Intraday financial data typically contain very strong diurnal or periodic patterns. For most stock markets volatility, the frequency of trades, volume, and spreads all typically exhibit a U-shaped pattern over the course of the day. For an ealry reference see McInish and Wood (1992). Volatility is systematically higher near the open and generally just prior to the close. Volume and spreads have a similar pattern. The time between trades, or durations, tend to be shortest near the open and just prior to the close. This was first documented in Engle and Russell (1998).

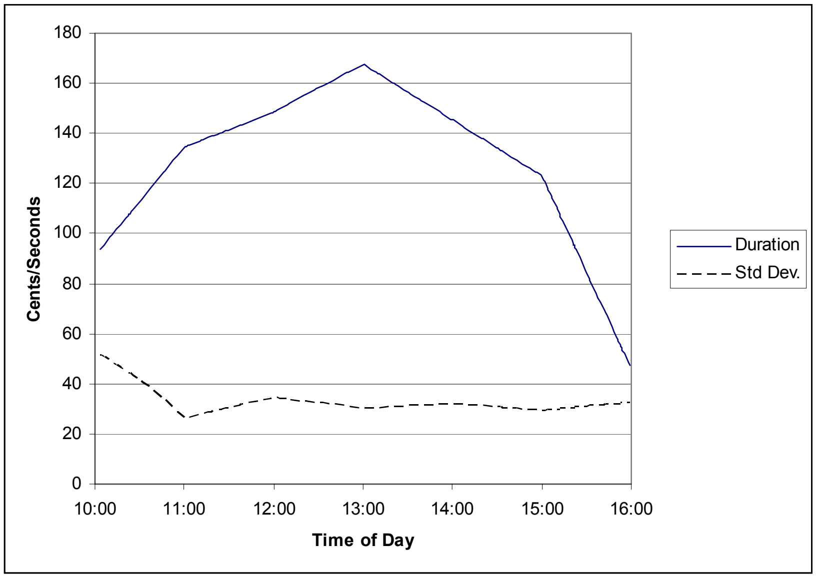

Figure 3: Diurnal pattern for durations and standard deviation of mid-quote price changes.

Figure 3 presents the diurnal patterns estimated for the ARG data. The diurnal patterns were estimated by fitting a piecewise linear spline to the duration between trades and the squared midquote price changes. The vertical axis is measured in seconds for the duration and cents for the standard deviation of price changes.

Diurnal patterns are also typically present in the foreign exchange market although here there is no opening and closing of the market. These markets operate 24 hours a day seven days a week. Here the pattern is typically driven by "active" periods of the day. See Andersen and Bollerslev (1997) for patterns in foreign exchange volaitlity. For example, prior to the induction of the Euro, US dollar exchange rates with European countries typically exhibit the highest volatility during the overlap of time that both the US markets and the European markets were active. This occurred in the late afternoon GMT when it is morning in the US and late afternoon in Europe.

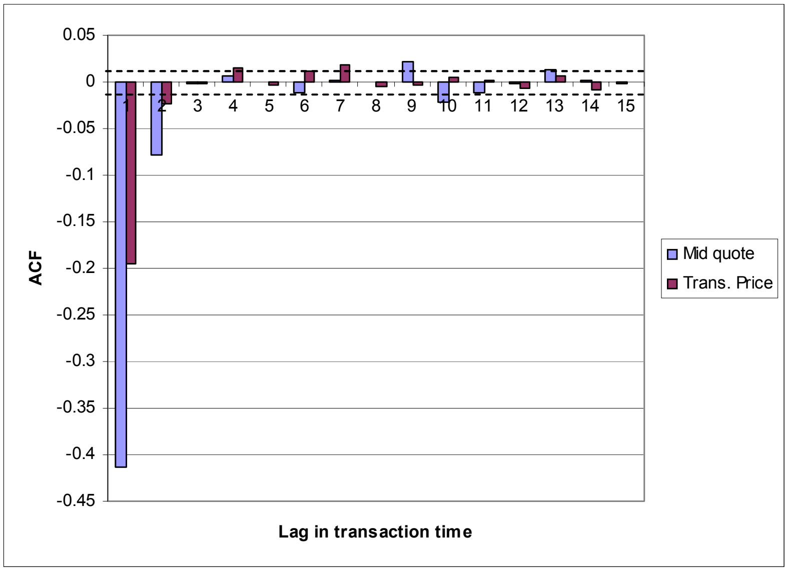

Figure 4: Autocorrelation function for the mid quote and transaction price changes

1.1.4 Temporal Dependence

Unlike their lower frequency counterparts, high frequency financial returns data typically display strong dependence. The dependence is largely the result of price discreteness and the fact that there is often a spread between the price paid by buyer and seller initiated trades. This is typically referred to as bid-ask bounce and is responsible for the large first order negative autocorrelation. Bid ask bounce will be discussed in more detail in section 2.3. Other factors leading to dependence in price changes include traders breaking large orders up into a sequence of smaller orders in hopes of transacting at a better price overall. These sequences of buys or sells can lead to a sequence of transactions that move the price in the same direction. Hence at longer horizons we sometimes find positive autocorrelations. Figure 4 contains a plot of the autocorrelation function for changes in the transaction and midpoint prices from one transaction to the next for the Airgass stock using the 10 months of data. Again, overnight price changes have been deleted.

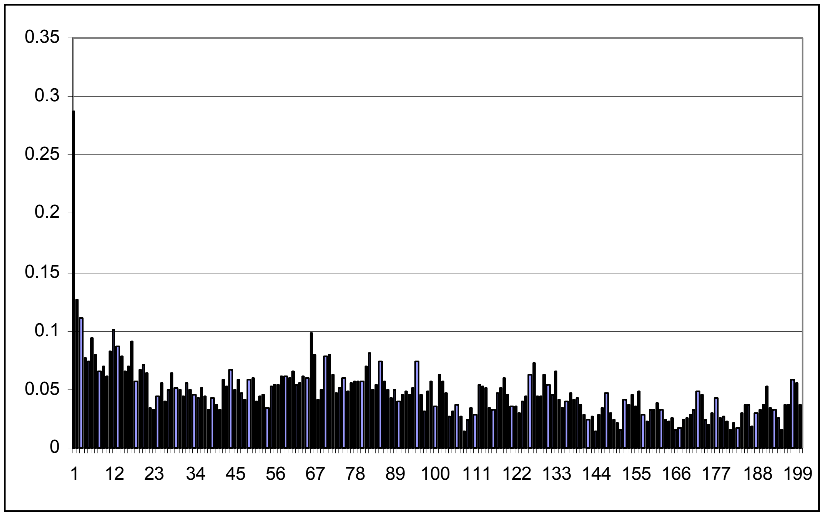

Figure 5: Autocorrelations for the squared midquote price changes

Similar to lower frequency returns, high frequency data tends to exhibit volatility clustering. Large price changes tend to follow large price changes and vice-versa. The ACF for the absolute value of the transaction price change for ARG is shown in figure 5. Since the diurnal pattern will likely influence the autocorrelation function it is first removed by dividing the price change by the square root of its variance by time of day. The variance by time of day was estimated with linear splines. The usual long set of positive autocorrelations is present.

The transaction rates also exhibit strong temporal dependence. Figure 6 presents a plot of the autocorrelations for the durations between trades after removing the deterministic component discussed above. Figure 7 presents the autocorrelations for the log of volume. Both series exhibit long sets of positive autocorrelation spanning many transactions. These autocorrelations indicate clustering of durations and volume respectively.

Under temporal aggregation the dependence in the price changes tends to decrease. However, even at intervals of a half hour or longer negative first order autocorrelation often remains.

1.2 Types of economic data

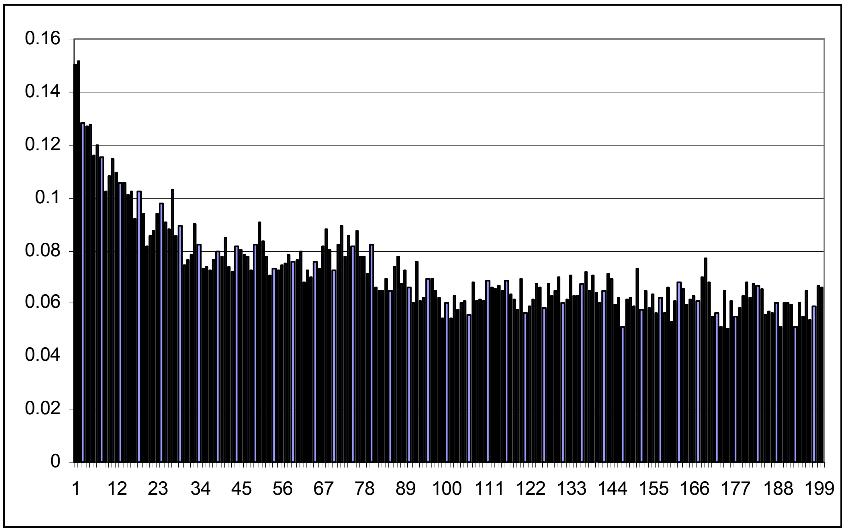

Figure 6: Autocorrelations for durations

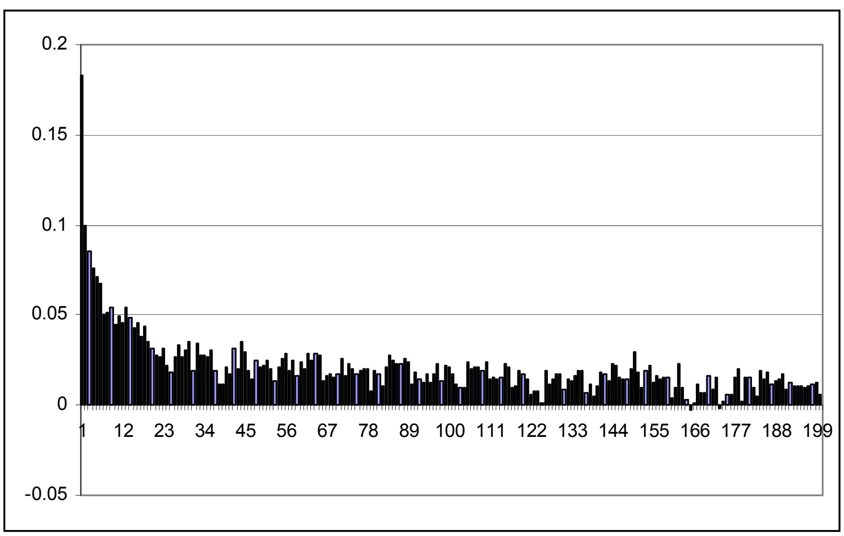

Figure 7: Autocorrelations for the log of volume

We begin this section with a general discussion of the types of high frequency data currently available. With the advancement and integration of computers in financial markets data sets containing detailed information about market transactions are now commonplace. Many financial assets are traded in centralized markets which potentially contain the most detailed information. The NYSE and the Paris Bourse are two such markets commonly analyzed. These centralized markets might be order driven (like the Paris Bourse) where a computer uses algorithms to match market participants or they might be open outcry (like the NYSE) where there is a centralized trading floor that market participants must trade through. In either case, the data sets constructed from these markets typically contain detailed information about the transactions and quotes for each asset traded on the market. The exact time of a transaction usually down to the second, and the price and quantity transacted are common data. Similarly, the bid and ask quotes are also generally available along with a time stamp for when the quotes became active. The trades and quotes (TAQ) data set distributed by the NYSE is an example of such a data set.

The currency exchange market is a commonly analyzed decentralized market. Here the market participants are banks that communicate and arrange transactions on a one on one basis with no central recording institution. Quotes are typically fed through reuters for customers with a reuters account to view continuous updates of quotes. Olsen and Associates has been downloading and storing this quote data and made it available for academic research. The most comprehensive data sets contain all quotes that pass through the Reuters screens and an associated time stamp. As the transactions do not pass through a centralized system there is no comprehensive source for transaction data.

The definition of the quote can vary across markets with important implications. Foreign exchange quotes are not binding. Quotes from the NYSE are valid for a fixed (typically small) quantity or depth. Quotes in electronic markets can come in various forms. For example, quotes for the Paris bourse are derived from the limit order book and represent the best ask and bid prices in the book. The depth here is determined by the quantity of volume at the best prices. Alternatively, the Taiwan stock exchange, which is an electronic batch auction market, posts a "reference price" derived from the past transaction and is only a benchmark from which to gauge where the next transaction price may fall. These difference are important not only from an economic perspective but also determine the reliability of the data. Non-binding quotes are much more likely to contain large errors than binding ones.

Many empirical studies have focused on the stock and currency exchange high frequency data. However, data also exists for other markets, most notably the options and futures markets. These data sets treat each contract as a separate asset reporting time quotes and transactions just as for the stocks. The Berkeley Options data base is a common source for options data.

Other specialized data sets are available that contain much more detailed information. Perhaps the most well known data set is the TORQ data set put together by Joel Hasbrouck and the NYSE. This data set is not very comprehensive in that it contains only 144 stocks traded on US markets covering three months in the early 1990's. However, it contains detailed information regarding the nature of the transactions including the order type (limit order, market order etc.) as well as detailed information about the submission of orders. The limit order information provides a widow into the limit order book of the specialist although it cannot be exactly replicated. The Paris Bourse data typically contain detailed information about the limit order book near the current price.

1.3 Economic Questions

Market microstructure economics focuses on how prices adjust to new information and how the trading mechanism affects asset prices. In a perfect world, new information would be immediately disseminated and interpreted by all market participants. In this full information setting prices would immediately adjust to a new equilibrium value determined by the agents preferences and the content of the information. This setting, however, is not likely to hold in practice. Not all relevant information is known by all market participants at the same time. Furthermore, information that becomes available is not processed at the same speed by all market participants implying variable lag time between a news announcement and the agents realization of price implications. Much of modern microstructure theory is therefore driven by models of asymmetric information.

In the simplest form, there is a subset of the agents that are endowed with superior knowledge regarding the value of an asset. These agents are referred to as privately informed or simply informed agents. Agents without superior information are referred to as noise or liquidity traders and are assumed to be indistinguishable from the informed agents. Questions regarding the means by which the asset price transitions to reflect the information of the privately informed agents can be couched in this context. Early theoretical papers utilizing this framework include Glosten and Milgrom (1985), Easley and O'Hara (1992), Copland and Galai (1983) and Kyle (1985). A very comprehensive review of this literature can be found in O'Hara (1995).

The premise of these models is that market makers optimally update bid and ask prices to reflect all public information and remaining uncertainty. For the NYSE it is the specialist that plays the role of market maker. Even in markets without a designated specialist bid and ask quotes are generally inferred either explicitly or implicitly from the buy and sell limit orders closest to the current price. Informed and uniformed traders are assumed to be indistinguishable when arriving to trade so the difference between the bid and the ask prices can be viewed as compensation for the risk associated with trading against potentially better informed agents. Informed traders will make profitable transactions at the expense of the uninformed.

In a rational expectations setting market makers learn about private information by observing the actions of traders. Informed traders only transact when they have private information and would like to trade larger quantities to capitalize on their information before it becomes public. The practical implications is that characteristics of transactions carry information. An overview of the predictions of these models is that prices adjust more quickly to reflect private information when the proportion of uninformed traders is higher. Volume is higher, transaction rates are higher when the proportion of uninformed traders is higher. The bid ask spread is therefore predicted to be increasing in volume and transaction rates.

It is unlikely that aggregated daily data will be very useful in empirically evaluating market microstructure effects on asset prices. Using high frequency data we seek to empirically asses the effects of market microstructure on asset price dynamics. Although private information is inherently unobservable the impact of transaction characteristics on subsequent price revisions can be measured and quantified. We can asses which trades are likely to move the price or what characteristics of the market should we expect to see large rapid price changes. A related and very practical issue is liquidity loosely defined as the ability to trade large volume with little or no price impact. How much will it cost to transact and how is this cost related to transaction characteristics? An empirical understanding of the answers to these questions might lead to optimal order submission strategies.

Answers to these questions necessarily involve studying the relationship between transaction prices and market characteristics. The bid ask spread is often considered a measure of liquidity since it measures the cost of purchasing and then immediately reselling. The wider the spread the higher the cost of transacting. Of course spreads vary both cross sectionally across different stocks as well as temporally. Using high frequency data the variation in the spread can be linked to the variation in market characteristics.

The price impact of a trade can be measured by relating quote revisions to characteristics of a trade. Again with high frequency transactions data the magnitude and possibly the direction of price adjustments can be linked to characteristics of the market. Volatility models for high frequency data are clearly relevant here.

A privately informed agent may have the ability to exploit that information in multiple markets. In some cases a single asset is traded in multiple markets. A natural question is in which market does price discovery take place? Do price movements in one market precede price movements in the other market. If the two series are cointegrated these questions can be addressed in the context of error correction models. Another possibly for informed agents to have multiple outlets to exploit their information is the derivative market. From an empirical perspective this can be reduced to causality questions. Do price movements in one market precede price movements in the other? Do trades in one market tend to have a greater price impact across both markets?

Implementing the econometric analysis, however, is complicated by the data features discussed in the previous section. This chapter provides a review of the techniques and issues encountered in the analysis of high frequency data. The chapter is organized as follows.

2 Econometric Framework

One of the most salient features of high frequency transactions data is that transactions do not occur at regularly spaced time intervals. In the statistics literature this type of process has been referred to as a point process and a large body of literature has been produced studying and applying models of point processes. Examples of applications include the study of firing of neurons or the study of earthquake occurrences. More formally, let $t_1, t_2, ..., t_i, ...$ denote a sequence of strictly increasing random variables corresponding to event arrival times such as transactions. Jointly, these arrival times are referred to as a point process. It is convenient to introduce the counting function $N(t)$ which is simply the number of event arrivals that have occurred at or prior to time t. This will be a step function with unit increments at each arrival time.

Often, there will be additional information associated with the arrival times. In the study of earthquake occurrences there might be additional information about the magnitude of the earthquake associated with each arrival time. Similarly, for the financial transactions data there is often a plethora of information associated with the transaction arrival times including price, volume, bid and ask quotes, depth, and more. If there is additional information associated with the arrival times then the process is referred to as a marked point process. Hence, if the marks associated with the i^{th} arrival time are denoted by an M-dimensional vector y_i then the information associated with the i^{th} event is summarized by its arrival time and the value of the marks $[t_i, y_i]$.

Depending on the economic question at hand, either the arrival time, the marks, or both may be of interest. Often economic hypothesis can be couched in the framework of conditional expectations of future values. We denote the filtration of arrival times and marks at the time of the i^{th} event arrival by $\widehat{t}_i = \{t_i, t_{i-1}, ..., t_0\}$ and $\widehat{y}_i = \{y_i, y_{i-1}, ..., y_0\}$ respectively. The probability structure for the dynamics associated with a stationary, marked point process is can be completely characterized and conveniently expressed as the joint distribution of marks and arrival times given the filtration of past arrival times and marks:

$$f\left(t_{N(t)+1}, y_{N(t)+1} | \widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) \tag{1}$$

While this distribution provides a complete description of the dynamics of a marked point process it is rarely specified in practice. Often the question of economic interest can be expressed in one of four ways. When will the next event happen? What value should we expect for the mark at the next arrival time? What value should we expect for the mark after a fixed time interval? Or, how long should we expect to wait for a particular type of event to occur?

The answers to the first two questions are immediately obtained from (1). Alternatively, if the contemporaneous relationship between y_i and t_i is not of interest, then the analysis may be greatly simplified by restricting focus to the marginalized distributions provided that the marks and arrival times are weakly exogneous. If the waiting time until the next event regardless of the value of the marks at termination is of interest, then the marginalized distribution given by

$$f_t\left(t_{N(t)+1}|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) = \int f\left(t_{N(t)+1}, y|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) dy \tag{2}$$

may be analyze. This is simply a point process where the arrival times may depend on the past arrival times and the past marks. We will refer to this as a model for the event arrival times, or simply a point process. Examples here include models for the arrival of traders.

Alternatively, many economic hypotheses involve the dynamics of the marks such as models for the spread, or prices. In this case, one may be interested in modeling or forecasting the value for the next mark, regardless of when it occurs, given the filtration of the joint process. This is given by

$$f_y\left(y_{N(t)+1}|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) = \int f\left(t, y_{N(t)+1}|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) dt \tag{3}$$

Here, the information set is updated at each event arrival time and we refer to such models as event time or tick time models of the marks. Of course, multiple step forecasts from 2 would require, in general, a model for the marks and multiple step forecasts for the mark in 3 would generally require a model for the durations.

Yet another alternative approach is to model the value of the mark to be at some future time $t + \tau$ ($\tau > 0$) given the filtration at time t. That is

$$g\left(y_{N(t+\tau)}|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) \tag{4}$$

Here the conditional distribution associated with the mark over a fixed time interval is the object of interest. Theoretically, specification of (1) implies a distribution for (4), only in very special cases,however, will this exist in closed form. Since the distribution of the mark is specified over discrete fixed calendar time intervals we refer to this type of analysis as fixed interval analysis. A final approach taken in the literature is to study the distribution of the length of time it will take for a particular type of event, defined by the mark, to occur. For example, one might want to know how long will it take for the price to move by more than a specified amount, or how long will it take for a set amount of volume to be transacted. This can be expressed as

$$g\left(t + \tau_{\min}|\widehat{t}_{N(t)}, \widehat{y}_{N(t)}\right) \tag{5}$$

where if E_t defines some event associated with the marks, then $\tau_{\min} = \min_{\tau>0} y_{N(t+\tau)} \in E_t$. Again, only in special cases can this distribution be derived analytically from (1). $t_{\min}$ is called the hitting time in the stochastic process literature. Since the marks are associated with arrival times, the first crossing times will simply be a subset of the original set of arrival times. In the point process literature the subset of points is called a thinned point process.

This section proceeds to discuss each of the above approaches. We begin with a discussion and examples of point processes. Next, we consider tick time models. We then consider fixed interval analysis by first discussing methods of converting to fixed time intervals and then give examples of various approaches used in the literature.

2.1 Examples of Point Processes

It is convenient to begin this section with a discussion of point processes with no marks. A point process is referred to as a *simple point process* if, as a time interval goes to zero, the probability of multiple events occurring over that time interval can be made an arbitrarily small fraction of the probability of a single event occurring. In this case, characterization of the instantaneous probability of a single event dictates the global behavior of the process. A convenient way of characterizing a simple point process, therefore is by the instantaneous arrival rate of the intensity function given by:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr\left(N(t+\Delta t > N(t))\right)}{\Delta t} \tag{6}$$

Perhaps the most well known simple point process is the homogenous Poisson process. For a homogeneous Poisson process the probability of an event arrival is constant. A homogenous Poisson process can therefore be described by a single parameter where $\lambda(t) = \lambda$. For many types of point process the assumption of a constant arrival rate is not likely realistic. Indeed, for financial data we tend to observe bursts of trading activity followed by lulls. This feature becomes apparent when looking at the series of the time between transactions, or durations. Figure 6 presents the autocorrelations associated with the intertrade durations of Airgas. The plot indicates strong temporal dependence in the durations between transaction events. Clearly the homogenous Poisson model is not suitable for such data.

For a point process with no marks, Snyder and Miller (1991) conveniently classify point processes into two categories, those that evolve with *after-effects* and those that do not. A point process on $[t_0, \infty]$ is said to evolve without after effects if for any $t > t_0$ the realization of events on $[t, \infty)$ does not depend on the sequence of events in the interval $[t_0, t)$. A point process is said to be *conditionally orderly* at time $t \geq t_0$ if for a sufficiently short interval of time and conditional on any event P defined by the realization of the process on $[t_0, t)$ the probability of two or more events occurring is infinitessimal relative to the probability of one event occurring. Our discussion here focuses on point processes that evolve with after-effects and are conditionally orderly A point process that evolves with after-effects can be conveniently described using the *conditional intensity function* which specifies the instantaneous probability of an event arrival conditional upon filtration of event arrival times. That is, the conditional intensity is given by

$$\lambda(t|N(t), t_{i-1}, t_{i-2}, ..., t_0) = \lim_{\Delta t \to 0} \frac{P\left(N(t+\Delta t) > N(t)|N(t), t_{N(t)}, t_{N(t)-1}, ..., t_0\right)}{\Delta t} \tag{7}$$

The conditional intensity function associated with any single waiting time has traditionally been called a hazard function in the econometrics literature. Here, however, the intensity function is defined as a function of t across multiple events, unlike much of the literature in macroeconomics that tends to focus on large cross sections with a single spell.

Perhaps the simplest example of a point process that evolves with aftereffects is a first order homogeneous point process where $\lambda\left(t|N(t), t_{N(t)-1}, t_{N(t)-2}, ..., t_0\right) = \lambda\left(t|N(t), t_{N(t)}\right)$ and the durations between events $x_i = t_i - t_{i-1}$form a sequence of independent random variables. If in addition, the durations are identically distributed then the process is referred to as a renewal process. More generally, for an m^{th} order self exciting point process the conditional intensity depends on $N(t)$ and the m most recent event arrivals.

As discussed in Snyder and Miller (1975), for example, the conditional intensity function, the conditional survivor function, and the durations or "waiting times" between events each completely describe a conditionally orderly point process. Letting p_i be a family of conditional probability density functions for arrival time t_i, the log likelihood can be expressed in terms of the conditional density or intensity as

$$L = \sum_{i=1}^{N(T)} \log p_i(t_i|t_0, ..., t_{i-1}) \tag{8}$$

$$L = \sum_{i=1}^{N(T)} \log \lambda(t_i|t_0, ..., t_{i-1}) - \int_{t_0}^{T} \lambda\left(u|N(t), t_0, ..., t_{i-1}\right) \tag{9}$$

Equation (**??**) is referred to as a self exciting point process. It was originally proposed by Hawkes (1971) and by Rubin (1972) and are sometimes called Hawkes self exciting processes. Numerous parameterizations have been proposed in the statistics literature.

2.1.1 The ACD model

Engle and Russell [**?**] propose the Autoregressive Conditional Duration (ACD) which is particularly well suited for high frequency financial data. This parameterization is most easily expressed in terms of the waiting times between events. Let $x_i = t_i - t_{i-1}$ be the interval of time between event arrivals which will be called the duration. The distribution of the duration is specified directly conditional on the past durations. The ACD model is then specified by two conditions. Let ψ_i be the expectation of the duration given the past arrival times which is given by

$$E\left(x_i|x_{i-1}, x_{i-2}, ..., x_1\right) = \psi_i\left(x_{i-1}, x_{i-2}, ..., x_1;\right) = \psi_i \tag{10}$$

Furthermore, let

$$x_i = \psi_i \varepsilon_i \tag{11}$$

where ε_i˜i.i.d. with density $p(\epsilon; \phi)$ with non-negative support, and θ and ϕ are variation free. The baseline intensity, or baseline hazard, is given by

$$\lambda_0 = \frac{p(\epsilon; \phi)}{S(\epsilon; \phi)} \tag{12}$$

where $S_0(\epsilon; \phi) = \int_{\epsilon}^{\infty} p(u; \phi) du$ is the survivor function. The intensity function for an ACD model is then given by

$$\lambda(t|N(t), t_{i-1}, t_{i-2}, ..., t_0) = \lambda_0 \left(\frac{t - t_{N(t)-1}}{\psi_{N(t)}}\right) \frac{1}{\psi_{N(t)}} \tag{13}$$

Since ψ_i enters the baseline hazard this type of model is referred to as an accelerated failure time model in the duration literature. The rate at which time progresses through the hazard function is dependent upon ψ_i and therefore can be viewed in the context of time deformation models. During some periods the pace of the market is more rapid than other periods.

The flexibility of the ACD model stems from the variety of choices for parameterizations of the conditional mean in (10) and the i.i.d. density $p(\epsilon;\phi)$. Engle and Russell (1998) suggest and apply linear parameterizations for the expectation given by

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{14}$$

Since the conditional expectation of the duration depends on p lags of the duration and q lags of the expected duration this is termed an ACD(p,q) model. Popular choices for the density $p(\epsilon;\phi)$ include the exponential and the Weibull distributions suggested in Engle and Russell (1998). These models are termed the Exponential ACD (EACD) and Weibull ACD (WACD) models respectively. The exponential distribution has the property that the baseline hazard is monotonic. The Weibull distribution relaxes this assumption and allows for a hump-shaped baseline intensity. An appropriate choice of the distribution, and hence the baseline intensity will depend on the characteristics of the data at hand. Other choices include the Gamma distribution Lunde (1998) and Zhang Russell and Tsay (2001) or the Burr distribution suggested in Grammig and Maurer (2000). These distributions allow for even greater flexibility in the baseline hazard. Given a choice for (10) and $p(\epsilon;\phi)$ the likelihood function is constructed from (8).

For each choice of $p(\epsilon;\phi)$ from (12) and (13) there is an implied intensity function. Since the exponential distribution implies a constant hazard, the intensity function takes a particularly simple form given by

$$\lambda(t|N(t), t_{i-1}, t_{i-2}, ..., t_0) = \frac{1}{\psi_{N(t)}} \tag{15}$$

and for the Weibull distribution the intensity is slightly more complicated

$$\lambda(t|N(t), t_{i-1}, t_{i-2}, ..., t_0) = \gamma \left(\frac{\Gamma\left(1+\frac{1}{\gamma}\right)}{\psi_{N(t)}} \right)^{\gamma} \left(t - t_{N(t)}\right)^{\gamma-1} \tag{16}$$

which reduces to (15) when $\gamma = 1$.

The ACD(p,q) specification in (14) appears very similar to a ARCH(p,q) models of Engle (1982) and Bollerslev (1986) and indeed the two models share many of the same properties. From (11) and (14) it follows that the durations x_i follow an ARMA$(max(p,q),q)$. Let $\eta_i \equiv x_i - \psi_i$ which is a martingale difference by construction then

$$x_i = \omega + \sum_{j=1}^{\max(p,q)} \left(\alpha_j + \beta_j\right) x_{i-j} - \sum_{j=1}^{q} \beta_j \eta_{i-j} + \eta_i$$

If $\alpha(L)$ and $\beta(L)$ denote polynomials in the lag operator of orders p and q respectively then the persistence of the model can be measured by $\alpha(1) + \beta(1)$. For most duration data this sum is very close to (but less than) one indicating strong persistence but stationarity. It also becomes clear from this representation that restrictions must be placed on parameter values to ensure non-negative durations. These restrictions impose that the infinite AR representation implied by inverting the MA component must contain non-negative coefficients for all lags. These conditions are identical to the conditions derived in Nelson and Cao (1992) to ensure non-negativity of GARCH models. For example, for the ACD(1,1) model this reduces to $\omega \geq 0$, $\alpha \geq 0$, $\beta \geq 0$. Similarly,

The most basic application of the ACD model to financial transactions data is to model the arrival times of trades. In this case it denotes the arrival of the i^{th} transaction and x_i denotes the time between the i^{th} and (i-1)th transactions. Engle and Russell (1998) propose using an ACD(2,2) model with Weibull errors to model the arrival times of IBM transactions. Like volatility, the arrival rate of transactions on the NYSE can have a strong diurnal (intraday) pattern. Volatility tends to be relatively high just after the open and just prior to the close; that is, they have volatility for stocks tends to exhibit a U shaped diurnal pattern. Similarly, Engle and Russell (1998) document that the durations between trades have a diurnal pattern with high activity just after the open and just prior to the close; that is, the durations exhibit an inverse U shaped diurnal pattern. Let $\phi_{N(t)+1} = E\left(x_{N(t)+1} | t_{N(t)}\right)$ denote the expectation of the duration given time of day alone. Engle and Russell (1998) suggest including an additional term on the right hand side of (11) to account for a diurnal pattern so that the i^{th} duration is given by:

$$x_i = \phi_i \psi_i \varepsilon_i \tag{17}$$

Now, ψ_i is the expectation of the duration after partialing out the deterministic pattern and is interpreted as the fraction above or below the average value for that time of day. The expected (non-standardized) duration is now given by $\phi_i \psi_i$. It is natural to refer to ϕ_i as the deterministic component and ψ_i as the stochastic component. Engle and Russell (1998) suggest using cubic splines to model the deterministic pattern.

The parameters of the two components as well as any parameters associated with ε_i can be estimated jointly by maximizing (8) or, a two step procedure can be implemented in which first the terms of the deterministic pattern are estimated and in a second stage the remaining parameters are estimated. The two step procedure can be implemented by first running an OLS regression of durations on a cubic spline. Let $\widehat{\phi}_i$ denote the prediction for the i^{th} duration obtained from the OLS regression. Then let $\widetilde{x}_i = \frac{x_i}{\phi_i}$ denote the normalized duration. This standardized series should be free of any diurnal pattern and should have a mean near unity. An ACD model can then be estimated by MLE using the normalized durations $\widetilde{x}_i$ in place of x_i in (8). While this is not efficient the two step procedure will provide consistent estimates under correct specification. (3) plots the estimated diurnal pattern for ARG. This plot was constructed by regressing the duration on a linear spline for the time of day at the start of the duration. We find the typical inverted U shaped pattern with durations longest in the middle of the day and shortest near the open and close.

The similarity between the ACD model . Indeed the link is close as detailed in the following corollary proven in Engle and Russell (1998).

Corollary 1 *QMLE results for the EACD(1,1) model*

If

1. $E_{i-1}(x_i) = \psi_i = \omega + \alpha x_{i-1} + \beta \; \psi_{i-1}$,
2. $\epsilon_i = \frac{x_i}{\psi_i}$ is
   (a) strictly stationary
   (b) nondegenerate
   (c) has bounded conditional second moments
   (d) $\sup_i E_{i-1}[\ln(\beta + \alpha \epsilon_i)] < 0$
3. $\theta_0 \equiv (\omega, \alpha, \beta)$ is in the interior of Θ

4. $L(\theta) = -\sum_{i=1}^{N(T)} \left(\log(\psi_i) + \frac{x_i}{\psi_i} \right)$

Then the maximizer of L will be consistent and asymptotically normal with a covariance matrix given by the familiar robust standard errors from Lee-Hansen (1994).

This result is a direct corollary from the Lee and Hansen (1994 and Lumsdaine (1996) proofs for the class of GARCH(1,1) models. The theorem is powerful since under the conditions of the theorem we can estimate an ACD model assuming an exponential distribution and even if the assumption is false we still obtain consistent estimates although the standard errors need to be adjusted as in White (1982). Furthermore, the corollary establishes that we can use standard GARCH software to perform QML estimation of ACD models. This is accomplished by setting the dependent variable equal to the square root of the duration and imposing a conditional mean equation of zero. The resulting parameter values provide consistent estimates of the parameters used to forecast the expected duration.

Additionally, an estimate of the conditional distribution can be obtained non-parametrically by considering the residuals $\widehat{\epsilon}_i = \frac{x_i}{\widehat{\psi}_i}$ where $\widehat{\psi}_i = E_{i-1}\left(x_i | \widehat{\theta} \right)$. Under correct specification the standardized durations $\widehat{\epsilon}_i$ should be i.i.d. and the distribution can be estimated using non-parametric methods such as kernal smoothing. Alternatively, it is often more informative to consider the baseline hazard. Given an estimate of the density the baseline hazard is obtained from (12). Engle and Russell (1998) therefore propose a semiparametric estimation procedure where in the first step QMLE is performed using the exponential distribution and in a second state the density of ϵ is estimated nonparametrically. This is referred to as a semiparametric ACD model.

ACD Model Diagnostics The properties of the standardized duration also provide a means to asses the goodness of fit of the estimated model. For example, the correlation structure, or other types of dependence can be tested. Engle and Russell (1998) suggest simply examining the Ljung-Box statistic although other types of nonlinear dependence can be examined.

Engle and Russell (1998) suggest examining autocorrelations associated with nonlinear transformations of the residuals $\widehat{\epsilon}_i$, for example, squares or square roots. An alternative test of nonlinearity advocated in Engle and Russell (1998) is to divide the diurnally adjusted durations into bins. Then regress $\widehat{\epsilon}_i$on a constant and indicators for the magnitude of the previous duration. One indicator must be omitted to avoid perfect multicollinearity. If the $\widehat{\epsilon}_i$ are indeed i.i.d. then there should be no predictability implied from this regression. Often these tests suggests that the linear specification tends to over predict the duration following extremely short or extremely long durations. This suggests that a model where the expectation is more sensitive following short durations and less sensitive following long durations may work well.

Engle and Russell (1998) suggest examining autocorrelations associated with nonlinear transformations of the residuals $\widehat{\epsilon}_i$, for example, squares or square roots. An alternative test of nonlinearity advocated in Engle and Russell (1998) is to divide the diurnally adjusted durations into bins. Then regress $(\widehat{\epsilon}_i - 1)$ on indicators for the magnitude of the previous duration. If the $\widehat{\epsilon}_i$ are indeed i.i.d. then there should be no predictability implied from this regression. Often these tests suggests that the linear specification tends to over predict the duration following extremely short or extremely long durations. This suggests that a model where the expectation is more sensitive following short durations and less sensitive following long durations may work well.

Tests of the distributional assumptions of ϵ can also be examined. A general test is based on the fact that the integrated intensity over the duration

$$u_i = \int_{s=t_{i-1}}^{t_i} \lambda(s|N(s), t_{i-1}, t_{i-2}, ..., t_0) ds \tag{18}$$

will be distributed as a unit exponential as discussed in Russell (1999). Often this takes a very simple form. For example, substituting the exponential intensity (15) into (18) simply yields the residual $\widehat{u}_i = \widehat{\epsilon}_i = \frac{x_i}{\widehat{\phi}_i \widehat{\psi}_i}$. Similarly, for the substituting the Weibull intensity (16) into (15) yields $\widehat{u}_i = \left(\frac{\Gamma\left(1+\frac{1}{\gamma}\right) x_i}{\widehat{\phi}_i \widehat{\psi}_i} \right)^{\gamma}$. The variance of u_i should be unity leading Engle and Russell (1998) to suggest the test statistic

$$\sqrt{N(T)} \frac{(\widehat{\sigma}_u - 1)}{\sqrt{8}} \tag{19}$$

which should have a limiting standard normal distribution. This is a formal test for remaining excess dispersion often observed in duration data. Furthermore, since the survivor function for an exponential random variable U is simply $exp(-u)$ a plot of the the negative of the log of the empirical survivor function should be linearly related to u_i with a slope of unity hence providing a graphical measure of fit.

Nonlinear ACD models The tests for nonlinearity discussed above often suggest nonlinearity. Zhang Russell and Tsay (2000) propose a nonlinear threshold ACD model with this feature in mind. Here the dynamics of the conditional mean are given by

$$\psi_i = \begin{cases} \omega^1 + \alpha^1 x_{i-1} + \beta^1 \psi_{i-1} & \text{if } x_{i-1} \le a_1 \\ \omega^2 + \alpha^2 x_{i-1} + \beta^2 \psi_{i-1} & \text{if } a_1 < x_{i-1} \le a_2 \\ \omega^3 + \alpha^3 x_{i-1} + \beta^3 \psi_{i-1} & \text{if } a_2 < x_{i-1} \end{cases}$$

where $a_1$and a_2 are parameters to be estimated. Hence the dynamics of the expected duration depend on the magnitude of the previous duration. Indeed, using the same IBM data as analyzed in Engle and Russell (1998) they find $\alpha^1 > \alpha^2 > \alpha^3$ as expected from the nonlinear test results. Estimation is performed using a combination of a grid search across a_1 and a_2 and maximum likelihood for all pairs of a_1 and a_2. The MLE is the pair of a_1 and a_2 and the corresponding parameters of the ACD model for each regime that produces the highest maximized likelihood.

Another nonlinear ACD model applied in Engle and Lunde (1999), Russell and Engle (2002) is the Nelson Form ACD model. The properties of the Nelson form ACD are developed in Bauwens and Giot (2000). We refer to this as the Nelson form ACD model because it is in the spirit of Nelson (1991) EGARCH model and we want to minimize confusion with the version of the ACD model that uses the exponential distribution for ε. Here the log of the expected duration follows a linear specification.

$$\ln(\psi_i) = \omega + \sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln(\psi_{i-j})$$

This formulation is particularly convenient when other market variables are included in the ACD model since non-negativity of the expected duration is directly imposed.

An interesting approach to nonlinearity is taken in Fernandes and Grammig (2002) who propose a class of nonlinear ACD models. The parametrization is constructed using the Box Cox transformation of the expected duration and a flexible non-linear function of ϵ_{i-j} that allows the expected duration to respond in a distinct manner to small and large shocks. The model nests many of the common ACD models and is shown to work well for a variety of duration data sets.

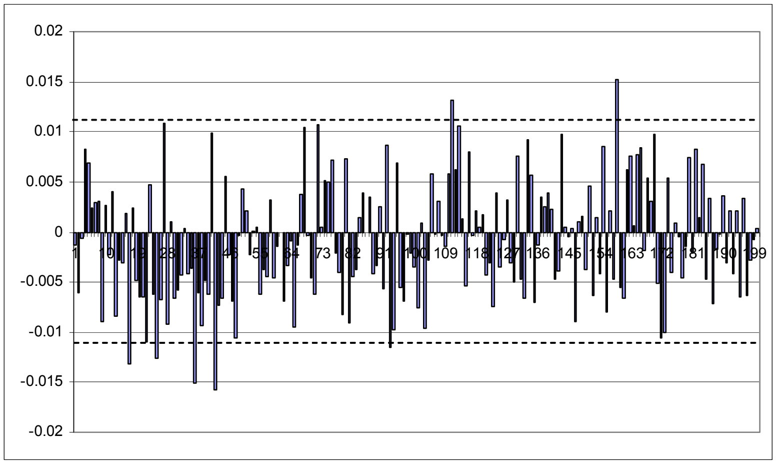

Figure 8: Autocorrelations for ACD duration residuals $\widehat{\epsilon}_i = \frac{x_i}{\widehat{\psi}_i}$

ACD example Appendix (A) contains parameter estimates for an EACD(3,2) model estimated using the GARCH module of EVIEWS. The durations were first adjusted by dividing by the time of day effect estimated by linear splines in figure 3. The sum of the α and β coefficients is in excess of .999 but less than one indicating strong persistence but the impact of shocks dies off after a sufficient period of time. A plot of the autocorrelations of the residuals is presented in figure 8. The autocorrelations are no longer all positive and appear insignificant. A formal test for the null that the first 15 auto correlations are zero yields a 12.15 with a p-value of 67%.

A test for remaining excess dispersion in (19) yields a test statistic of $\frac{\sqrt{32365}(1.41-1)}{\sqrt{8}} = 26.07$. There is evidence of excess dispersion indicating that it is unlikely that ε is exponential. However, under the conditions of the corollary the parameter estimates can be viewed as QML estimates. A plot of the nonparametric hazard is given in figure 9. The estimate was obtained using a nearest neighbor estimator. The hazard is nearly monotonically decreasing indicating that the longer it has been since the last transaction the less likely it is for a transaction to occur in the next instant. This is a common finding for transactions data.

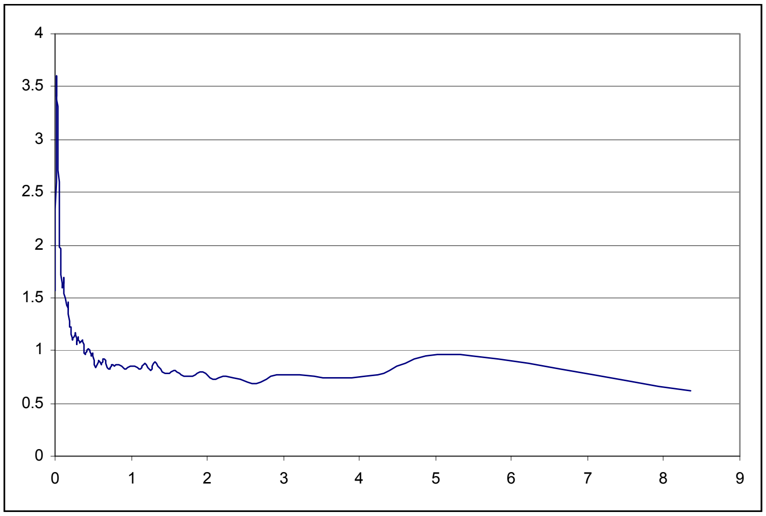

Figure 9: Non-parametric estimate of baseline intensity.

2.1.2 Thinning point processes

The above discussion focuses on modeling the arrival of transaction events. For some economic questions it may not be necessary to examine the arrival of all transactions. Rather, we might want to focus on a subset of all transactions which have special meaning. For example, when foreign exchange data are examined many of the posted quotes appear to be simply noisy repeats of the previous posted quotes. Alternatively, for stock data, often times many transactions are recorded with the exact same transaction price. Another example might be the time until a set amount of volume has been transacted - a measure related to liquidity.

Depending on the question at hand, we might be able to focus on the subset of event arrival times for which the marks take special importance - such as a price movement or attained a specified cumulative volume. In this case, the durations would correspond to the time interval between marks of special importance. If the timing of these special events is important, then econometrically this corresponds to focusing on the distribution of $\tau_{\min}$ in (5). The sequence of arrival times corresponding to the times at which the marks take special values is called a thinned point process since the new set of arrival times will contain fewer events than the original series.

Engle and Russell (1998) refer to the series of durations between events constructed by thinning the events with respect to price and volume as price based and volume based durations. They suggest that the ACD model might be a good candidate for modeling for these thinned series. In Engle and Russell (1997) a Weibull ACD model is applied to a thinned series of quotes arrival times for the Dollar Deutchemark exchange rate series. More formally if t_i denotes the arrival times of the original series of quotes then let $\tau_0 = t_0$ Next, let $N^*(t)$denote the counting function for the thinned process defined by $t_{N^*(t)+1} = t + \tau_{N^*(t)+1}$ where $\tau_{N^*(t)+1} = \min_{\tau>0} \left|p_{N(t+\tau)} - p_{N^*(t)}\right| > c$ and $N(t_0) = N^*(t_0) = 0$. So, the sequence of durations τ_i corresponds to the price durations defined by price movements greater than a threshold value c. An ACD Weibull model appears to provide a nice fit for the thinned series.

The authors suggest that this provides a convenient way of characterizing volatility when the data are irregularly spaced. The intuition is that instead of modeling the price change per unit time, as is typically done for volatility models constructed using regularly spaced data, the model for price durations models the time per unit price change. In fact, assuming that the price process locally follows a geometric Brownian motion leads to implied measures of volatility using first crossing time theory.

Engle and Lange (2001) combine the use of price durations discussed above with the cumulative signed volume transacted over the price duration to measure liquidity. For each of the US stocks analyzed, the volume quantity associated with each transaction is given a positive sign if it is buyer initiated and negative sign if it is seller initiated using the rule proposed by Lee and Ready (1991). This signed volume is then cumulated for each price duration. The cumulative signed volume, referred to as VNET, is the total net volume that can be transacted before inducing a price move hence providing a time varying measure of the depth of the market. For each price duration several other measures are also constructed including the cumulative (unsigned) volume and number of transactions. Regressing VNET on these market variables suggest that market depth is lower following periods high transaction rates, and high volatility. While VNET increases with past volume it does so less than proportionally indicating that order imbalance as a fraction of overall (unsigned) volume decreases with overall volume. Jointly these results suggest that market depth tends to be lower during periods of high transaction rates, high volatility, and high volume. In an asymmetric information environment this is indicative of informed trading transpiring during period of high transaction rates and high volume.

2.2 Modeling in Tick Time - the Marks

Often, economic hypothesis of interest are cast in terms of the marks associated with the arrival times. For example, many hypothesis in the asymmetric information literature focusing on the mechanism by which private information becomes impounded asset prices. In a rational expectations environment, the specialist will learn about a traders private information from the characteristics of their transactions. Hence, many asymmetric information models of financial participants have implications about how price adjustments should depend on the characteristics of trades such as volume or frequency of transactions. By the very nature of market micro structure field, these theories often need to be examined at the transaction by transaction frequency This section of the paper examines transaction by transaction analysis of the marks. We refer to this approach generally as tick time analysis.

2.2.1 VAR Models for Prices and Trades in Tick Time

Various approaches to tick time modeling of the marks have been considered in the literature. The approaches are primarily driven by the economic question at hand as well as assumptions about the role of the timing of trades. Many times the hypothesis of interest can be expressed as how the prices adjust given characteristics of past order flow. In this case, it is not necessary to analyze the joint distribution in (1) but only the marginal distribution of the mark given in (3).

Perhaps the simplest approach in application is to assume that timing of past of transactions has no impact on the distribution of the marks, that is $f_y\left(y_{i+1}|\widehat{t}_i, \widehat{y}_i\right) = f_y\left(y_{i+1}|, \widehat{y}_i\right).$ This is the approach taken in Hasbrouck (1991) where the price impact of a trade on future transaction prices is examined. Hasbrouck focuses on the midpoint of the bid and ask quotes as a measure of the price of the asset. We refer to this as the midprice and would appear to be a good approximation to the value of the asset given the information available. Hasbrouck is interested in testing and identifying how buyer and seller initiated trades differ in their impact on the expectation of the future price.

Let Δm_i denote the change in the midprice from the (i-1)th to the i^{th} transactions or $m_i - m_{i-1}$. The bid and ask prices used to construct the midprice are those prevailing just prior to the transaction time t_i. Let w_i denote the signed volume of a transaction taking a positive value if the i^{th} trade is buyer initiated and a negative value if it is seller initiated. The direction of trade is inferred using the Lee and Ready rule discussed in section . Hasbrouck persuasively argues that market frictions induce temporal correlations in both the price and volume series. Regulations require the specialist to operate an "orderly" market meaning that the price should not fluctuate dramatically from one trade to the next. Hence in the face of a large price move the specialist will have to take intervening transactions at intermediate prices to smooth the price transition. Volume might also be autocorrelated as a result of the common practice of breaking up large orders into multiple small orders to achieve at a better overall price than had the order been executed in one large transaction. Finally, since neither price nor direction of trades can be viewed as exogenous the series must be analyzed jointly to get a full picture of the series dynamics. Hasbrouck analyzes the bivariate system using the following VAR:

$$\begin{aligned} \Delta m_i &= \sum_{j=1}^{J} a_j \Delta m_{i-j} + \sum_{j=0}^{J} b_j w_{i-j} + v_{1i} \\ w_i &= \sum_{j=1}^{J} c_j \Delta m_{i-j} + \sum_{j=1}^{J} d_j w_{i-j} + v_{2i} \end{aligned} \tag{20}$$

Notice that the signed volume appears contemporaneously on the right hand side of the quote update equation. The quote revision equation is therefore specified conditional on the contemporaneous trade. In reality, it is likely that the contemporaneous quote revision will influence the decision to transact as the marginal trader might be enticed to transact when a new limit order arrives improving the price. In fact, our application suggests some evidence that this may be the case for the Airgas stock. We estimate a VAR is for price changes and a trade direction indicator variable taking the value 1 if the trade is deemed buyer initiated and -1 if the trade is deemed seller initiated using the Lee and Reedy rule. Volume effects are not considered here. As expected, the b_j coefficients tend to be postive meaning that buys tend to lead to increasing quote revisinos and sells tend to lead to decreasing quote revisions. The market frictions suggest that the full impact of a trade may not be instantaneous, but rather occur over a longer period of time. To gauge this effect the VAR can be expressed as an infinite vector moving average model. The coefficients then form the impulse response function.

The cumulants of the impulse response functions then provide a measure of the total price impact of a trade. Since the model operates in transaction

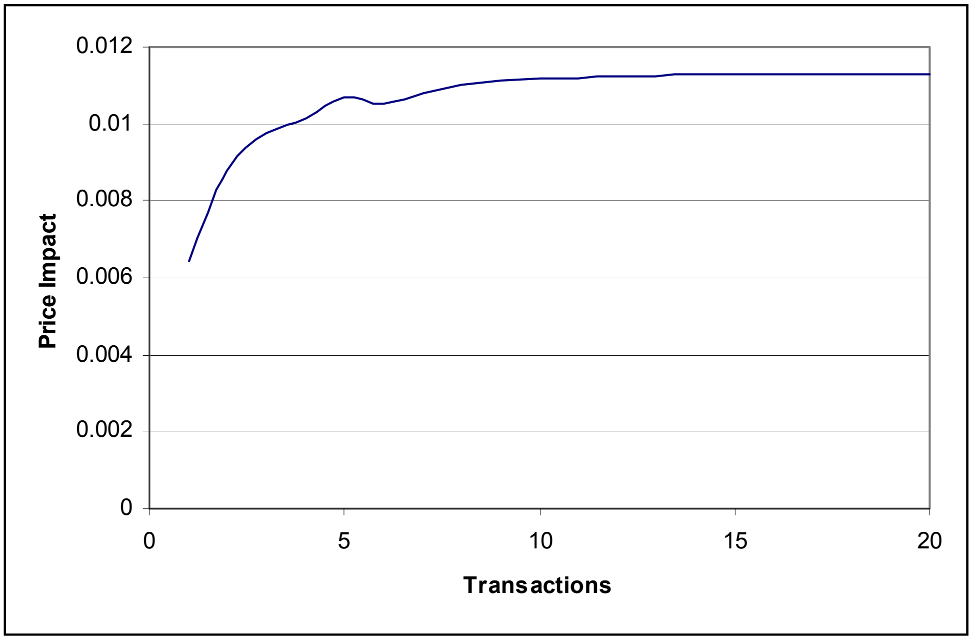

Figure 10: Cummulative price impact of an unexpected buy

time these price impacts are therefore also measured in transaction time. The asymptote associated with the cummulative impulse response function is then defined as the total price impact of a trade. Since the data are indexed in tick time the price impact is measured in units of transactions. The results indicate that it can take several transactions before the full price impact of a transaction is realized. Figure 1 below presents a price impact plot for the stock Airgas. Similar to Hasbrouck's findings using the earlier data sets we find that the price impact can take many periods to be fully realized and that the function is concave. The price impact is in the expected direction - buys increase the price and sells decreased the priced. Looking at a cross section of stocks, Hasbrouck constructs measures for the information asymmetry by taking the ratio of the price impact of a 90th percentile volume trade over the average price. This measure is decreasing with market capitalization suggesting that firms with smaller market capitalization have larger information asymmetries.

Engle and Dufour also analyze the price impact of trades, but relax the assumption that the timing of trades has no impact on the marginal distribution of price changes. Easley and O'Hara (1992) propose a model with informed and uninformed traders. On any given day private information may or may not exist. Informed and uninformed traders are assumed to arrive in a random fashion. Informed traders only transact when private information is present so on days with no private information all transactions are by the uninformed. Days with high transaction rates are therefore viewed as days with more informed trading. Admatti and Pfeiderer (1988) suggest that in the presence of short sales constraints the timing of trades should also carry information. Here, long durations imply bad news and suggest falling prices. The important thread here is that the timing of trades should carry informative. This leads Dufour and Engle (2000) to consider expanding Hasbrouck's VAR structure to allow the durations to impact price updates. The duration between trades is treated as a predetermined variable that influences the informativeness of past trades on future quote revisions. This is done with by allowing the b_j parameters in (20) to be time varying parameters. In particular,

$$b_j = \gamma_j + \sum_{k=1}^{K} \delta_k D_{j,i-k} + \eta_i \ln(x_{i-j})$$

where $D_{j,i-k}$ are dummy variables for the time of day and $x_i$is the duration. Since the b_j dictate the impact of past trades on quote revisions it is clear that these effects will be time varying whenever the coefficients d_k or η are non-zero. The model therefore extends the basic VAR of Hasbrouck by allowing the impact of trades to depend on the time of day as well the trading frequency as measured by the elapsed time between trades. A similar adjustment is made to the coefficients d_j in the trade equation.

The modified VAR is specified conditional on the durations and may therefore be estimated directly. Impulse response function, however, will require complete specification of the trivariate system of trades quotes and arrival times. Dufour and Engle propose using the ACD model for the arrival times.

Parameters are estimated for 18 stocks. As in the simple Hasbrouck VAR the impact of past transactions on quote revisions tends to be positive meaning that buys tend to lead to increasing quote revisions and sells lead to decreasing quote revisions. The magnitude of the d_k indicate some degree of time of day effects in the impact of trades. Trades near the open tend to be more informative, or have a larger price impact than trades at other times during the day although this effect is not uniform across all 18 stocks. The coefficients on the durations tend to be negative indicating that the longer it has been since the last trade, the smaller the price impact will be. The cumulative impulse response from an unanticipated order can be examined only now these functions will depend on the state of the market as dictated by transaction rates measured by the durations. The result is that the price impact curves will shift up when the transaction occurs with a short duration and shift down when transactions occur with long durations.

The VAR approach to modeling tick data is particularly appealing due to its ease of use. Furthermore, information about the spacing of the data can be included in these models as suggested in Dufour and Engle (2000). These VAR models can, of course, be expanded to include other variables such as the bid ask spread or measures of volatility.

2.2.2 Volatility Models in Tick Time

The VARs of the previous section where proved useful in quantifying the price impact of trades. As such, they focus on the predictable change in the quotes given characteristics of a transaction. Alternatively, we might want to ask how characteristics of a transaction affect our uncertainty about the quote updates. Volatility models provide a means of quantifying our uncertainty.

The class of GARCH models of Engle (1982) and Bollerslev (1986) have proven to be a trusted work horse in modeling financial data at the daily frequency. Irregular spacing of transaction by transaction data seems particularly important for volatility modeling of transaction by transaction data since volatility is generally measured over fixed time intervals. Furthermore, it is very unlikely that, all else equal, the volatility of the asset price over a one hour intertrade duration should be the same as the volatility over a 5 second intertrade duration. Engle (2000) proposes adapting the GARCH model for application to irregularly spaced transaction by transaction data.

Let the return from the $i-1^{th}$ to the i^{th} transaction be denoted by r_i. Define the conditional variance per transaction as

$$V_{i-1}\left(r_i|x_i\right) = h_i \tag{21}$$

where this variance is defined conditional on the contemporaneous duration as well as past price changes. The variance of interest, however, is the variance per unit time. This is related to the variance per transaction as

$$V_{i-1}\left(\frac{r_i}{\sqrt{x_i}}|x_i\right) = \sigma_i^2 \tag{22}$$

so that the relationship between the two variances is $h_i = x_i\sigma_i^2$.

The volatility per unit time is then modeled as a GARCH process. Engle proposes an ARMA(1,1) model for the series $\frac{r_i}{\sqrt{x_i}}$. Let e_i denote the innovation to this series. If the durations are not informative about the variance per unit time then the GARCH(1,1) model for irregularly spaced data is simply

$$\sigma_i^2 = \dot{\omega} + \dot{\alpha}e_{i-1}^2 + \dot{\beta}\sigma_{i-1}^2 \tag{23}$$

where we have placed dots above the GARCH parameters to differentiate these from the parameters of the ACD model with similar notation. Engle terms this model the UHF-GARCH model or Ultra High-Frequency GARCH model.

A more general model is inspired by the theoretical models of Easley and O'Hara (1992) and Admatti and Pleiderer (1985) discussed in section (2.2.1) above. These models suggest that the timing of transactions is related to the likelihood of asymmetric trading and hence more uncertainty. Engle therefore proposes augmenting the GARCH(1,1) models with additional information about the contemporaneous duration and perhaps other characteristics of the market that might be thought to carry information about uncertainty such as spreads and past volume.

While the model specifies the volatility per unit time it is still operates in transaction time updating the volatility on a time scale determined by transaction arrivals. If calendar time forecasts of volatility are of interest then a model for the arrival times must be specified and estimated. Toward this end Engle proposes using an ACD model for the arrival times. If the arrival times are deemed exogenous then the ACD model and the GARCH model can be estimated separately although this estimation may be inefficient. In particular, under the exogeneity assumption the ACD model could be estimated first and then the volatility model could be specified conditional on the contemporaneous duration and expected duration in a second step using canned GARCH software that admits additional explanatory variables. This is the approach taken in Engle (2000) where estimation is performed via (Q)MLE. Engle considers the following specification

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

where ψ_i is the expected duration obtained from an ACD model and ξ_{i-1} characterizes the long run volatility via exponential smoothing of the squared return per unit time.

An alternative approach to modeling volatility of irregularly spaced data was simultaneously and independently developed in Ghysels and Jasiak (1998). Here the authors suggest using temporal aggregation to handle the spacing of the data. GARCH models are not closed under temporal aggregation so the authors propose working with the weak GARCH class of models proposed by Drost and Nijman (1993). For the weak GARCH class of models Drost and Nijman derive the implied low frequency weak GARCH model implied by a higher frequency weak GARCH model. Ghysels and Jasiak propose a GARCH model with time varying parameters driven by the expected spacing of the data. This approach is complicated, however, by the fact that the temporal aggregation results apply to aggregation from one fixed interval to another, exogenously specified, fixed interval. The spacing of the transactions data is not fixed and it is unlikely that transaction arrival times are exogenous. Nevertheless, the authors show that the proposed model is an exact discretization of a time deformed diffusion with ACD as the directing process. The authors propose using GMM to estimate the model.

2.3 Models for discrete prices

Discrete prices in financial markets pose an additional complication in the analysis of financial data. For the US markets the graduation to decimalization is now complete, but we still find price changes clustering on just a handful of values. This discreteness can have an important influence on analysis of prices. Early analysis of discrete prices focused on the notion of a "true" or efficient price. The efficient price is defined as the expected value of the asset given all currently available public information. The focus of these early studies, therefore, was on the relationship between the efficient price and observed discrete prices. In particular, much emphasis was placed on how inference about the efficient price is influenced by measurement errors induced by discreteness.

Let P_t denote the observed price at time t and let P_t^e denote the efficient or "true" price of the asset at time t. Early models for discrete prices can generally be described in the following setting:

$$
\begin{aligned}
&P_t^e = P_{t-1}^e + v_t \\
&P_t = round\left(P_t^e + c_t Q_t, d\right) \\
&v_t \tilde{} N\left(0, \sigma_t^2\right)
\end{aligned} \tag{24}
$$

where $d \geq 0$ is the tick size and *round* is a function rounding the argument to the nearest tick. Q_t is an unobserved i.i.d. indicator for whether the trade was buyer or seller initiated taking the value 1 for buyer initiated and -1 for seller initiated trades and probability given by 1/2. The parameter $c_t \geq 0$ denotes the cost of market making. It includes both tangible costs of market making as well as compensation for risk.

With $c = 0$ and $\sigma_t^2 = \sigma^2$we obtain the model of Gottlieb and Kalay (1985)[1]. When $d = 0$ (no rounding) we obtain the model of Roll (1984). Harris (1990) considers the full model in (24). In this case, we can write

$$
\Delta P_t = c\left(Q_t - Q_{t-1}\right) + \eta_t - \eta_{t-1} + v_t \tag{25}
$$

where $\eta_t = P_t^e - P_t$ is the rounding error. The variance of the observed price series is therefore given by

$$
E\left(\Delta P_t\right)^2 = \sigma^2 + 2c^2 + E\left(\eta_{t+1} - \eta_t\right)^2 \tag{26}
$$

. Hence the variance of the observed transaction price will exceed the variance by an amount that depends on the cost of market making and the discrete rounding errors. Furthermore, the first order serial correlation is given by

$$
E\left(\Delta P_t \Delta P_{t-1}\right) = -c^2 + E\left(\eta_{t+1} - \eta_t\right)\left(\eta_t - \eta_{t-1}\right) \tag{27}
$$

which is shown to be negative in Harris (1990). The first order serial correlation will be larger when the cost of market making is larger and depends on the discreteness rounding errors. The Harris model goes a long way in describing key features of price discreteness and the implications regarding inference on the efficient price dynamics but it is still very simplistic in several dimensions since it assumes that both the volatility of the efficient price and the cost of market making is constant. As new information hits the market the volatility of the efficient price will change. Since part of the cost of making market is the risk of holding the asset the cost of making market will also be time varying.

[1]For simplicity we have neglected a drift term in the efficient price equation and the possiblity of dividend payments considered in Gottlieb and Kalay (1985).

Hasbrouck (1999) builds on the previous discrete price literature by relaxing these two assumptions - both the volatility of the efficient price and the cost of making market are time varying. He also proposes working with bid and ask prices as opposed to the transaction prices circumventing the need to sign trades as buyer or seller initiated. P_t of 24 is therefore replaced by two prices P_t^a and P_t^b, the bid and ask prices respectively where

$$\begin{array}{c} P_t^a = ceiling\left(P_t^e + c_t^a, d\right) \\ P_t^b = floor\left(P_t^e - c_t^b, d\right) \end{array} \tag{28}$$

$c_t^a > 0$ and $c_t^b > 0$ are the cost of exposure on the ask and bid side respectively. These costs are the economic cost to the specialist including both the fixed cost of operation and the expected cost incurred as a result of the obligation to trade a fixed quantity at these prices with potentially better informed traders. The ceiling function rounds up to the nearest discrete tick and the floor function rounds down to the nearest tick recognizing the market maker would not set quotes that would fail to cover the cost.

The dynamics of this model are not suited for the type of analysis presented by Harris due to it's more complex dynamic structure. Instead, Hasbrouck focuses on characterizing the dynamics of the both cost of exposure and, second, estimating models for the volatility of the efficient price given only observations of the perturbed discrete bid and ask prices. Hasbrouck proposes using a GARCH model for the dynamics of the efficient price volatility σ_t^2. The dynamics of the cost of exposure is assumed to be of an autoregressive form.

$$\begin{aligned} \ln\left(c_t^a\right) &= \mu_t + \alpha\left(\ln\left(c_t^a\right) - \mu_{t-1}\right) + \nu_t^\alpha \\ \ln\left(c_t^b\right) &= \mu_t + \alpha\left(\ln\left(c_t^b\right) - \mu_{t-1}\right) + \nu_t^\beta \end{aligned} \tag{29}$$

where μ_t is a common deterministic function of time of day and α is the common autoregressive parameter. ν_t^α and ν_t^β are assumed to be i.i.d. and independent of the efficient price innovation v_t.

Estimation of the model is complicated by the fact that the efficient price is inherently unobserved - only the discrete bid and ask quotes are observed. Hasbrouck (1999) proposes using a non-Gaussian, non-linear state space of Kitagawa (1987) and Hasbrouck (1999b) and Manrique and Sheppard (1997) propose using MCMC methods which treat the price at any given date as an unknown parameter. We refer the reader to this literature for further details on the estimation.

The early work with no time varying parameters focused on the impact of discrete rounding errors on inference regarding the efficient price. Hasbrouck also treats the object of interest as the dynamics of the efficient price and demonstrates a methodology to study the second moment dynamics of the efficient price while accounting for the discrete rounding errors. In addition, the model allows for asymmetric cost of exposure. In some states of the world the specialist may set quotes more conservatively on one side of the market than the other.

More recently Zhang Russell and Tsay (2000) appeal to the asymmetric information microstructure theory that suggests that transaction characteristics should influence the market makers perception of the exposure risk. They include in the cost of exposure dynamics (29) measures of the order imbalance and overall volume and find that the cost of exposure is affected by order imbalance. In particular, unexpectedly large buyer initiated volume tends to increase the cost of exposure on the ask side and decrease the cost of exposure on the bid side with analogous results for unexpected seller initiated volume. These effects are mitigated, however, the larger the total volume transacted.

If the structural parameters are not of primary interest then an alternative is to directly model transaction prices with a reduced form model for discrete valued random variables. This is the approach taken in Hausman Lo and MacKinlay (1992). They propose modeling the transaction by transaction price changes with an ordered Probit model. In doing so, the structural models linking the unobserved efficient price to the observed transaction price is replaced by a reduced from Probit link. The model is applied to transaction by transaction price dynamics. Let k denote the number of discrete values that the price changes can take which is assumed to be finite. Let s_i denote a vector of length k taking the j^{th} column of the kxk identity matrix if the j^{th} state occurs on the i^{th} transaction. Let π_i denote a k dimensional vector with j^{th} where $\pi_i = E\left(s_i|I_{i-1}\right)$ where I_i is the information set associated with the i^{th} transaction. Clearly, the j^{th} element of x_i denotes the conditional probability of the j^{th} state occurring. At the heart of the Probit model lies the assumption that the observed discrete transaction price changes can be represented as a transformation of a continuous latent price given by $\Delta P_i^{*\,\tilde{}}N\left(\mu_i, \sigma_i^2\right)$ where μ_i and σ_i^2 are the mean and variance of the latent price given I_{i-1}. The Hausman Lo and MacKinlay model assumes that the j^{th} element of π_i is given by

$$\pi^{j\,\tilde{}}F_{P_i^*}\left(c_{j-1}\right) - F_{P_i^*}\left(c_j\right) \tag{30}$$

where $F_{P_i^*}$ is the cdf associated with the price changes ΔP_i^* and c_j, $j =$

$1, k-1$ are time invariant parameters.

Since bid ask bounce induces dependence in the price changes it is natural to allow the conditional mean of price changes to depend on past price changes. Hausman Lo and Mackinlay are particularly interested in testing asymmetric information theories regarding the information content in a sequence of trades. In particular, they study how transaction prices respond to a sequence of buyer initiated trades versus a sequence of seller initiated trades. By conditioning on recent buys and sells the authors find evidence that persistent selling predicts falling prices and persistent buying predicts rising prices. The authors also suggest that the conditional variance may depend on the contemporaneous duration so that the variance associated with a price change over a long duration may not be the same as the variance of a price change associated with a relatively short duration. Indeed, they find that long durations lead to higher variance per transaction.

Russell and Engle (2002) are also interested in reduced form models for discrete prices that explicitly account for the irregular spacing of the data. They propose joint modeling the arrival times and the price changes as a marked point process. The joint likelihood for the arrival times and the price changes is decomposed into the product of the conditional distribution of the price change given the duration and the marginal distribution of the arrival times which are assumed to be given by an ACD model. More specifically if x_i denotes the i^{th} intertransaction duration and $\widehat{z}_i = (z_i, z_{i-1}, z_{i-2}...)$ then

$$f\left(x_{i+1}, \Delta p_{i+1} | \widehat{x}_i, \Delta \widehat{p}_i\right) = \varphi\left(\Delta p_{i+1} | \widehat{x}_i, \Delta \widehat{p}_{i+1}\right) \chi\left(x_i | \widehat{x}_i, \Delta \widehat{p}_i\right)$$

where φ denotes the distribution of price changes given the past price changes and durations as well as the contemporaneous duration. χ denotes the distribution of price changes given the past price changes and returns. Engle and Russell propose using the ACD model for the durations and propose the Autoregressive Conditional Multinomial (ACM) model for the conditional distribution of the discrete price changes. A simple model for the price dynamics might assume a VARMA model for the state vector s_i. Since the state vector is simply a vector of ones and zeros its expectation should be bounded between zero and one. Russell and Engle use the logistic transformation to directly impose this condition. Using the logistic link function the VARMA model is expressed in terms of the log odds. Let h_j denote a $k-1$ vector with j^{th} element given by $ln\left(\pi^j/\pi^k\right)$. Let $\widetilde{s}_i$ and $\widetilde{\pi}_i$ denote $k-1$ dimensional vectors consisting of the first k-1 elements of s_i and π_i. Hence the k^{th} element has been omitted and is referred to as the base state. Then the ACM(u, v) model with duration dependence is given by:

$$h_i = c + \sum_{m=1}^{u} A_m \left(\widetilde{s}_{i-m} - \widetilde{\pi}_i\right) + \sum_{m=1}^{v} B_m h_{i-m} + \sum_{m=1}^{w} \chi_m \ln\left(x_{i-m+1}\right) \qquad (31)$$

where A_m and B_m are $(k-1)\,x\,(k-1)$ parameter matrices and ω and χ_m are $k-1$ parameter vector. Given the linear structure of the log odds VARMA the choice of the the base state is arbitrary. The first $k-1$ probabilities are obtained by applying the logistic link function

$$\pi_i = \frac{1}{1 + \iota' \exp\left(h_i\right)} \exp\left(h_i\right) \qquad (32)$$

where ι' is a $k-1$ vector of ones and $\exp(h_i)$ should be interpreted as applying the exponential function element by element. The omitted state is obtained by imposing the condition that the probabilities sum to 1. The ACD(p,q) specification for the durations allows feedback from the price dynamics into the duration dynamics as follows:

$$\ln\left(\psi_i\right) = \omega + \sum_{m=1}^{p} \alpha_m \frac{x_{i-m}}{\psi_{i-m}} + \sum_{m=1}^{q} \beta_m \ln\left(\psi_{i-m}\right) + \sum_{m=1}^{r} \left(\rho_m \Delta p_{i-m} + \zeta_m \Delta p_{i-m}^2\right)$$

For the stocks analyzed, the longer the contemporaneous duration the lower the expected price change and large price changes tend to be followed by short durations.

Another reduced form model for discrete prices is proposed by Rydberg and Sheppard (1999). The model decomposes the discrete price changes into the trivariate process $\Delta P_i = Z_i D_i M_i$ where Z_i is an indicator for the i^{th} transaction price change being non-zero and is referred to as activity. Conditional on a price move $(Z_i \neq 0)$ D_i is takes the value of 1 or -1 denoting an upward or downward price change respectively. Given a non-zero price change and its direction M_i is the magnitude of the price change given it is non-zero and the direction. The authors suggest decomposing the distribution of price changes given an information set I_i $\Pr\left(\Delta P_i | I_{i-1}\right) = \Pr\left(Z_i | I_{i-1}\right) \Pr\left(D_i | Z_i, I_{i-1}\right) \Pr\left(M_i | Z_i, D_i, I_{i-1}\right)$. The authors propose modeling the binary variables $Z_i$and $D_i$following an autoregressive logistic process first proposed by Cox (1958). A simple version for the activity variable is given by:

$$\ln\left(\frac{\Pr\left(Z_i = 1\right)}{1 - \Pr\left(Z_i = 1\right)}\right) = c + \sum_{m=1}^{u} Z_{i-m} \qquad (33)$$

The direction indicator variable is modeled in a similar fashion. Finally, the magnitude of the price change is modeled by a distribution for count data. Hence it is positively valued over integers. The integers here are measured in units of ticks, or the smallest possible price change.

The Russell Engle ACM approach and the Rydberg Sheppard components model are very similar in spirit both implementing an autoregressive structure. While this decomposition breaks the estimation down into a sequence of simpler problems it comes with a cost. In order to estimate the model sequentially the first model for the activity can not be a function of lagged values of $Pr\left(D_i|I_{i-1}\right)$ or $\Pr(M_i|I_{i-1})$. Similarly the model for the direction cannot depend on the past probability associated with the magnitude. The importance of this restriction surly depends on the application at hand. A second advantage of the Rydberg Sheppard model easily accounts for a large number of states (possibly infinite).

2.4 Calendar Time Conversion

Most financial econometric analyses are carried out in fixed time units. These time intervals for many years were months or weeks or days but now time intervals of hours, five minutes or seconds are being used for econometric model building. Once the data are converted from their natural irregular spacing to regular spaced observations, econometric analysis typically proceeds without considering the original form of the data. Models are constructed for volatility, price impact, correlation, extreme values, and many other financial constructs. In this section we discuss the most common approaches used to convert irregularly spaced data to equally spaced observations and in the next sections we will examine the implications of this conversion.

Suppose the data on prices arrive at times $\{t_i; i = 1, ..., N(T)\}$ so that there are N observations occurring over time $(0, T)$. These times could be times at which transactions occur and the price could be either the transaction price or the prevailing midquote at that time. An alternative formulation would have these times as the times at which quotes are posted and then the prices are naturally considered to be midquotes. Let the log of the price at time t_i be denoted p_i^*.

The task is to construct data on prices at each fixed interval of time. Denoting the discrete time intervals by integers of $t = 1, \ldots, T$, a task is to estimate p_t. The most common specification is to use the most recent price at the end of the time interval as the observation for the interval. Thus:

$$p_t = p_i^* \text{ where } t_i \leq t < t_{i+1} \tag{34}$$

For example, Huang and Stoll (1994) use this scheme where p is the prevailing midquote at the time of the last trade. Andersen Bollerslev Diebold and Ebens (2001) use the last trade price.

Various alternative schemes have been used. One could interpolate the price path from some or all of the p^* observations and then record the value at time t. For example smoothing splines could be fit through all the data points. A particularly simple example of this uses the weighted average of the last price in one interval and the first price in the next interval:

$$\widetilde{p}_t = \left(\lambda p_i^* + (1-\lambda)\, p_{i+1}^*\right) \text{ where } t_i \leq t < t_{i+1}, \text{ and } \lambda = \frac{t - t_i}{t_{i+1} - t_i} \tag{35}$$

Andersen, Bollerslev, Diebold and Labys (2001,2002) use this procedure with midquotes to get 5 minute and 30 minute calendar time data. The advantage of this formulation is supposed to be its reduced sensitivity to measurement error in prices. Clearly this comes at a cost of using future information. The merits of such a scheme must be evaluated in the context of a particular data generating process and statistical question.

A third possibility is adopted by Hasbrouck(2002). Since the time of a trade is recorded only to the nearest second, then if t is measured in seconds, there is at most one observation per time period. The calendar price is either set to this price or it is set to the previous period price. This version follows equation 34 but does not represent an approximation in the same sense.

Returns are defined as the first difference of the series.

$$y_i^* = p_i^* - p_{i-1}^*,\ y_t = p_t - p_{t-1},\ \text{and } \widetilde{y}_i = \widetilde{p}_i - \widetilde{p}_{i-1} \tag{36}$$

thus y^* defines returns over irregular intervals while y defines returns over calendar intervals. In the case of small calendar intervals, there will be many zero returns in y. In this case, there are some simplifications. Means and variances are preserved if there never is more than one trade per calendar interval

$$\sum_{i=1}^{N(T)} y_i^* = \sum_{t=1}^{T} y_t \text{ and } \sum_{i=1}^{N(T)} \left(y_i^*\right)^2 = \sum_{t=1}^{T} y_t^2 \tag{37}$$

If the calendar time intervals are larger, then means will still be preserved but not variances as there may be more than one price in a calendar interval.

$$\sum_{i=1}^{N(T)} y_i^* = \sum_{t=1}^{T} y_t \text{ and } \sum_{i=1}^{N(T)} \left(\sum_{\substack{multiple \\ trades}} y_i^* \right)^2 = \sum_{t=1}^{T} y_t^2 \tag{38}$$

However, if the prices are Martingales, then the expectation of the cross products is zero and the expected value and probability limit of the calendar time and event time variances is the same.

When prices are interpolated, these relations no longer hold. In this scheme there would be many cases of multiple observations in a calendar interval. The mean will approximately be the same in both sets of returns, however the variances will not. The sum of squared transformed returns is given by:

$$\begin{aligned} \sum_{i=1}^{N(T)} [\widetilde{y}_i]^2 &= \sum_{i=1}^{N(T)} \left[\lambda_i p_i^* + (1-\lambda_i) p_{i-1}^* - \lambda_j p_j^* - (1-\lambda_j) p_{j-1}^*\right]^2 \\ &= \sum_{i=1}^{N(T)} \left[\lambda_i \left(p_i^* - p_{i-1}^*\right) + \left(p_i^* - p_j^*\right) + (1-\lambda_j)\left(p_j^* - p_{j-1}^*\right)\right]^2 \end{aligned} \tag{39}$$

where i and j are the events just after the end points of the calendar intervals. In the right hand side of expression 39, the returns will all be uncorrelated if the y^* are Martingale differences, hence the expected variance and the probability limit of the variance estimators will be less than the variance of the process. Furthermore, the returns will be positively autocorrelated because they are formulated in terms of future prices. This is easily seen in equation 39 because the change in price around the interval endpoints is included in both adjacent returns.

2.4.1 Bivariate Relationships

We now consider two correlated assets with Martingale prices. One of these asset prices is only observed at random time periods while the other is continuously observable. In this case the stochastic process of the infrequently observed process is defined on times $\{t_i; i = 1, ..., N(T)\}$. Let the log price of the second asset be q_t, and let its return, measured respectively in the first asset trade time and in calendar time, be:

$$z^*_{t_i} \equiv z^*_i = q_{t_i} - q_{t_{i-1}}, z_t = q_t - q_{t-1}$$

The return on the first asset is given by

$$y^*_i = \beta z^*_i + \epsilon^*_i \tag{40}$$

where the innovation is a Martingale difference sequence, independent of z, with potential heteroskedasticity since each observation may have a different time span. Since this model is formulated in transaction time, it is natural to estimate the unknown parameter beta with transaction time data. It is straightforward to show that least squares will be consistent.

Calendar time data on y can be constructed from 40 and 36. Consider the most disaggregated calendar time interval and let d_t be a dummy variable for the time periods in which a price is observed on the first asset. Then a useful expression for y_t is

$$y_t = d_t \begin{pmatrix} (\beta z_t + \epsilon_t) + (1 - d_{t-1}) (\beta z_{t-1} + \epsilon_{t-1}) \\ + (1 - d_{t-2}) (\beta z_{t-2} + \epsilon_{t-2}) + ... \end{pmatrix} \tag{41}$$

With this data and the comparable calendar data on z, we are in a position to estimate beta by ordinary least squares in calendar time. The estimator is simply

$$\widehat{\beta} = \frac{\sum\limits_{t=1}^{T} z_t y_t}{\sum\limits_{t=1}^{T} z_t^2} \tag{42}$$

which has an interesting probability limit under simplifying assumptions.

Theorem 1

If

a) (z_t, ϵ_t) are independent Martingale difference sequences with finite variance

b) d_t ~ independent Bernoulli with parameter π

Then

$$\text{plim}\widehat{\beta} = \pi\beta \tag{43}$$

Proof:

Substituting and taking probability limits:

$$\text{plim}\widehat{\beta} = \frac{1}{\sigma_z^2} E\left[z_t d_t \left((\beta z_t + \epsilon_t) + (1 - d_{t-1}) (\beta z_{t-1} + \epsilon_{t-1}) + ...\right)\right] \tag{44}$$

Writing the expectation of independent variables as the product of their expectation gives

$$\text{plim}\widehat{\beta} = \frac{1}{\sigma_z^2} \left[\begin{array}{c} E(d_t) E\left(\beta z_t^2 + z_t \epsilon_t\right) + \\ E(d_t) E(z_t) E(1 - d_{t-1}) E(\beta z_{t-1} + z \epsilon_{t-1}) + ... \end{array} \right] = \beta\pi$$

QED.

The striking result is that the regression coefficient is heavily downward biased purely because of the non-trading bias. If trading is infrequent, then the regression coefficient will be close to zero.

In this setting, researchers will often regress on many lags of z. Suppose the regression computed incorporates k lags of z.

$$y_t = \beta_0 z_t + \beta_1 z_{t-1} + ... + \beta_k z_{t-k} + \epsilon_t$$

The result is given by Theorem II.

THEOREM II:

Under the assumptions of Theorem I the regression in (11) has a probability limit

$$\text{plim}\widehat{\beta} = \pi (1 - \pi)^{j-1} \beta$$

and the sum of these coefficients approaches beta as k gets large.

PROOF:

Because z is a Martingale difference sequence, the matrix of regressors approaches a diagonal matrix with the variance of z on the diagonals. Each row of the $z\prime y$ matrix has dummy variables $d_t (1 - d_{t-1}) (1 - d_{t-2}) ... (1 - d_{t-j})$ multiplying the square of z. The result follows from independence.

QED.

The regression coefficients decline from the contemporaneous one but ultimately summing up to the total impact of asset one on two. The result is however misleading because it appears that the price movements in asset two predict future movements in asset one. There appears to be causality or price discovery between these assets merely because of the random trade times.

Similar results can be found in more general contexts including dynamic structure in the price observations and dependence with the z's. Continuing research will investigate the extent of the dependence and how results change with the economic structure.

3 Conclusion

The introduction of widely available ultra high frequency data sets over the past decade has spurred interest in empirical market microstructure. The black box determining equilibrium prices in financial markets has been opened up. Intraday transaction by transaction dynamics of asset prices, volume, and spreads are available for analysis. These vast data sets present new and interesting challenges to econometricians.

Since transactions data are inherently irregularly spaced we view the process as a marked point process. The arrival times form the points and the characteristics of the trades form the marks. We first discuss models for the timing of events when the arrival rate may be time varying. Since the introduction of the ACD model of Engle and Russell (1998) numerous other models for the timing of event arrivals have been proposed and applied to financial data. The models have been applicable to transaction arrival times or, if some arrival times are thought to be more informative than others the point process can be "thinned" to contain only those arrival times with special information. Examples include volume based durations which correspond to the time it takes for a specified amount of volume to be transacted. Another example is price durations which correspond to the time it takes for the price to move a specified amount. These models can be thought of as models of volatility where the volatility is intuitively the inverse of our usual measures of volatility - namely the time it takes for the price to move a specified amount.

Models for the marks are also discussed. Often the focus is on the transaction price dynamics or joint modeling of transaction prices and volume. If the spacing of the data is ignored then the modeling problem can be reduced to standard econometric modeling procedures of VARs, simple linear regression, or GARCH models. Models that address the inherent discreteness in transaction by transaction prices are also discussed.

Alternatively, if the spacing of the data is thought to carry information then the simple approaches may be mispecified. Choices then include conditioning the marks on the arrival times as in Hausman Lo and Mackinlay, or, if forecasting is of interest joint modeling of the arrival times. The latter approach is considered in Engle (2000), Russell and Engle (2002), Rydberg and Sheppard (1999), or Ghysels (1999) among others.

Finally, while artificially descretizing the time intervals at which prices (or other marks) is a common practice in the literature, it does not come without cost. Different descretizing schemes trade of bias associated with temporally aggregating with variance. Averaging reduces the variability but blurs the timing of events. We also show, in a stylized model, that causal relationships can be artificially induced by descretizing the data. Care should be taken in interpreting results from this type of analysis.

References

[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] Andersen, Torben and Tim Bollerslev, 1997, Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, The Journal of Finance, 52, No. , pp. 975-1005.

[3] Andersen, T., Bollerslev, T., Diebold, F.X. and Ebens, H. 2001, The Distribution of Realized Stock Return Volatility,"Journal of Financial Economics , 61, 43-76.

[4] Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. 2002 , Modeling and Forecasting Realized Volatility, Econometrica , forthcoming.

[5] Luc Bauwens, and P. Giot, 2000, The Logarithmic ACD Model: An Applicaiton to the bid-ask quotes process of three NYSE stocks, Annales d'conomie et de Statistique, 60, 117-149

[6] Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327

[7] Copeland and Galai, 1993, Information Effects and the Bid-Ask Spread, Journal of Finance 38, 1457-1469

[8] Easley and O'Hara, 1992, Time and the Process of Security Price Adjustment. The Journal of Finance 19, 69-90

[9] Drost, F. C. and T.E. Nijman 1993, Temproal Aggregation of GARCH Processes, Econometrica, 61, 909-927

[10] Engle, Robert, 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, Econometrica, 50, 987-1008

[11] Engle, Robert 2000, The Econometrics of Ultra-High Frequency Data, Econometrica, 68, 1, 1-22

[12] Engle, Robert and Dufour, 2000, Time and the Price Impact of a Trade, Journal of Finance, 55, 2467-2498

[13] Engle, Robert and J. Lange, 2001, Predicting VNET; A Model of the Dynamics of Market Depth, Journal of Financial Markets, V4N2, 113-142

[14] Engle, Robert and A. Lunde, 1999, Trades and Quotes a Bivariate Process, NYU working paper.

[15] Engle, Robert and J. Russell, 1998, Autoregressive Conditional Duration: A New Model for Irregularly Spaced Data, Econometrica 66, 5, 1127-1162

[16] Engle, Robert and J. Russell, 1997, Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the Autoregressive Conditional Duration Model, Journal of Empirical Finance, 4 187-212

[17] Engle, Robert and J. Russell, 1995b, Autoregressive Conditional Duration: A New Model for Irregularly Spaced Data, University of California, San Diego, unpublished manuscript

[18] Fernandes, Marcelo and J. Grammig, 2002, A Family of Autoregressive Conditional Duration Models Working paper Graduate School of Economics, Fundacao Getulio Vargas Praia de Botafogo.

[19] 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

[20] Ghysels, Eric and Joanna Jasiak, 1998, GARCH for Irregularly Spaced Data: the ACD-GARCH Model, Studies in Nonlinear Dynamics and Econometrics v2, n4 133-49

[21] Gottlieb, Gary and Avner Kalay, 1985, Implications of the Discreteness of Observed Stock Price, The Journal of Finance, Vol. 40, No. 1., 135-153.

[22] Grammig, J. and K.-O. Maurer, 2000, Non-monotonichazard functions and the Autoregressive Condtional Duration Model. The Econometrics Journal 3, 16-38

[23] Hasbrouck, J., 1999, The Dynamics of Discrete Bid and Ask Quotes, Journal of Finance, 54, 6, 2109-2142

[24] Hasbrouck, J., 1999a, Security Bid/Ask Dynamics with Discreteness and Clustering, Journal of Financial Markets, 2, 1, 1-28

[25] Hausman, J., A. Lo, and C. MacKinlay, 1992, An Ordered Probit Analysis of Transaction Stock Prices, Journal of Financial Economics

[26] Hawkes, A. G., 1971, Spectra of Some Self-Exciting and Mutually Excinting Point Process, Biometrika, 58, 83-90

[27] Harris, L., 1990, Estimation of Stock Price Variances and Serial Covariances from Discrete Observations, Journal of Financial And Quantiatative Analysis 25, 291-306

[28] Hasbrouck, J., 1991, Measuring the Information Content of Stock Trades, The Journal of Finance 66,1, 179-207.

[29] Roger D. Huang, Roger, and Hans R. Stoll, 1994, Market Microstructure and Stock Return Predictions, The Review of Financial Studies, Vol. 7, No. 1. 179-213.

[30] Kitagawa, G. 1987, Non-Gaussian State-Space Modeling of Nonstationary Time Series, Journal of American Statistical Association, 82, 1032-1041

[31] Lee, Charles, and M. Ready, 1991, Inferring Trade Direction from Intraday Data, The Journal of Finance, Vol. 46, No. 2., pp. 733-746

[32] Lee, S., and B. Hansen, 1994, Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, Econometric Theory 10, 29-52

[33] Lumsdaine, R., 1996, Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, Econometrica 64, 575-596.

[34] Lunde, A. 1998, A generalized Gamma Autoregressive Conditional Duration Model, Working paper Department of Economics, University of Aarhus

[35] Manrique, A. and N. Shephard, 1997, Likelihood Analysis of a Discrete Bid/Ask Price Model for a Common Stock. Working Paper. Nuffield College, Oxford University

[36] Nelson, D., 1991, Conditional Heteroskedasticity in asset returns: A new Approach, Econometrica 59, 347-370

[37] Nelson, D. and C. Cao, 1992, Inequality Constraints in the GARCH(1,1) Model, Journal of Business and Economic Statistics, 10, 229-235

[38] O'Hara, Maureen, 1995, Market Microstructure Theory, Blackwell Publishers

[39] Roll, R. 1984, A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, Journal of Finance, 39, 1127-1139

[40] Rubin, I., (1972), Regular Point Processes and Their Detection, IEEE Transactions on Information Theory, ITT-18, 547-557

[41] Russell, Jeffrey, and R. Engle, 2002, Econometric Analysis of Discrete Valued, Irregulalry Spaced Financial Transactions Data, Working Paper University of Chicago, Graduate School of Business

[42] Rydberg, T.H. & N. Shephard 1999, Dynamics of trade-by-trade price movements: decomposition and models, Working Paper. Nuffield College, Oxford University

[43] Snyder, Donald, and Michael Miller, Random Point Processes in Time and Space, Springer Verlag

[44] White, H. 1982, Maximum Likelihood Estimation in Mispecified Models, Economtrica, 50, 1-25

[45] Zhang, M. J. Russell, and R. Tsay, 2000, Information Determinants of Bid and Ask Quotes: Implications for Market Liquidity and Volatility, Working Paper University of Chicago, Graduate School of Business

[46] Zhang, M. J. Russell, and R. Tsay, 2001, A Nonlinear Conditional Autoregressive Duration Model with Applications to Financial Transactions Data, Journal of Econometrics, 104, 179-207

Appendix

A EACD(3,3) parameter estimates using EVIEWS GARCH module.

	Coefficient	Robust Std. Err.
ω	0.004244	0.000855
α_1	0.070261	0.007157
α_2	0.038710	0.012901
α_3	-0.055966	0.008640
β_1	0.835806	0.125428
β_2	0.107894	0.118311

where $\psi_i = \omega + \sum_{j=1}^{3} \alpha_j x_{i-j} + \sum_{j=1}^{2} \beta_j \psi_{i-j}$

Model diagnostics

B VAR parameter estimates

	Price Equation		Trade Equation	
Variable	Coefficient	Std. Error	Coefficient	Std. Error
c	-0.006553	0.000284	0.509648	0.004785
w_i	0.014230	0.000430		
w_{i-1}	0.000891	0.000493	0.298146	0.005557
w_{i-2}	-0.000175	0.000493	0.059228	0.005797
w_{i-3}	-0.000533	0.000493	0.036385	0.005803
w_{i-4}	0.000176	0.000493	0.026645	0.005798
w_{i-5}	-0.001295	0.000425	0.035205	0.005558
Δm_{i-1}	-0.262310	0.005734	0.250909	0.071635
Δm_{i-2}	-0.121951	0.005934	0.108735	0.081696
Δm_{i-3}	-0.054038	0.005968	-0.000260	0.084009
Δm_{i-4}	-0.026460	0.005934	-0.022889	0.081695
Δm_{i-5}	-0.011011	0.005734	-0.220448	0.071634