Trades and Quotes: A Bivariate Point Process*

ROBERT F. ENGLE

Department of Economics, University of California, San Diego,
9500 Gilman Drive La Jolla, CA 92093-0508, USA

ASGER LUNDE

Department of Economics, Politics and Public Administration,
Aalborg University, Fibigerstraede 1, DK-9220 Aalborg Ø, Denmark
Email: alunde@socsci.auc.dk

This version: February 1999, First version: November 1997

Recent empirical work has studied point processes of transactions in financial markets and observed clear time dependent patterns in these arrival times. However, these studies do not examine the timing of quoted price changes. This paper formulates a bivariate point process to jointly analyze transaction and quote arrivals. In microstructure models, transactions may reveal private information that is then incorporated into new prices. This paper examines the speed of this information flow and the circumstances that govern it.

A joint likelihood function for trade and quote arrivals is specified which is a generalization of the ACD model of Engle and Russell (1998) and which recognizes that an intervening trade sometimes censors the time between a trade and the subsequent quote. Models of trades and quotes are estimated for 8 stocks using TAQ transaction data. A series of diagnostic tests and alternative estimates are presented.

The essential finding for the arrival of price quotes, is that information flow variables, such as high transaction arrival rates, large volume per trade, and wide bid/ask spreads, all predict more rapid price revisions. This means prices respond more quickly to trades when information is flowing so that the price impacts of trades and ultimately the volatility of prices, are high in such circumstances.

When quote arrivals that modify only depth are included as well, a somewhat different answer is found. Now the information variables such as volume and trade arrival rate and level of spread have a slowing or ambiguous effect on quote arrivals. This result is an indication that depth revisions are relatively more frequent when the market is slower. A possible explanation is that depth quotes are derived directly from the limit order book and limit orders are submitted more cautiously when the market shows informational trading.

Journal of Economic Literature Classification Numbers: C1, C2, C39, D41, G4.

*The authors would like to thank the National Science Foundation for support in grant SBR-9422575. The second author thanks the Center for Nonlinear Modelling in Economics, University of Aarhus. Valuable comments have been received from Ken Nyholm, seminar participants at University of Aarhus, the Center for Analytical Finance in Aarhus, University of California San Diego and Camp Econometrics IV at Catalina Island.

1. Introduction

Financial markets are designed to rapidly match buyers and sellers of assets at mutually agreeable prices. When this process is examined in detail, there are two types of events, which are observable at most financial exchanges. Traders buy and sell assets and specialists post quotes. Traders observe the posted prices and previous transactions to determine their strategies, and similarly, the specialists observe past transactions and prices to decide what quotes to post. Since transactions and quote revisions do not occur simultaneously, the times of each event presumably represent some optimization and potentially convey information. This paper analyses these two time scales and estimates a model that relates them.

The detailed records of financial transactions typically include two prices with different interpretations. A quote reflects one market participant's willingness to trade. It is firm only up to a given size and may be improved, both in terms of price and/or quantity. It may well reflect limit orders that are known to the specialist but often not to other participants. Transaction prices are agreed prices between counter parties; however, they do not reflect opportunities to trade. For example, a transaction which occurs at the ask price is likely to be between a buyer and the specialist or a limit order. This price is not available to a seller. Similarly, a transaction price for a small volume transaction is not generally available to a large transaction trader.

In analyzing the information content of a transaction, it is now common not only to examine the impact on prices of the direction of the trade, i.e. Hasbrouck (1996) and his many references, but also of the timing of the trade as in Easley and O'Hara (1992), Engle (1996), and Dufour and Engle (1997). The timing of the quote response however has not been examined. How long do specialists wait until they post new quotes? This is relevant in determining the speed of the price response to transactions and ultimately to the rapidity of information absorption and market clearing.

In this paper the arrival of trades and quotes is treated as a bivariate, dependent point process. The arrival of each type of event is influenced by the past history of both processes and also by information such as the bid/ask spread, volume of transactions and other predetermined variables. In the next section, the economic background is sketched, and then the statistical framework is presented. Section 4 presents the data and basic statistics while section 5 gives the results and discussion. Section 6 concludes.

2. Economic Motivation

Empirical findings of recent studies such as Engle (1996), Engle and Russell (1998) and Dufour and Engle (1997) are consistent with the predictions of the literature studying the *market microstructure* of financial markets (see (O'Hara 1995)).

In the early microstructure models time does not matter *per se*. In Kyle (1985) orders are batched together and cleared at predetermined points in time. Hence, the arrival times of individual orders are of no relevance to the market marker. The sequential trade framework suggested by Glosten and Milgrom (1985) has orders arriving according to some stochastic fashion independent of any time parameters. Thus the timing of trades is also irrelevant in this model. If, however, time can be correlated with any factor related to the asset price, then the rate of trade arrival conveys information to the agents. And as the agents learn from watching the flow of trades the adjustment of prices to information will also depend on time (O'Hara 1995 p. 168).

The notion of time was introduced into economic models by Diamond and Verrecchia (1987) and Easley and O'Hara (1992). Put very shortly the first model predicts that observing a low rate of trade arrival implies the presence of *bad news*. This result is derived from short sell constraints. Easley and O'Hara (1992) introduce event uncertainty into the sequential trade framework. The uncertainty is whether informed traders received a signal about the value of the asset. Their model implies that a low trading intensity means *no news*, because the informed traders only trade when they get a signal.

The empirical studies mentioned above seem to favor the Easley and O'Hara model. Dufour and Engle (1997) found that time durations are negatively correlated with the absolute value of the following quote revision, and that the spread is negatively correlated with lagged durations. Engle (1996) and Engle and Russell (1998) derived a relationship between arrival rates and volatility. Engle (1996) modelled both the arrival times of transactions and characteristics of these events sometimes called marks. He modelled time according to the ACD model and the marks are modelled conditional on the times. Thus the estimated expected durations are included in the volatility equation of an ARCH-type model. It was found that longer durations were associated with lower volatilities, and interpreted as no news reduces volatility; this supports Easley and O'Hara.

In an asymmetric information framework, the specialist quotes bid and ask prices to offset the expected losses from trading with informed traders. Once a trade has occurred, the specialist can reevaluate his quotes. If the trade was a buy, then there is a slightly increased probability that the information possessed by a fraction of the traders was positive for the asset. The specialist will increase both bid and ask prices at this time and possibly change the spread. The amount by which the specialist moves the quotes depend on the information he has from trades thus far and the assessment of the fraction of traders who are informed. The higher the fraction, the greater the response to the trade.

A central question in market microstructure is how fast and how completely, new information is incorporated into prices. A key but unnoticed part of this question is the timing of quote changes in response to transactions. The timing of the specialist's response is assumed to be immediate in models such as Glosten and Milgrom (1985). However, in Easley and O'Hara (1992), the calendar time between his revisions can change. If there is no information event, then trading will slow down and consequently the time between quote revisions will become longer. However there is still no delay from a trade to a quote revision. Only in the case where trades have no information, or where the discreteness of quotes is greater than the size of the desired revision, will there be a delay between transactions and revisions. Thus, there is a prediction that in a market with fewer information traders and slower transaction rates, the time to revise quotes should be longer.

A deeper analysis of the timing of quote setting must be tied to the supply of limit orders. Since on the NYSE, the specialist participates in a relatively small number of transactions, his quotes reflect the tightness of the limit order book. If limit buy orders are all above his asking quote; he may increase the quote to execute new market orders against the book. Similarly if there are limit buy orders within his spread, he may reduce the ask to again reflect the prices at which transactions can be executed. In this interpretation, quotes may change in the absence of transactions or other news simply because of changes in the order book. Similarly, transactions may not result in a quote change if the limit order book is unchanged.

A quote consists of four numbers, a bid and an asking price, and a bid and an asking quantity, called the quoted depth. The specialist guarantees to transact at least the quoted depth at the quoted prices. The specialist not only can post a quote to signify new prices but also new depth. While both reflect the response to trades and to the limit order book, the economic incentives to adjust the quotes may be different for prices and quantities. In the empirical work, quotes which alter the prices are called mid-quotes which is the average of the bid and ask prices[1], and are analyzed separately from all quotes.

3. Statistical Formulation

In transaction or quote datasets only one type of event can occur; namely a trade or the post of a quote, respectively. When combining trade and quote data a more complicated situation arises. Now two types of events will be occurring as time passes and the associated marks may be different variables for the two types. There are very few general accounts on multivariate point processes in the literature. A comprehensive treatment was given by Cox and Lewis (1972) from whom we adopted some of the terminology and notation.

Denote by $t_1, \ldots, t_{i-1}, t_i, \ldots$ the sequence of clock times at which a transaction of a given asset occurred, and by $q_1, \ldots, q_{i-1}, q_i, \ldots$ denote the timing of bid/ask quote revisions for this particular asset. A general bivariate model for these processes would involve associating a counting process $N(s_1, s_2) = \{(N^t(s_1), N^q(s_2)\}$ with the bivariate point process. Here $N^t(s_1)$ and $N^q(s_2)$ are the number of trades and quotes in $(0, s_1]$ and $(0, s_2]$ respectively. Further a bivariate sequence of intervals $\{T_i, Q_j\}$ is defined. Here T_i is the time between the $(i-1)$'th and the i'th trade; and the $\{Q_j\}$ sequence is defined similarly. This specification might suggest constructing a bivariate model from $\{(T_i, Q_i)\}_{i=1}^N$. This could be fruitful in some situations, but in general it is not a useful approach, because in the present case events in the two processes with a common serial number will be far apart in real time. This leaves the specification of the dependence between pairs of T_i and Q_i very tricky. Below it is seen how our model circumvents this problem of asynchronous starting points of duration pairs.

3.1. The Model

Let $t_1, \ldots, t_{i-1}, t_i, \ldots$ be defined as above. Using these arrival times we define the sequence $t_1^{'}, \ldots, t_{i-1}^{'}, t_i^{'}, \ldots$, with $t_i^{'}$ denoting the clock time of the first quote arriving after the transaction at t_{i-1}. Given these point processes two types of durations are defined. Both types start with a transaction occurring at time $t_{i-1.}$ Hence, define by $X_i = t_i - t_{i-1}$ the forward recurrence time to the next trade, and denote by $Y_i = t_i^{'} - t_{i-1}$ the forward recurrence time to the next quote. We call these random variables a *forward trade duration* and a *forward quote duration*

[1] There are no cases in our data where bid and ask movements offset to leave the mid-quote constant.

respectively. Together X_i and Y_i constitute a bivariate duration process which eliminates the synchronization problem mentioned above. The transaction times each initiate a waiting time for the next quote to occur. The structure does not comprise the assumption that the process of forward trade durations is independent of the timing of quotes or information associated with the quote, such as spreads. This is merely a possibility that can be tested. Giving the model this structure has economic intuition, as in most transaction datasets the transaction times are the only observable events that make the specialist reflect upon his current bid-price, ask-price or the depth at these prices. The series constructed in this manner now have the property that durations with a common serial number have the same time origin. While it is clearly possible to model the time from a quote to the following trade, this would be an entirely different model designed to answer questions different from the ones we are addressing.

It will often be the case for very frequently traded stocks that a new transaction occurs before the next quote, that is t_i might be less than $t_i^{'}$. The transaction conveys information that is likely to change our beliefs about when the next quote will occur and especially our beliefs about what will happen at the next quote. This means that at time t_i our expectation of the arrival of the next quote will change even though the initiated quote spell was not completed. To embed this feature in the model, cases of $t_i^{'} > t_i$ are treated as *censored* forward quote durations. This is done by defining

$$\widetilde{Y}_i = \min(Y_i, X_i) = \widetilde{t}_i - t_{i-1}, \qquad \text{where } \widetilde{t}_i = \min(t_i^{'}, t_i)$$

We call $\widetilde{Y}_i$ the *observed forward quote duration*, and associate with it an indicator d_i taking the value 1 if $\widetilde{Y}_i$ was censored. Note that in this case we only know that the i'th forward quote duration was longer than $\widetilde{Y}_i$.

The statistical model can now be build by specifying the parameterization of the bivariate duration process given by $\{(X_i, \widetilde{Y}_i)\}_{i=1}^N$. Assume that the i'th observation has a joint density conditional on all relevant and available information as of time t_{i-1}. Modelling this distribution directly would be a very complex matter, but fortunately a simpler expression can easily be obtained. Without loss of generality, the joint density can be written as the product of the conditional density and the resulting marginal density. Hence, we write this as

$$p(x_i, \widetilde{y}_i | \mathcal{H}_{i-1}; \boldsymbol{\omega}) = g(x_i | \mathcal{H}_{i-1}; \boldsymbol{\omega}_1) f(\widetilde{y}_i | x_i, \mathcal{H}_{i-1}; \boldsymbol{\omega}_2) \tag{1}$$

and call $g(\cdot|\mathcal{H}_{i-1}; \boldsymbol{\omega}_1)$ the trade density and $f(\cdot|x_i, \mathcal{H}_{i-1}; \boldsymbol{\omega}_2)$ the quote density.

Before we turn to the actual parameterization a few words about the model defined so far are required. The process for the forward trade duration is assumed to be of the *Autoregressive Conditional Duration* (ACD) type suggested by Engle and Russell (1998). However, the observed forward quote durations are censored and this must be modelled carefully. The process of censoring times is in fact the forward trade duration process, and hence the censoring times will be a dependent process. This is not a problem even though there will be clusters of short censoring time and periods with longer ones, and these periods may correspond to similar periods in quote intensities. Cox and Oakes (1984) require that, conditionally on the values of any explanatory variables, the prognosis for any forward quote duration not terminated at the censoring time should not be affected if it is censored. This is clearly not true in the model we propose. The reason is that the censoring thresholds are a dependent process, namely the transaction times, which may or may not depend on the other variables in the model. But this is exactly the feature the model is constructed to deal with. The model explicitly states how our prognosis about the next occurrence of a quote should be altered in the case of an intervening trade. Hence the introduction of this type of censoring is a way of updating our beliefs about the expected forward quote duration. It is this feature that gives the effect equivalent of time-dependent covariates in a Cox regression for inter quote arrival times. Unlike the Cox and Oakes model and competing risk models we observe the censoring threshold for each observation, and can model it directly.

We now return to the parametric specification of the model. In specifying the trade density let $\psi_i(\mathcal{H}_{i-1};\boldsymbol{\omega}_1) = E(X_i|\mathcal{H}_{i-1};\boldsymbol{\omega}_1)$, then

$$g(x_i|\mathcal{H}_{i-1};\boldsymbol{\omega}_1) = \psi_i(\mathcal{H}_{i-1};\boldsymbol{\omega}_1)^{-1}\exp\left\{\frac{-x_i}{\psi_i(\mathcal{H}_{i-1};\boldsymbol{\omega}_1)}\right\} \tag{2}$$

where the expected duration follows an exponential linked ACD-type model, which we call the *Nelson form*[2]. It is given by

$$\psi_i(\mathcal{H}_{i-1};\boldsymbol{\omega}_1) \equiv \psi_i = \exp\left\{\alpha + \delta\ln(\psi_{i-1}) + \gamma\frac{x_{i-1}}{\psi_{i-1}} + \boldsymbol{\beta}\boldsymbol{Z}_{i-1}\right\} \tag{3}$$

with $\boldsymbol{Z}_{i-1}$ being a vector of explanatory variables known at time t_{i-1}. Equation (3) will be referred to as the trade equation. Note that ψ^{-1}, the inverse of the expected duration, is the trading intensity. It gives the instantaneous rate at which trades arrive.

The quote density takes into account that some of the observations are censored. This is done in the usual way for models with censored observations. Thus we have

$$f(\widetilde{y}_i|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2) = h_Y(\widetilde{y}_i|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)^{1-d_i}S_Y(\widetilde{y}_i|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)^{d_i} \tag{4}$$

where h_Y and S_Y are the density function and the survivor function [3] respectively for the forward quote duration. $h_Y(\cdot|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)$ is the actual forward quote density and will be termed so. The density is similar to the density for the forward trade durations, except for the important feature that it is conditional on x_i. Let $\varphi_i(x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2) = E(Y_i|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)$, then

$$h_Y(\widetilde{y}_i|x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2) = \varphi_i(x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)^{-1}\exp\left\{\frac{-\widetilde{y}_i}{\varphi_i(x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_2)}\right\} \tag{5}$$

Again the expected duration follows an exponential ACD-type model, hence the quote equation is given by

$$\varphi_i(x_i,\mathcal{H}_{i-1};\boldsymbol{\omega}_1) \equiv \varphi_i = \exp\left\{\mu + \rho\ln(\varphi_{i-1}) + \delta_1\frac{\widetilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2\frac{\widetilde{y}_{i-1}}{\varphi_{i-1}}d_{i-1} + \tau\frac{x_i}{\psi_i} + \boldsymbol{\eta}\boldsymbol{V}_{i-1}\right\} \tag{6}$$

---

[2]The name is due to the resemblance with the EGARCH of Nelson (1991) . The main reason for choosing this form is it's computational convenience, as it implies that the parameter space is unrestricted. This contrasts with the ACD model, which requires parameter restrictions to keep the expected duration positive.

[3]The survivor function, $S(t)$, is one minus the cumulative density function, that is the probability that the duration will last longer than t.

where $\boldsymbol{V}_{i-1}$ typically contains some of the variables of $\boldsymbol{Z}_{i-1}$. Here φ^{-1} is the quoting intensity, that is the rate at which the specialist posts his quotes. Note that the inclusion of $\frac{\widetilde{y}_{i-1}}{\varphi_{i-1}} d_{i-1}$ allows us to asses the impact of having some observations censored.

The definitions given above imply that the expected forward trade duration $\{\ln(\psi_i)\}$ and the expected forward quote duration, $\{\ln(\varphi_i)\}$, both follow ARMAX type processes, with the usual ARMAX stationarity conditions. Specifically, $\{\ln(\psi_i)\}$ and $\{\ln(\varphi_i)\}$ are stationary if δ and ρ are strictly less than one respectively.

3.2. Estimation and Inference

Under the specification (1), the log likelihood can be expressed as:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\omega}; \boldsymbol{X}, \widetilde{\boldsymbol{Y}}) &= \sum_{i=1}^{N} \left[\ln g(x_i | \mathcal{H}_{i-1}; \boldsymbol{\omega}_1) + \ln f(\widetilde{y}_i | x_i, \mathcal{H}_{i-1}; \boldsymbol{\omega}_2) \right] \\ &= \sum_{i=1}^{N} l_i^g(\boldsymbol{\omega}_1) + \sum_{i=1}^{N} l_i^f(\boldsymbol{\omega}_2) \end{aligned} \tag{7}$$

which has to be maximized with respect to the parameters $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$. We take the approach of maximizing $\sum_{i=1}^{N} l_i^g(\boldsymbol{\omega}_1)$ first and then conditional on this $\sum_{i=1}^{N} l_i^f(\boldsymbol{\omega}_2)$ is maximized. This reduces the estimation time considerably and lowers the requirement to the computational device, which is very important considering the huge datasets analyzed in this paper. This two-step approach is not equivalent to maximum likelihood as possible constraints are ignored, which may result in less efficient estimates. Still, our two-step approach is pretty close to maximum likelihood. We will substantiate this claim by estimating the parameters jointly as well as using the two-step approach.

The estimation approach is semiparametric, in that we do not assume that the true densities of g and h are exponential as stated in (2) and (5)[4]. The log likelihood function is called a quasi-likelihood function. This method only requires specifying the mean of the distribution. Then QML estimators can be obtained which are consistent for $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ and have a well-defined asymptotic covariance matrix. QML methods were introduced into econometrics by White (1982) and the results that justify the present application are analogous to Bollerslev and Wooldridge (1992), as shown by Engle and Russell (1998). The robust covariance matrix for $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ is calculated as:

$$\left[\sum_{i=1}^{N} \frac{\partial^2 l_i^s}{\partial \boldsymbol{\omega}_j \partial \boldsymbol{\omega}_j'} (\widehat{\boldsymbol{\omega}}_j) \right]^{-1} \left[\sum_{i=1}^{N} \frac{\partial l_i^s}{\partial \boldsymbol{\omega}_j} (\widehat{\boldsymbol{\omega}}_j) \frac{\partial l_i^s}{\partial \boldsymbol{\omega}_j} (\widehat{\boldsymbol{\omega}}_j)' \right] \left[\sum_{i=1}^{N} \frac{\partial^2 l_i^s}{\partial \boldsymbol{\omega}_j \partial {\boldsymbol{\omega}_j}'} (\widehat{\boldsymbol{\omega}}_j) \right]^{-1}_{(s,j)=(g,1),(f,2)} \tag{8}$$

Hence the two-step estimation procedure assumes that $\boldsymbol{\omega}$ is block diagonal in $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$.

It is important to note that it is only for the trade equation that the QMLE's of the parameters are consistent for the mean, as this part satisfies the requirement that the expected score must

[4]It is clearly possible to use a general density such as the generalized gamma distribution as suggested by Lunde (1998). But we have no economic theory suggesting what the shapes of these densities should be. And it is not a question we are going to address in this paper.

equal zero. For the quote equation these estimates are, in general, only consistent for the mean if the true duration density is exponential. Otherwise, it is not true that the expected score is equal to zero. However if this assumption is violated the QMLE's will still be consistent in the sense of White, that is, consistent for whatever they are estimating. It is easy to test the fit of quote durations, and misspecification may be resolved by using a more general distribution as suggested in Lunde (1998). Still this is beyond of the aim of this paper as the model is quite complex already.

3.3. Residuals and Specification tests

The residual analysis assesses the validity of the exponential duration distributions used in the QML approach, and the amount of remaining autocorrelation not explained by the specified model. Hence, under the null that the model is truly exponential and that the expected duration is correctly specified, the residuals should be identical and independent unit exponentially distributed.

Generally residuals with a unit exponential distribution are defined as follows. If T_i has survivor function $S(t|\mathcal{H}_{i-1};\boldsymbol{\omega})$ then $S(T_i|\mathcal{H}_{i-1};\boldsymbol{\omega})$ is uniformly distributed and $-\ln(S(t|\mathcal{H}_{i-1};\boldsymbol{\omega}))$ has a unit exponential distribution. Thus, for the trade part we define the residual to be

$$\begin{aligned}\xi_i &= -\ln(S(x_i|\mathcal{H}_{i-1};\widehat{\boldsymbol{\omega}}_1)) \\ &= x_i\psi_i(\mathcal{H}_{i-1};\widehat{\boldsymbol{\omega}}_1)^{-1} \qquad (9)\end{aligned}$$

which is identical to the residual defined in Engle and Russell (1998). These residuals are often called Cox-Snell residuals as they are derived from the general definition of residuals given by Cox and Snell (1968) [5].

If the i'th individual is censored, so too is the corresponding residual and thus in general we obtain a set of uncensored and a set of censored residuals which cannot be regarded on the same footing. We may therefore seek to modify the Cox-Snell residuals taking explicit account of the censoring. Suppose that $\widetilde{y}_i$ is censored. The Cox-Snell residual for this observation is then given by

$$r_i = \widetilde{y}_i\varphi_i(x_i,\mathcal{H}_{i-1};\widehat{\boldsymbol{\omega}}_2)^{-1}$$

If the fitted model is correct, then the values r_i can be taken to have a unit exponential distribution. The cumulative hazard function of this distribution increases linearly with time, hence, the greater the value of the duration, the greater the value of that residual. It hereby follows that the residual for this duration at the actual unobserved termination time will be greater than the residual evaluated at the observed censoring time. To account for this the Cox-Snell residual can be modified by adding a positive constant, called the excess residual. Using the lack of memory property of the exponential distribution, we know that because r_i has a unit exponential distribution the excess residual will also have a unit exponential distribution. The expected value of the excess residual is therefore one. This suggests defining the residuals for the quote part to be

$$\varepsilon_i = \widetilde{y}_i\varphi_i(x_i,\mathcal{H}_{i-1};\widehat{\boldsymbol{\omega}}_2)^{-1} + d_i$$

[5] A very readable exposition of residuals for survival models may be found in Collett (1994).

A number of LM tests defined as given in Wooldridge (1994) section 4.6 are presented. The same type of comments on the effect of the censoring in the quote part, as given in the end of the previous section applies here. We used the heteroskedasticity consistent covariance matrix (8) when computing these tests.

4. Data Description

The data are extracted from the Trade and Quote (TAQ) database. The TAQ database is a collection of all trades and quotes in New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotation (Nasdaq) securities. We only consider trades and quotes on the NYSE. Schwartz (1993) and Hasbrouck, Sofianos, and Sosebee (1993) document NYSE trading and quoting procedures.

Among the fifty stocks with the highest capitalization value on December 13 1996 eight stocks were randomly selected. The names and some summary statistics are given in Tables 1 and 2. Trades reported within the same second were treated as one trade, with the volumes of the multiple trades aggregated. For the quotes we also filtered out multiple occurrences of quotes. The sample period is the two months from August 4, 1997 to September 30, 1997, which gives a total of 42 trading days. All stocks in the period had more quotes than trades. By comparing the last two columns of Table 1 it appears less than half of the quotes comprised revisions (changes) of the mid-quote[6], the rest of the revisions being pure depth revisions.

In Table 2 it can be seen that the average spread when mid-quotes are revised is substantially higher than for all quotes. Consequently, depth only revisions occur predominantly when spreads are low. In Table 3, mid-quote durations are systematically negatively correlated with lagged volume, measured as shares transacted on the previous trade, and negatively correlated with the lagged spread. All quote durations are however systematically positively correlated with the spread and have a mixed positive relation with lagged volume. So spread and volume have essentially opposite effects on the two types of quote arrivals. Both measures are positively correlated with trade durations and with the change in spreads. Thus a rise in the spread predicts a delay in the next quote of either type. However spreads change on just a few percent of the observations so this is a very transient effect.

The different behavior of the price and quantity quotes will become more obvious in the empirical section of the paper and have an intuitively appealing explanation. We associate pure depth revisions with the movements in the limit order book, and hence the speed of such depth revisions is expected to be dictated by the behavior of the liquidity providers in the market. In contrast, the pure mid-quote revisions are presumed to be caused by the specialist's ever-changing perceptions about the value of the stock and especially his fear of facing informed agents. Thus it is the pure mid-quote model that we should expect to fit the predictions of asymmetric information models.

To construct the bivariate duration process of forward trade and quote durations we begin

[6]It is of no importance whether one consider mid-quote revisions or spread revisions as these coincide. In the data considered, the specialist almost exclusively changed the bid or the ask, and very rarely both at the same revision.

as outlined in section 2 by calculating the forward trade durations simply as the inter-trade arrival times. The first duration every day is the duration from the second to the third trade that day. Thus, the NYSE opening and the high volume associated with this are excluded from the analysis. This is important because the opening trade is fundamentally different from all other trades during the trading day. The procedure is that the market for a stock opens when the specialist finds a price that balances the buy- and sell-orders for the issue. The specialist does this by matching market orders that come in through the Opening Automated Report Service (OARS), a feature which accepts preopening market orders, public limit orders and non-OARS-eligible market orders that come into the electronic display book or printers, and orders from the trading crowd. In this way the NYSE opening resembles a call market where orders are batched together for execution at a predetermined time at a single price, in contrast with the rest of the day which is a continuous market where market orders are executed immediately upon submission. This motivates why the opening requires special attention and modelling, a task, which is beyond the scope of this paper.

Overnight durations were also omitted from the sample. Now for every transaction the prevailing quote is the most recent quote that occurred at least five seconds before the transaction. As on the NYSE floor posting new quotes is given priority over recording completed transactions, a quote revision will often precede the trade from which it was instigated. Hence, to compute the forward quote durations we delay every quote time five seconds and then the forward quote duration is the time from the present trade to the next quote. Matching transactions with quotes in this way overcome the concern for mis-timed recordings. This *five second* rule was suggested by Lee and Ready (1991). To build the observed forward quote durations every pair of forward trade and quote durations are compared. If the forward quote is longer than the forward trade duration, the observed forward quote duration is censored.

With the dependent process defined we need to specify the explanatory variables to be put into $\boldsymbol{Z}_{i-1}$ and $\boldsymbol{V}_{i-1}$. There are surely a lot of possibilities. More lagged values of the dependent process, the time since the most recent quote, the spread, the volume etc. Of course it is preferable to include variables that have economic interpretations. There is a huge amount of literature on the importance of information contained in spreads and trading volume which is nicely reviewed in Hasbrouck (1996) and Goodhart and O'Hara (1997). The selection is mainly derived from these former studies and the motivation and interpretation of the included variables will become clear when we discuss the results of the estimation. Table 4 presents and explains the computation of the explanatory variables associated with the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ of equations (3) and (6). It is important to note that variables are lagged with respect to the trade time. Hence, the third lagged spread would be the prevailing spread three trades ago.

Trades may be classified into buys and sells using the technique also used by Lee and Ready (1991). Trades at prices above the mid-quote are associated with buys (initiated by a buyer) and are marked 1; trades below the mid-quote are called sells (initiated by a seller) and given the mark -1 . This variable is often referred to as a buy-sell trade indicator variable. The rational for this classification is that trades originating from buyers are most likely to be executed at or near the ask, while sell orders trade at or near the bid. This scheme classifies all trades except those that occur at the mid-quote. We do not apply the *tick* rule; trades at the mid-quote are given a zero mark. On average 15.9% of the volumes traded are given a zero mark. Using these sign marks we compute the accumulated signed volume. This is calculated use a moving window of ten trades. We include the absolute value of this variable as an explanatory variable. This variable is a measure of the imbalance of trades and related to the depth measure *VNET* introduced by Engle and Lange (1997). Giving the sizes corresponding to trades executed at the mid-quote zero weight excludes these sizes from the constructed variable. As the trades at the mid-quote often correspond to crossed orders these do not contribute to imbalance of the specialists inventory, because he is not trading on his own account. We find the tick rule arbitrary and using our method will deliver a conservative estimate of the imbalance effect.

Typically the market exhibits high activity in the morning and before closure. Around lunchtime the activity is mostly lower. The daily pattern in the dependent variable must be reflected in patterns in the independent variables. It is unlikely that the diurnal pattern of trades is due to similar movements in spreads, or quotes and thus the autoregressive nature of the model is left to explain these shapes. Either time of day exogenous variables are needed or the data must be filtered to eliminate these effects from all series. For the trade equation this has been done by estimating $E\left[\cdot_i | t_{i-1}\right]$ for every second of the day using the Splus routing called smooth.spline. This one-dimensional cubic smoothing spline uses a basis of B-splines as discussed in chapters 1,2 & 3 of Green and Silverman (1994). The estimated splines for trade-trade and quote-quote duration all have the characteristic inverse U-shaped form found in similar studies. The same methodology was used to filter the spread and volume variables, all of which are shown in Figure 1 and referred to in the equations with a check; " $\check{\cdot}$ ".

For the quote equations we do not filter out the diurnal effects but assume that the daily periodicity in quote arrivals is a natural economic result of the periodicity in the trade arrivals, and other exogenous and predetermined variables. This is a particularly convenient assumption since the censoring of quotes by intervening trades could be seriously distorted by attempting to adjust each separately. Since the conditional duration estimated in the trade equation has been purged of its diurnal effect before estimation, the total estimate of the expected duration must be multiplied by the daily periodic effect before introducing it into the quote equation. Hence the following estimate of the expected trade duration is computed:

$$\widehat{\psi}_i = \check{\psi}_i(\widehat{\boldsymbol{\omega}}_1) E\left[X_i | t_{i-1}\right]$$

Thus the diurnal pattern of the trades is naturally allowed to influence the diurnal pattern of quotes. As most other variables also have daily periodicities, trades are not the only source of diurnal movements. One of the diagnostic tests and further estimates examine the validity of these assumptions.

The estimated functional forms are given as follows[7]:

$$\begin{aligned}\ln(\check{\psi}_i) &= \alpha + \delta \ln(\check{\psi}_{i-1}) + \gamma \frac{\check{x}_{i-1}}{\tilde{\psi}_{i-1}} + \beta_1 \mathsf{lev.Q\check{Q}}_{i-1} + \beta_2 \Delta \mathsf{S\check{p}r}_{i-1} + \beta_3 \mathsf{lev.S\check{p}r}_{i-1} \\ &+ \beta_4 \sqrt{\mathsf{v\check{o}l}}_{i-1} + \beta_5 \mathsf{Abs(s.v\check{o}l)}_{i-1} + \beta_6 \mathsf{Ba\check{c}k.Q}_{i-1}\end{aligned}$$

[7] All variables are in deviations from their mean value.

and

$$
\begin{aligned}
\ln(\varphi_i) \;=\;& \mu + \rho\ln(\varphi_{i-1}) + \delta_1\frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2\frac{\tilde{y}_{i-1}}{\varphi_{i-1}}d_{i-1} + \tau\frac{x_i}{\widehat{\psi}_i} + \eta_1\frac{x_{i-1}}{\widehat{\psi}_{i-1}} + \eta_2\ln(\widehat{\psi}_{i-1}) \\
+\;& \eta_3\mathsf{lev.QQ}_{i-1} + \eta_4\Delta\mathsf{Spr}_{i-1} + \eta_5\mathsf{lev.Spr}_{i-1} + \eta_6\sqrt{\mathsf{vol}}_{i-1} \\
+\;& \eta_7\mathsf{Abs(s.vol)}_{i-1} + \eta_8\mathsf{Back.Q}_{i-1}
\end{aligned}
$$

see Table 4 to refresh the definition of the explanatory variables. Table 3 gives some simple correlations of the dependent variable and the explanatory variables in the quote equations.

5. Estimation and Results

Maximization of the log likelihood as outlined in section 2.2 was performed in C++ using the simplex method as found in Press, Teukolsky, Vetterling, and Flannary (1992). We often formulate the discussion in terms of intensities, which are the reciprocals of the expected durations, since these may have particularly clear interpretations. In the next two sections, the model estimated using the mid-quote arrivals, is presented and discussed. Then the all quote model is presented and results are contrasted. Finally, several alternative specifications are introduced to test the robustness of the models to time of day effects and to the two step estimation approach.

5.1. The Trade Equation

The estimates of the trade equation are presented in Table 5. We present the equation, stressing the signs of the coefficients and boldface the generally significant ones:

$$
\begin{aligned}
\ln(\check{\psi}_i) \;=\;& \overset{-}{\boldsymbol{\alpha}} + \overset{+}{\boldsymbol{\delta}}\ln(\check{\psi}_{i-1}) + \overset{+}{\boldsymbol{\gamma}}\frac{\check{x}_{i-1}}{\check{\psi}_{i-1}} + \overset{+}{\beta}_1\mathsf{lev.Q\check{Q}}_{i-1} + \overset{+}{\boldsymbol{\beta}}_2\Delta\mathsf{S\check{p}r}_{i-1} \\
+\;& \overset{-}{\beta}_3\mathsf{lev.S\check{p}r}_{i-1} + \overset{-}{\boldsymbol{\beta}}_4\sqrt{\mathsf{v\check{o}l}}_{i-1} + \overset{+}{\beta}_5\mathsf{Abs(s.v\check{o}l)}_{i-1} + \overset{-}{\boldsymbol{\beta}}_6\mathsf{Ba\check{c}k.Q}_{i-1}
\end{aligned}
$$

The trading intensity shows a very high degree of persistence, with δ bigger than 0.95 for most stocks. All stocks have γ, the coefficient on the surprise term positive and significant. Hence, these two parameters are as expected from earlier studies.

The most conspicuous of the explanatory variables is the square root of the volume. The systematically and significantly negative coefficients reveal that large trades initiate shorter durations than small trades so that the arrival intensity rises after a big trade. Presumably this is because large trades are more likely to be information based and therefore signal the presence of information trading as for example in Easley and O'Hara (1987). Since informed traders will be in a hurry to exploit deteriorating informational advantages, information flow leads to fast trading.

A high bid ask spread is another sign of informational traders and it has a predominantly negative but not significant sign in the level as would be expected. However it has a positive and generally significant coefficient in the change. Thus rising spreads initially discourage trades but once the spread has risen permanently to a new higher level, trades can increase again.

Lagged quote durations have a mixed positive impact on trade durations and the time since the last quote has a negative and systematic effect on trade arrivals. When quotes have not been revised for a long time, transaction intensities rise. When demand and supply are out of balance, measured by $\mathsf{Abs}(\mathsf{s.}\check{\mathsf{vol}})_{i-1}$, there is a positive but not significant effect on the duration to the next trade.

The diagnostic tests in the bottom of the table are surprisingly well behaved. The Ljung Box tests for autocorrelation in durations are dramatically reduced and are potentially clean depending upon the appropriate p value for such large sample sizes. Further, only a first order model is estimated and probably higher order models would be even more satisfactory from this point of view. Neither the LM test for another lagged quote duration or for time of day dummies proved significant.

Some excess dispersion is still present as the standard deviation of the residuals exceeds 1. To assess the significance of this excess dispersion we apply a simple test suggested by Engle and Russell (1998). The null of no excess dispersion is based on the statistics $\sqrt{N}(\frac{\widehat{\sigma}_{\xi}^{2}-1}{\sigma_{\nu}})$, where $\widehat{\sigma}_{\xi}^{2}$ is the sample variance of $\widehat{\xi}$, which should be 1 under the null hypothesis. σ_{ν} is the standard deviation of $(\xi-1)^2$ which equals $\sqrt{8}$ under the under the null of a unit exponential distribution. This statistic has a limiting normal distribution under the null with a 5% critical value of 1.645. Performing this test on our samples reveals that excess dispersion is still left in the residuals. In Lunde (1998) a generalized gamma version of trade equation is estimated. This model is able to remove the excess dispersion completely. But our model is complex enough as it is, and our experiments showed that it did not change any of the conclusions.

Overall the trade equation is consistent with our economic models and quite stable across stocks. It reveals the importance of volume and spreads in predicting transaction intensities, and finds that quote revisions are less important.

5.2. The Quote Equation

Table 6 reports the estimates of the quote equation estimated for mid-quote arrivals. The dependent variable can be interpreted as the time from a trade to a mid-quote revision taking account of the censoring effect of a trade that arrives before the next quote. The equation is reproduced here with the typical sign;

$$\ln(\varphi_i) = \overset{-}{\boldsymbol{\mu}} + \overset{+}{\boldsymbol{\rho}} \ln(\varphi_{i-1}) + \overset{+}{\boldsymbol{\delta}}_1 \frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \overset{-}{\delta}_2 \frac{\tilde{y}_{i-1}}{\varphi_{i-1}} d_{i-1} + \overset{+}{\boldsymbol{\tau}} \frac{x_i}{\widehat{\psi}_i} + \overset{-}{\boldsymbol{\eta}}_1 \frac{x_{i-1}}{\widehat{\psi}_{i-1}}$$

$$+ \overset{+}{\boldsymbol{\eta}}_2 \ln(\widehat{\psi}_{i-1}) + \overset{+}{\boldsymbol{\eta}}_3 \mathsf{lev.QQ}_{i-1} + \overset{+}{\boldsymbol{\eta}}_4 \Delta\mathsf{Spr}_{i-1} + \overset{-}{\boldsymbol{\eta}}_5 \mathsf{lev.Spr}_{i-1}$$

$$+ \overset{-}{\boldsymbol{\eta}}_6 \sqrt{\mathsf{vol}}_{i-1} + \overset{+}{\boldsymbol{\eta}}_7 \mathsf{Abs}(\mathsf{s.}\check{\mathsf{vol}})_{i-1} + \overset{+}{\boldsymbol{\eta}}_8 \mathsf{Back.Q}_{i-1}$$

the variable definitions are given in Table 4. The conditional intensity is again persistent but much less so than trades at least as judged by the autoregressive coefficient. The lagged standardized forward quote duration is very significant but there is no evidence that censored durations need to be treated differently from uncensored durations.

Trade timing is very important in explaining quote timing. The current standardized forward trade duration is probably the most significant variable, which is not particularly surprising since it is also the censoring threshold. Although the lagged value is negative, the total effect of lagged trade durations is also positive. Recognizing that this is a distributed lag relation between quote durations and standardized trade durations, the lag coefficients are readily calculated. The impact coefficient is simply τ, the first lag effect is $\rho\tau + \eta_1$ which is positive for all stocks, and the long run effect is $(\tau + \eta_1) / (1 - \rho)$which is also clearly positive for all stocks. Thus short trade durations predict short quote durations. Similarly, because is η_2 positive, short expected trade durations also lead to short quote durations. In contrast with the trade equation, the level of quote durations is also very positively significant.

The spread variables enter in a very interesting fashion. Whenever the spread has recently increased, quotes are more slowly adjusted but when the level of the spread is high, quotes are more rapidly adjusted. Conversely, recent decreases in the spread lead to fast adjustment but permanent decreases lead to slow adjustment. Presumably, the specialist prefers a wide spread in order to profit from his monopoly position. Hence he may choose to leave the spread wide as long as possible until either demand falls or competition from the limit order book rises to erode this position. We see that a temporary rise in spread slows trades but we don't see the response of the limit order book. Lange (1997) shows that limit orders arrive quickly when spreads temporarily widen but slow down for permanently wide spreads.

The second effect reveals that permanent high spreads are associated with fast price adjustment. This is consistent with both theoretical and empirical models. For example, the Easley and O'Hara (1992) model as discussed by Dufour and Engle (1997) finds that when the proportion of informed traders is high, then the spread will be high and the price impact of a trade will be high. This will appear as large price revisions following trades when spreads are high. Since only price revisions that exceed the tick size will appear as new quotes, this matches the empirical finding. Similarly, using different empirical methods, Dufour and Engle (1997) and Engle and Lange (1997) show that market liquidity whether measured by price impact or depth, is generally reduced when markets are faster. Engle (1996) finds that price volatility is greater when time between trades is shorter and when spreads are wider.

The strong negative volume effect is again consistent with both theoretical and empirical asymmetric information models. Large volume is typically a signal of informational trading and consequently should imply high price impacts and price volatility which in this continuous time framework means short quote durations.

The final two variables, the buy-sell imbalance and time since last quote, are both very significantly positive, although there appear to be theoretical arguments to support either sign. Trade imbalance can be interpreted as evidence that liquidity suppliers are willing to tolerate one-sided demands and continue to take the other side. Hence it is an indicator that asymmetric information is not a current fear and consequently price changes can be postponed. Similarly, long times since prices were revised can be interpreted as evidence that there is little information flow and therefore even longer durations can be expected. This is an implication of a declining hazard function which has frequently been found in this literature, see e.g. Engle and Russell (1998). Of course, both of these variables could be interpreted as evidence that prices need to be changed. In the first case, there is presumed to be some latent variable that we might call liquidity, which is only measured by its consequences such as the spread or order imbalance. Therefore, one consequence appears to predict another.

In Table 7, several diagnostic tests are presented for the quote equation. The extent of autocorrelation in quote durations in the first column is quite extraordinary; even though the second column reveals that the final model does not pass at a 1% level, the improvement is dramatic. The excess dispersion test is passed and the lagged quote test is passed by all but two stocks. The time of day test reveals evidence for 5 of the stocks that the daily periodicity is not adequately represented. This is remedied below.

The quote equation nevertheless tells a very consistent story. Evidence of information flows, whether it is past short quote durations, or current and past high trade intensities, or high levels of spread or volume, or low levels of market liquidity, all lead to short quote revision times. This is the underlying model of price dynamics in response to trades trading and news. This gives us a microstructure view of price volatility and its correlates.

5.3. The Quote Equation with All Quotes

The specification of this equation is identical to the Mid-Quote equation although there are approximately twice as many observations. The main differences, which will be found from examining Table 8, are three variables with changed signs. These variables are the information variables; many of the economic implications are quite different. Short expected trade durations now forecast long quote durations, high spreads predict long quote durations and high volume has an ambiguous effect. Thus the central indicators of information trading no longer indicate that rapid flow leads to faster quote revision. Since these quotes are not price changes however this does not mean that the implication for price volatility or price impacts has changed.

Instead, the delay in quote setting during information flow is a delay in depth adjustment. In order to understand the difference, it is useful to conceptualize the specialist as merely recording the depth measures, if not the price quotes, from the limit order book. As this limit order book is constantly changing, he constantly revises his quotes, but about half of these are simply depth revisions. This interpretation indicates how there can be many more quotes than trades; the limit order book is the additional source of information. Under this interpretation of specialist behavior, then the findings of slow response when information is flowing are findings that the limit orders respond slowly when information is flowing. Since the costs of submitting a limit order rather than a market order are great when there are informed traders able to pick-off orders, it is natural to be more cautious when such indicators are present.

The empirical results for the All Quote equation can thus be interpreted as showing that the specialist revises his depth more slowly when information trading is occurring, and this is probably because limit orders are supplied more cautiously in such circumstances.

5.4. Joint Estimation and Time-of-day Dummies

As mentioned earlier the quote equation was not filtered to remove the time-of-day features. It was trusted that the seasonal in the explanatory variables would take care of that in the observed forward quote durations. This was not the case for all stocks and the worst results were obtained for the mid-quote model, which only had three of the LM2 statistics insignificant in Table 7. In Table 10 the estimates of the mid-quote models with dummies are reported and these are very similar to those of Table 6, except for the coefficient on $\ln(\widehat{\psi}_{i-1})$ which has seven positive and significant compared to four in Table 6. This just reinforces the conclusion of the previous section. There are no interesting changes to the diagnostics, which are left out.

To be cautious in the inference about the estimated coefficients, a reduced model leaving out $\mathsf{lev.QQ}_{i-1}$, $\mathsf{Abs(s.\check{vol})}_{i-1}$ and $\mathsf{Back.Q}_{i-1}$ of both the trade and the quote equation was estimated jointly. Joint estimation means that the full likelihood

$$\mathcal{L}(\boldsymbol{\omega}; \boldsymbol{X}, \widetilde{\boldsymbol{Y}}) = \sum_{i=1}^{N} \left[\ln g(x_i|\mathcal{H}_{i-1}; \boldsymbol{\omega}_1) + \ln f(\widetilde{y}_i|x_i, \mathcal{H}_{i-1}; \boldsymbol{\omega}_2)\right]$$

is maximized in one step. The results are compared with the two-step method and reported in Table 11 and 12 for the mid-quote model. It is clear that the coefficients and their standard errors have changed slightly in all cases, but the changes are so small that they do not alter any of our conclusions. Further it seems to be safe to use the less computational and time demanding two-step approach.

6. Conclusion

This paper has developed a bivariate model of the arrival of trades and quotes for stocks traded on the NYSE. The time between a trade and a new price quote is argued to be an interesting measure of the speed at which information is incorporated into prices. It is also related to the price impact of the trade, the liquidity of the stock, and ultimately, the volatility of the stock. The model seeks to determine under what circumstances price adjustment is rapid and when it is slow.

Since trade-quote durations are not observed when there is an intervening trade, the bivariate model must allow censoring of the quote durations by the trade durations. A bivariate likelihood function is formulated with censoring of one variable by the other. This likelihood function is a generalization of the ACD model of Engle and Russell (1998). It is maximized under a variety of specifications for a trade equation and for a quote equation conditional on the trade duration. Specification tests and robustness checks are carried out.

The essential finding for the arrival of quotes that revise prices, is that information flow variables, such as high transaction arrival rates, large volume per trade, and wide bid/ask spreads, all predict more rapid price revisions. This means prices respond more quickly to trades when information is flowing so that the price impacts of trades and ultimately the volatility of prices, are high in such circumstances.

When quote arrivals that modify only depth are included as well, a somewhat different answer is found. Now the information variables such as volume and trade arrival rate and level of spread have a slowing or ambiguous effect on quote arrivals. This result is an indication that depth revisions are relatively more frequent when the market is slower. A possible explanation is that depth quotes are derived directly from the limit order book and limit orders are submitted more cautiously when the market shows informational trading.

References

Bollerslev, T., and J. M. Wooldridge, 1992, "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariance," *Econometric Reviews*, 11, 143–172.

Collett, D., 1994, *Modelling Survival Data in Medical Research*. Chapman & Hall.

Cox, C. R., and E. J. Snell, 1968, "A General Definition of Residuals," *Journal of the Royal Statistical Society* **B**, 30, 248–75.

Cox, D. R., and P. A. W. Lewis, 1972, "Multivariate Point Processes," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. M. Le Cam, J. Neyman, and E. L. Scott. University of California Press, Berkeley and Los Angeles, vol. III, pp. 401–448, volume III.

Cox, D. R., and D. Oakes, 1984, *Analysis of Survival Data*. Chapman & Hall.

Diamond, D. W., and R. E. Verrecchia, 1987, "Constraints on short-selling and asset price adjustments to private information," *Journal of Financial Economics*, 18, 277–311.

Dufour, A., and R. F. Engle, 1997, "Time and the impact of a trade," Department of Economics, University of California, San Diego.

Easley, D., and M. O'Hara, 1987, "Price, trade size, and information in secturities markets," *Journal of Financial Economics*, 16, 69–90.

——— , 1992, "Time and the process of security price adjustment," *Journal of Finance*, 47, 905–927.

Engle, R. F., 1996, "The econometrics of ultra-high frequency data," Department of Economics, University of California, San Diego Working paper 96-15, committed to appear in Econometrica.

Engle, R. F., and J. Lange, 1997, "Measuring, Forecasting and Explaining Time Varying Liquidity in the Stock Market," Discussion paper 97-12R, Department of Economics, University of California, San Diego.

Engle, R. F., and J. Russell, 1998, "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66(5), 1127–1163.

Glosten, L., and P. Milgrom, 1985, "Bid, ask, and transactions prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 13, 71–100.

Goodhart, C. A. E., and M. O'Hara, 1997, "High frequency data in financial markets: Issues and applications," *Journal of Empirical Finance*, 4, 73–114.

Green, P. J., and B. W. Silverman, 1994, *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall.

Hasbrouck, J., 1996, "Modelling market microstructure time series," in *Handbook of Statistics, Vol. 14*, ed. by G. S. Maddala, and C. R. Rao. North-Holland, pp. 647–692.

Hasbrouck, J., G. Sofianos, and D. Sosebee, 1993, "Orders, Trades, Reports and Quotes at the New York Stock Exchange," working paper, NYSE, Research and Planning Section.

Kyle, A. S., 1985, "Continuous auctions and insider trading," *Econometrica*, 53(3), 1315–1335.

Lee, C. M., and M. J. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733–746.

Lunde, A., 1998, "A Generalized Gamma Autoregressive Conditional Duration Model," Working Paper, Dept. of Economics, University of Aarhus.

Nelson, D. B., 1991, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59(2), 347–370.

O'Hara, M., 1995, *Market Microstructure Theory*. Blackwell.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannary, 1992, *Numerical Recipes in C*. Cambrigde University Press, 2 edn.

Schwartz, R. A., 1993, *Reshaping the Equity Markets*. Business One Irwin.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Economietrica*, 50, 1–26.

Wooldridge, J. M., 1994, "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden. Elsevier Science B.V., vol. IV, pp. 2640–2738.

APPENDIX A: TABLES AND FIGURES.

Table 1
Selected NYSE stocks

Company/Symbol		Shares	Value	Trades	All Quotes	Mid-quotes
Procter & Gamble Company	**PG**	694	74,595	46932	59658	28086
Disney (Walt) Company (The)	**DIS**	682	47,496	28390	39299	15344
Federal National Mortgage Ass.	**FNM**	1129	42,053	24910	34527	10806
General Motors Corporation	**GM**	757	42,182	32618	39007	14067
Bank-American Corporation	**BAC**	387	38,632	34764	47515	19184
McDonald's Corporation	**MCD**	830	37,572	24720	27693	9513
Monsanto Company	**MTC**	822	31,954	25324	30208	11587
Schlumberger Limited	**SLB**	309	30,848	27787	40363	18193

Table 1 shows the eight randomly selected stocks from the fifty leading NYSE stocks in market value, as of December 31, 1996. Shares and values are in millions. Column 4 and 5 reports the number of trades and quotes after cleaning the data. The last column gives the number of quotes associated with a change in the mid-quote.

Table 2
Summary statistics

Panel A: Trade statistics

FIRM	Av. forward trade	Average price	Average volume	Sells	Buys
PG	20.73	129.39	1148.08	38.0	49.8
DIS	34.27	78.72	1473.07	37.5	44.9
FNM	38.93	45.53	2915.62	39.3	42.3
GM	29.88	64.92	2459.75	36.8	43.7
BAC	27.99	71.30	1840.89	39.1	48.4
MCD	39.30	48.61	2623.38	43.8	41.3
FNM	38.93	45.53	2915.62	39.3	42.3
SLB	34.96	77.71	1658.56	38.2	48.3

Panel B: Quote statistics

FIRM	Av. Obs. fw. quote duration		Av. forward quote duration		Average spread(%)		Amount censored(%)	
	All Q	Mid Q	All Q	Mid Q	All Q	Mid Q	All Q	Mid Q
PG	7.24	13.18	10.47	30.29	0.128	0.142	33.9	56.0
DIS	10.30	22.36	16.02	61.76	0.148	0.169	36.3	59.8
FNM	11.32	30.06	17.03	104.59	0.199	0.234	31.9	64.2
GM	10.83	22.32	18.70	71.89	0.146	0.172	52.2	85.1
BAC	7.76	17.43	11.26	46.71	0.174	0.202	46.5	81.9
MCD	12.77	30.00	21.47	128.45	0.185	0.222	34.1	66.2
MTC	11.47	27.84	17.46	87.38	0.272	0.313	36.3	64.2
SLB	9.82	19.82	14.31	48.90	0.177	0.197	33.0	54.7

Table 2 gives summary statistics for the datasets. Panel A presents statistics related to trades. Av. forward trade is the average length of the durations between successive trades measured in seconds. The next column gives the average transaction price, and the fourth column lists the average size of the amount of shares traded. The fifth and the sixth column report the amount of transactions classified as sells and buys respectively.

Panel B presents statistics related to quotes. Av. obs. fw. quote duration is the average length of the truncated forward quote durations measured in seconds. The next column is the average length of the observed forward quote durations also measured in seconds. Average spread is the average of the percentage spread calculated as the log of the ask price divided by the bid price times 100. The last two columns reports the amount of the observed forward quote durations that were censored. The four statistics are reported both including all quotes and only mid-quotes.

Table 3
Simple correlations with quote durations

Panel A: All quotes included

FIRM	lagged tr.dur	lagged volume	lagged spread	lagged Δspread
PG	0.0604	0.0229	0.0268	0.0155
DIS	0.0588	0.0039	0.0056	0.0035
FNM	0.0457	-0.0111	0.0256	0.0064
GM	0.0427	0.0390	0.0280	0.0178
BAC	0.0783	0.0082	0.0095	0.0123
MCD	0.0434	-0.0214	0.0224	0.0216
MTC	0.0382	-0.0115	0.0018	0.0035
SLB	0.0582	0.0105	0.0157	0.0143

Panel B: Only mid-quote revisions

FIRM	lagged tr.dur	lagged volume	lagged spread	lagged Δspread
PG	0.0659	-0.0067	-0.0165	0.0021
DIS	0.0571	-0.0284	-0.0213	0.0032
FNM	0.0592	-0.0472	0.0060	0.0197
GM	0.0557	-0.0127	0.0018	0.0142
BAC	0.0959	-0.0264	-0.0298	0.0069
MCD	0.0517	-0.0484	0.0082	0.0230
MTC	0.0464	-0.0386	-0.0126	-0.0048
SLB	0.0838	-0.0241	-0.0310	0.0055

Table 3 gives some simple correlations of the observed forward quote durations with lagged trade durations, lagged volume, lagged spread and the change in the lagged spread. In Panel A all quote revisions are included, whereas Panel B only includes mid-quote revisions.

Table 4

Description of explanatory variable

Variable	Name	Description
$Z_{i-1}^{\beta_1}$, $V_{i-1}^{\eta_3}$	lev.QQ_{i-1}	Mean of 10 lagged QQ durations
$Z_{i-1}^{\beta_2}$, $V_{i-1}^{\eta_4}$	ΔSpr_{i-1}	Change in spread between the trade which initialized the present duration and the most recent trade
$Z_{i-1}^{\beta_3}$, $V_{i-1}^{\eta_5}$	lev.Spr_{i-1}	Mean of 10 lagged spreads
$Z_{i-1}^{\beta_4}$, $V_{i-1}^{\eta_6}$	$\sqrt{\text{vol}}_{i-1}$	Square root of the size of the trade which initialized the present duration
$Z_{i-1}^{\beta_5}$, $V_{i-1}^{\eta_7}$	Abs(s.vol)_{i-1}	Absolute value of accumulated signed size of the 10 previous trades
$V_{i-1}^{\eta_1}$	$\frac{x_{i-1}}{\widehat{\psi}_{i-1}}$	lagged surprise trade duration
$V_{i-1}^{\eta_2}$	ψ_{i-1}	Expected forward trade duration
$V_{i-1}^{\eta_8}$	Back.Q_{i-1}	Time from the most recent quote to the trade which initialized the present duration
V^{κ_1}	D_{i-1}^{1}	9:30-10:00 time-of-day dummy
V^{κ_2}	D_{i-1}^{2}	10:00-11:00 time-of-day dummy
V^{κ_3}	D_{i-1}^{3}	11:00-12:00 time-of-day dummy
V^{κ_4}	D_{i-1}^{4}	15:00-16:00 time-of-day dummy

Table 4 defines the explanatory variables included in the two lagged information sets, $\boldsymbol{Z}_{i-1}$ and $\boldsymbol{V}_{i-1}$. The superscripts on these are the coefficient names applied in the analysis.

Table 5
Estimates for the trade equation. Only quotes where the mid-quote changes

$$\ln(\check{\psi}_i) \;=\; \alpha + \delta \ln(\check{\psi}_{i-1}) + \gamma \frac{\check{x}_{i-1}}{\check{\psi}_{i-1}} + \beta_1 \mathsf{lev.Q\check{Q}}_{i-1} + \beta_2 \Delta \mathsf{S\check{p}r}_{i-1} + \beta_3 \mathsf{lev.S\check{p}r}_{i-1} + \beta_4 \sqrt{\mathsf{v\check{o}l}}_{i-1} + \beta_5 \mathsf{Abs(s.v\check{o}l)}_{i-1} + \beta_6 \mathsf{Back.Q}_{i-1}$$

Panel A: Parameter Estimates

FIRM	$\hat{\alpha}$ ($t_{\alpha=0}$)	$\hat{\delta}$ ($t_{\delta=1}$)	$\hat{\gamma}$ ($t_{\gamma=0}$)	$\hat{\beta}_1$ ($t_{\beta_1=0}$)	$\hat{\beta}_2$ ($t_{\beta_2=0}$)	$\hat{\beta}_3$ ($t_{\beta_3=0}$)	$\hat{\beta}_4$ ($t_{\beta_4=0}$)	$\hat{\beta}_5$ ($t_{\beta_5=0}$)	$\hat{\beta}_6$ ($t_{\beta_6=0}$)
PG	***-0.0291***(-7.86)	***0.9780***(5.88)	***0.0386***(11.25)	***0.0046***(3.64)	***0.0482***(4.53)	***0.0061***(4.44)	***-0.0242***(-5.89)	0.0002(0.98)	-0.0014(-2.32)
DIS	-0.0039(-0.88)	***0.9804***(3.34)	***0.0338***(7.37)	-0.0004(-0.29)	0.0334(1.90)	***-0.0078***(-2.84)	***-0.0258***(-4.34)	0.0002(1.45)	-0.0022(-2.44)
FNM	-0.0081(-1.20)	***0.9506***(3.23)	***0.0381***(7.88)	0.0075(2.37)	0.0246(1.18)	-0.0025(-0.47)	***-0.0479***(-4.79)	***0.0007***(3.68)	***-0.0032***(-3.19)
GM	***-0.0207***(-6.04)	***0.9934***(3.17)	***0.0260***(8.75)	-0.0007(-0.68)	0.0356(2.14)	-0.0007(-0.42)	-0.0048(-2.11)	0.0000(0.14)	-0.0004(-0.94)
BAC	***-0.0268***(-5.08)	***0.9682***(3.82)	***0.0450***(8.77)	0.0021(1.04)	0.0164(1.11)	0.0008(0.32)	***-0.0273***(-4.67)	0.0002(1.10)	-0.0005(-0.63)
MCD	***-0.0101***(-2.65)	***0.9829***(3.46)	***0.0304***(6.88)	0.0010(0.95)	***0.0528***(2.87)	-0.0042(-1.58)	***-0.0232***(-4.51)	0.0002(1.68)	-0.0011(-2.23)
MTC	***-0.0142***(-5.39)	***0.9919***(4.44)	***0.0302***(9.66)	-0.0012(-1.51)	***0.0509***(3.16)	-0.0007(-0.44)	***-0.0192***(-4.98)	0.0001(0.58)	-0.0004(-0.84)
SLB	***-0.0181***(-6.45)	***0.9936***(3.74)	***0.0250***(8.69)	0.0001(0.09)	***0.0507***(3.30)	-0.0013(-0.94)	-0.0055(-2.08)	-0.0002(-1.49)	-0.0006(-1.18)

Panel B: Diagnostics

FIRM	$LB(\check{X})$	$LB(\xi)$	$E(\xi)$	$St(\xi)$	$St(\xi) - E(\xi)$	E-R EDT	LM 1	LM 2	Max Like
PG	***2919.4***	***41.9***	1.0000	1.0679	0.0679	***10.76***	0.26	1.26	-45677.60
DIS	***898.3***	8.7	0.9998	1.1754	0.1756	***22.73***	0.03	2.21	-27738.10
FNM	***582.1***	11.2	1.0000	1.1151	0.1151	***13.58***	1.57	0.59	-24506.88
GM	***1427.7***	***33.5***	1.0002	1.1327	0.1325	***18.07***	2.43	2.19	-31780.80
BAC	***1439.1***	20.7	1.0002	1.1667	0.1665	***23.81***	1.61	0.65	-33979.13
MCD	***1009.8***	***31.7***	1.0001	1.0422	0.0421	***4.79***	1.72	3.84	-24161.00
MTC	***2586.8***	***21.4***	1.0003	1.1129	0.1126	***13.42***	0.16	0.56	-23787.39
SLB	***1419.9***	***53.0***	1.0001	1.1539	0.1538	***19.53***	0.07	0.57	-26895.87

Table 5 reports estimation results for the trade equation defined as above using the thinned quote series containing only quotes with mid-quote changes. Panel A gives the parameters estimates with T-statistics reported in parentheses. In panel B some diagnostics are given. In the first column $LB(\check{X})$ is short for the Ljung-Box test with 15 lags on $\check{X}$, the second column is likewise for ξ. E-R EDT is short for: Engle-Russell Excess Dispersion Test. LM 1 and LM 2 are Lagrange Multiplier tests for: 1) first lagged quote-quote duration, and 2) four hourly dummies 9:30-10:00, 10:00-11:00, 11:00-12:00 and 15:00-16:00. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant.

Table 6
Estimates for the quote equation. Only quotes where the mid-quote changes

$$\begin{aligned}\ln(\varphi_i) &= \mu + \rho\ln(\varphi_{i-1}) + \delta_1\frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2\frac{\tilde{y}_{i-1}}{\varphi_{i-1}}d_{i-1} + \tau\frac{x_i}{\widehat{\psi}_i} + \eta_1\frac{x_{i-1}}{\widehat{\psi}_{i-1}} + \eta_2\ln(\widehat{\psi}_{i-1}) + \eta_3\mathsf{lev.QQ}_{i-1} \\ &+ \eta_4\Delta\mathsf{Spr}_{i-1} + \eta_5\mathsf{lev.Spr}_{i-1} + \eta_6\sqrt{\mathsf{vol}}_{i-1} + \eta_7\mathsf{Abs(s.vol)}_{i-1} + \eta_8\mathsf{Back.Q}_{i-1}\end{aligned}$$

FIRM	$\hat{\mu}$ ($t_{\mu=0}$)	$\hat{\rho}$ (SE)	$\hat{\delta}_1$ ($t_{\delta_1=0}$)	$\hat{\delta}_2$ ($t_{\delta_2=0}$)	$\hat{\tau}$ ($t_{\tau=0}$)	$\hat{\eta}_1$ ($t_{\eta_1=0}$)	$\hat{\eta}_2$ ($t_{\eta_2=0}$)
PG	***-0.3700***(-7.65)	*0.6877*(.046)	***0.2343***(11.22)	-0.0410(-1.65)	***0.8153***(23.18)	-0.1943(-2.43)	0.0152(0.90)
DIS	***-0.2718***(-3.74)	*0.5778*(.054)	***0.2514***(10.62)	-0.0287(-0.79)	***1.5099***(24.97)	***-0.6306***(-5.77)	***0.0810***(2.64)
FNM	-0.1912(-2.43)	*0.6091*(.051)	***0.2758***(8.36)	0.0522(1.05)	***1.8146***(23.55)	***-0.5853***(-4.12)	0.0354(0.86)
GM	***-0.4267***(-6.47)	*0.5249*(.054)	***0.3577***(8.74)	-0.1045(-1.87)	***1.3643***(26.20)	-0.2865(-2.52)	***0.1124***(3.88)
BAC	***-0.3507***(-6.38)	*0.5348*(.054)	***0.2654***(9.38)	-0.0815(-2.47)	***1.0524***(28.07)	-0.0886(-0.96)	***0.0932***(3.04)
MCD	-0.1490(-1.47)	*0.4924*(.054)	***0.2790***(13.08)	-0.0215(-0.41)	***1.9866***(23.89)	-0.4157(-2.53)	0.0695(1.61)
MTC	***-0.1654***(-2.74)	*0.6147*(.053)	***0.2769***(10.07)	-0.0227(-0.55)	***1.5655***(20.81)	***-0.6034***(-4.32)	-0.0316(-1.32)
SLB	***-0.5128***(-8.67)	*0.4241*(.043)	***0.2461***(12.77)	-0.0675(-2.28)	***1.1826***(23.96)	-0.1846(-1.85)	***0.1456***(5.06)

FIRM	$\hat{\eta}_3$ ($t_{\eta_3=0}$)	$\hat{\eta}_4$ ($t_{\eta_4=0}$)	$\hat{\eta}_5$ ($t_{\eta 5=0}$)	$\hat{\eta}_6$ ($t_{\eta_6=0}$)	$\hat{\eta}_7$ ($t_{\eta 7=0}$)	$\hat{\eta}_8$ ($t_{\eta_8=0}$)
PG	***0.0990***(5.29)	***0.0537***(4.49)	-0.0068(-1.06)	-0.0369(-1.93)	***0.0039***(2.62)	***0.0479***(4.45)
DIS	***0.0810***(3.98)	***0.0601***(3.63)	***-0.0709***(-4.53)	***-0.2379***(-10.11)	***0.0111***(3.33)	***0.0862***(5.12)
FNM	***0.0913***(4.00)	***0.2089***(9.61)	***-0.0606***(-2.74)	***-0.3791***(-13.93)	***0.0107***(3.83)	***0.1248***(4.90)
GM	***0.0772***(3.44)	***0.1532***(9.10)	-0.0162(-1.04)	***-0.2174***(-8.85)	***0.0087***(2.88)	***0.1250***(6.09)
BAC	***0.1211***(5.70)	***0.0862***(5.31)	***-0.0524***(-3.18)	***-0.2279***(-10.31)	0.0036(1.67)	***0.0947***(5.81)
MCD	***0.1368***(5.15)	***0.2053***(9.08)	***-0.0932***(-3.72)	***-0.4395***(-12.36)	***0.0110***(3.62)	***0.2172***(6.01)
MTC	***0.1452***(5.82)	0.0713(2.30)	***-0.0608***(-4.29)	***-0.3045***(-10.78)	0.0026(1.02)	***0.1067***(4.93)
SLB	***0.1148***(5.17)	0.0240(1.57)	-0.0311(-2.27)	***-0.1942***(-8.36)	***0.0099***(3.39)	***0.1426***(8.61)

Table 6 reports estimation results for the quote equation defined as above using the thinned quote series containing only quotes with mid-quote changes. T-statistics are reported in parentheses. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant. This does not apply to ρ.

Table 7

Diagnostics for the quote equation. Only quotes where the mid-quote changes

Diagnostics for quote equation of Table 9

FIRM	$LB(\tilde{Y})$	$LB(\xi)$	$E(\varepsilon)$	$St(\varepsilon)$	$St(\varepsilon) - E(\varepsilon)$	LM 1	LM 2	Max Like
PG	***65304.2***	***57.6***	1.0000	0.8127	-0.1873	1.19	***20.72***	-19175.34
DIS	***21293.2***	***41.7***	1.0000	0.7818	-0.2182	0.97	7.90	-11549.57
FNM	***30330.4***	***45.1***	1.0001	0.7866	-0.2135	3.19	***18.97***	-9316.12
GM	***64808.8***	***56.9***	1.0001	0.6968	-0.3033	***9.20***	12.61	-13166.11
BAC	***12446.3***	30.0	1.0000	0.8048	-0.1952	8.27	5.32	-15339.85
MCD	***14844.4***	***41.8***	1.0000	0.7575	-0.2425	***16.6***	***23.56***	-10327.43
MTC	***41795.8***	***33.0***	1.0001	0.8132	-0.1869	0.21	7.96	-8707.37
SLB	***4260.9***	***40.6***	1.0000	0.8230	-0.1771	1.59	10.10	-13587.36

Table 7 gives several diagnostics the for the quote equation corresponding to the model defined in Table 6. In the first column $LB(\tilde{Y})$ is short for the Ljung-Box test with 15 lags on $\tilde{Y}$, the second column is likewise for ε. LM 1 and LM 2 are Lagrange Multiplier tests for: 1) first lagged quote-quote duration, and 1) four hourly dummies 9:30-10:00, 10:00-11:00, 11:00-12:00 and 15:00-16:00. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant.

Table 8
Estimates for the quote equation

$$\ln(\varphi_i) = \mu + \rho\ln(\varphi_{i-1}) + \delta_1\frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2\frac{\tilde{y}_{i-1}}{\varphi_{i-1}}d_{i-1} + \tau\frac{x_i}{\widehat{\psi}_i} + \eta_1\frac{x_{i-1}}{\widehat{\psi}_{i-1}} + \eta_2\ln(\widehat{\psi}_{i-1}) + \eta_3\mathsf{lev.QQ}_{i-1}$$
$$+ \eta_4\Delta\mathsf{Spr}_{i-1} + \eta_5\mathsf{lev.Spr}_{i-1} + \eta_6\sqrt{\mathsf{vol}}_{i-1} + \eta_7\mathsf{Abs(s.vol)}_{i-1} + \eta_8\mathsf{Back.Q}_{i-1}$$

FIRM	$\hat{\mu}$ ($t_{\mu=0}$)	$\hat{\rho}$ (SE)	$\hat{\delta}_1$ ($t_{\delta_1=0}$)	$\hat{\delta}_2$ ($t_{\delta_2=0}$)	$\hat{\tau}$ ($t_{\tau=0}$)	$\hat{\eta}_1$ ($t_{\eta_1=0}$)	$\hat{\eta}_2$ ($t_{\eta_2=0}$)
PG	***-0.4808***(-8.25)	*0.6339*(.051)	***0.1257***(7.91)	-0.0199(-1.00)	-0.0029(-0.07)	***0.3983***(7.23)	-0.0222(-1.33)
DIS	-0.3455(-1.03)	*0.6391*(.407)	0.1047(2.32)	-0.0411(-0.83)	***0.4225***(7.41)	-0.1599(-0.49)	0.0050(0.16)
FNM	-0.2933(-2.55)	*0.7219*(.086)	***0.1067***(4.38)	0.0396(2.30)	***0.6445***(8.79)	-0.2287(-1.79)	***-0.0964***(-3.17)
GM	***-0.3083***(-6.32)	*0.7542*(.045)	***0.1454***(10.11)	-0.0146(-1.01)	***0.2475***(4.59)	-0.0444(-0.73)	***-0.0422***(-3.01)
BAC	***-0.4684***(-3.49)	*0.5110*(.152)	***0.1741***(3.96)	-0.0541(-1.35)	***0.1102***(3.06)	***0.2287***(3.18)	0.0181(0.73)
MCD	***-0.3124***(-3.26)	*0.6218*(.087)	***0.1255***(5.93)	0.0484(2.02)	***0.8394***(10.67)	-0.1376(-0.95)	***-0.1967***(-3.99)
MTC	***-0.2444***(-3.92)	*0.7617*(.055)	***0.1241***(5.73)	0.0199(1.13)	***0.4851***(5.96)	-0.0209(-0.16)	***-0.0802***(-3.29)
SLB	***-0.4724***(-3.87)	*0.6603*(.110)	***0.1087***(6.83)	0.0103(0.61)	***0.3533***(6.38)	0.0566(0.39)	0.0545(2.34)

FIRM	$\hat{\eta}_3$ ($t_{\eta_3=0}$)	$\hat{\eta}_4$ ($t_{\eta_4=0}$)	$\hat{\eta}_5$ ($t_{\eta 5=0}$)	$\hat{\eta}_6$ ($t_{\eta_6=0}$)	$\hat{\eta}_7$ ($t_{\eta 7=0}$)	$\hat{\eta}_8$ ($t_{\eta_8=0}$)
PG	***0.0972***(5.16)	***0.1329***(13.18)	***0.0303***(4.41)	***0.1335***(7.48)	0.0010(0.74)	***0.0304***(4.05)
DIS	0.1316(0.93)	***0.1051***(7.15)	0.0395(1.09)	-0.0242(-0.66)	0.0065(0.98)	0.0339(1.30)
FNM	***0.1172***(2.74)	***0.1723***(10.12)	0.0437(2.56)	***-0.0830***(-3.49)	***0.0056***(3.39)	***0.0301***(3.43)
GM	***0.0787***(4.19)	***0.1222***(9.88)	***0.0474***(5.03)	***0.0903***(5.34)	0.0009(0.77)	***0.0455***(5.79)
BAC	***0.1206***(2.89)	***0.0842***(6.06)	***0.0468***(2.85)	0.0110(0.62)	0.0005(0.28)	***0.0322***(3.05)
MCD	***0.2022***(3.73)	***0.1473***(6.98)	***0.0589***(3.05)	***-0.1930***(-4.33)	***0.0087***(3.24)	***0.0738***(3.77)
MTC	***0.0690***(2.63)	***0.1759***(9.07)	***0.0349***(3.52)	***-0.1047***(-4.68)	0.0013(0.78)	0.0119(1.48)
SLB	0.0688(2.36)	***0.0985***(6.72)	***0.0329***(2.72)	0.0358(1.37)	0.0081(2.22)	0.0079(0.97)

Table 8 gives the estimates of the quote equation defined as above. T-statistics are reported in parentheses. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant. This does not apply to ρ.

Table 9

Diagnostics for the quote equation

Diagnostics for quote equation of Table 6

FIRM	$LB(\tilde{Y})$	$LB(\xi)$	$E(\varepsilon)$	$St(\varepsilon)$	$St(\varepsilon) - E(\varepsilon)$	LM 1	LM 2	Max Like
PG	***7336.8***	***41.6***	1.0001	1.2158	0.2157	0.00	5.24	-21632.50
DIS	***790.3***	25.4	1.0000	1.0095	0.0095	0.22	4.19	-17761.31
FNM	***1527.9***	13.1	1.0001	1.2409	0.2408	2.58	6.11	-11084.83
GM	***3816.4***	14.0	1.0002	0.9894	-0.0108	0.53	5.75	-19583.90
BAC	***966.6***	***57.6***	1.0000	1.1346	0.1346	0.26	***16.50***	-20765.65
MCD	***2346.3***	16.8	1.0000	1.2992	0.2991	4.03	13.08	-9704.77
MTC	***14009.5***	23.0	1.0003	1.2968	0.2965	0.50	3.67	-7084.17
SLB	***1627.6***	20.0	1.0001	1.1706	0.1705	0.21	11.28	-14260.78

Table 9 gives several diagnostics the for the quote equation corresponding to the model defined in Table 8. In the first column $LB(\tilde{Y})$ is short for the Ljung-Box test with 15 lags on $\tilde{Y}$, the second column is likewise for ε. LM 1 and LM 2 are Lagrange Multiplier tests for: 1) first lagged quote-quote duration, and 2) four hourly dummies 9:30-10:00, 10:00-11:00, 11:00-12:00 and 15:00-16:00. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant.

Table 10
Estimates for the quote equation including time-of-day dummies. Only quotes where the mid-quote changes

$$\ln(\varphi_i) = \mu + \rho\ln(\varphi_{i-1}) + \delta_1\frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2\frac{\tilde{y}_{i-1}}{\varphi_{i-1}}d_{i-1} + \tau\frac{x_i}{\widehat{\psi}_i} + \eta_1\frac{x_{i-1}}{\widehat{\psi}_{i-1}} + \eta_2\ln(\widehat{\psi}_{i-1}) + \eta_3\mathsf{lev.QQ}_{i-1} + \eta_4\Delta\mathsf{Spr}_{i-1} + \eta_5\mathsf{lev.Spr}_{i-1}$$
$$+ \eta_6\sqrt{\mathsf{vol}}_{i-1} + \eta_7\mathsf{Abs(s.vol)}_{i-1} + \eta_8\mathsf{Back.Q}_{i-1} + \kappa_1\mathsf{D}^1_{i-1} + \kappa_2\mathsf{D}^2_{i-1} + \kappa_3\mathsf{D}^3_{i-1} + \kappa_3\mathsf{D}^3_{i-1}$$

FIRM	$\hat{\mu}$ ($t_{\mu=0}$)	$\hat{\rho}$ (SE)	$\hat{\delta}_1$ ($t_{\delta_1=0}$)	$\hat{\delta}_2$ ($t_{\delta_2=0}$)	$\hat{\tau}$ ($t_{\tau=0}$)	$\hat{\eta}_1$ ($t_{\eta_1=0}$)	$\hat{\eta}_2$ ($t_{\eta_2=0}$)	$\hat{\eta}_3$ ($t_{\eta_3=0}$)	$\hat{\eta}_4$ ($t_{\eta_4=0}$)
PG	***-0.4547***(-7.64)	*0.6624*(.048)	***0.2458***(11.29)	-0.0485(-1.87)	***0.8169***(23.55)	-0.1739(-2.13)	***0.0606***(2.75)	***0.1009***(5.32)	***0.0539***(4.56)
DIS	***-0.2835***(-3.71)	*0.5673*(.054)	***0.2563***(10.81)	-0.0280(-0.78)	***1.5223***(25.45)	***-0.6029***(-5.53)	0.0815(2.48)	***0.0839***(4.03)	***0.0630***(3.86)
FNM	***-0.4260***(-3.56)	*0.5414*(.057)	***0.2989***(8.60)	0.0298(0.59)	***1.7974***(25.01)	-0.3681(-2.50)	0.1281(2.23)	***0.0900***(3.66)	***0.2021***(9.62)
GM	***-0.4680***(-6.26)	*0.5294*(.051)	***0.3608***(9.17)	-0.0963(-1.84)	***1.3593***(26.45)	***-0.2962***(-2.79)	***0.1207***(3.59)	***0.0778***(3.53)	***0.1524***(9.22)
BAC	***-0.4080***(-6.04)	*0.5187*(.055)	***0.2704***(9.55)	-0.0797(-2.44)	***1.0564***(28.71)	-0.0775(-0.84)	***0.1275***(3.14)	***0.1203***(5.66)	***0.0842***(5.36)
MCD	***-0.3078***(-2.76)	*0.4937*(.052)	***0.2777***(14.69)	***0.1193***(2.62)	***2.0629***(25.42)	***-0.6164***(-4.15)	***0.1395***(2.73)	***0.1102***(4.42)	***0.1930***(9.01)
MTC	***-0.2199***(-3.53)	*0.5995*(.053)	***0.2904***(10.23)	-0.0221(-0.55)	***1.5795***(21.61)	***-0.5788***(-4.25)	-0.0188(-0.80)	***0.1433***(5.83)	0.0742(2.56)
SLB	***-0.5518***(-8.10)	*0.4226*(.043)	***0.2567***(12.65)	-0.0748(-2.51)	***1.2099***(24.67)	-0.2156(-2.18)	***0.1575***(4.75)	***0.1128***(5.10)	0.0249(1.66)

FIRM	$\hat{\eta}_5$	$\hat{\eta}_6$ ($t_{\eta_6=0}$)	$\hat{\eta}_7$ ($t_{\eta_7=0}$)	$\hat{\eta}_8$ ($t_{\eta_8=0}$)	$\hat{\kappa}_1$ ($t_{\kappa_1=0}$)	$\hat{\kappa}_2$ ($t_{\kappa_2=0}$)	$\hat{\kappa}_3$ ($t_{\kappa_3=0}$)	$\hat{\kappa}_4$ ($t_{\kappa_4=0}$)
PG	-0.0134(-1.86)	-0.0395(-2.04)	***0.0045***(2.95)	***0.0512***(4.54)	***0.0406***(2.74)	0.0139(1.39)	0.0064(0.65)	***0.0558***(3.58)
DIS	***-0.0704***(-4.45)	***-0.2356***(-10.12)	***0.0110***(3.32)	***0.0881***(5.21)	***-0.0585***(-2.63)	-0.0102(-0.62)	-0.0161(-0.96)	0.0233(1.07)
FNM	-0.0557(-2.42)	***-0.3825***(-14.66)	***0.0112***(3.73)	***0.1493***(5.29)	-0.0546(-1.93)	0.0093(0.42)	0.0519(2.45)	***0.1028***(3.07)
GM	-0.0069(-0.43)	***-0.2154***(-9.03)	***0.0082***(2.82)	***0.1235***(6.24)	-0.0483(-1.93)	0.0100(0.56)	0.0018(0.10)	0.0529(2.22)
BAC	***-0.0542***(-3.13)	***-0.2284***(-10.42)	0.0038(1.73)	***0.0964***(5.87)	0.0188(0.73)	-0.0034(-0.20)	0.0034(0.20)	0.0532(2.56)
MCD	***-0.0792***(-3.18)	***-0.4323***(-12.98)	***0.0110***(3.58)	***0.2098***(6.16)	***-0.1243***(-3.72)	-0.0341(-1.44)	0.0065(0.27)	***0.1271***(3.90)
MTC	***-0.0522***(-3.68)	***-0.3017***(-10.95)	0.0029(1.15)	***0.1092***(5.11)	***-0.0909***(-3.44)	-0.0252(-1.41)	0.0167(0.92)	0.0547(2.20)
SLB	-0.0266(-1.95)	***-0.1989***(-8.63)	***0.0099***(3.45)	***0.1445***(8.59)	-0.0548(-2.01)	0.0182(0.87)	0.0147(0.70)	0.0627(2.39)

Table 10 gives the estimates of the quote equation including time-of-the-day dummies defined as above, using the thinned quote series containing only quotes with mid-quote changes. The four hourly dummies are 9:30-10:00, 10:00-11:00, 11:00-12:00 and 15:00-16:00. T-statistics are reported in parentheses. Numbers in italic boldface are significant at the 99% level, number in normal font are significant at the 95% level. The numbers typed with very small types are insignificant. This does not apply to ρ.

Table 11
Joint estimation. The trade equation, only mid-quote changes

$$\ln(\check{\psi}_i) \quad = \quad \alpha + \delta \ln(\check{\psi}_{i-1}) + \gamma \frac{\check{x}_{i-1}}{\check{\psi}_{i-1}} + \beta_2 \Delta \check{\mathsf{Spr}}_{i-1} + \beta_3 \mathsf{lev.}\check{\mathsf{Spr}}_{i-1} + \beta_4 \sqrt{\check{\mathsf{vol}}}_{i-1}$$

Panel A: Parameter Estimates

FIRM	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
PG*	-0.0279(.0031)	0.9821(.0031)	0.0375(.0033)	0.0489(.0106)	0.0052(.0012)	-0.0180(.0025)
PG	-0.0294(.0033)	0.9801(.0035)	0.0384(.0034)	0.0549(.0108)	0.0054(.0013)	-0.0175(.0025)
DIS*	-0.0096(.0040)	0.9781(.0059)	0.0331(.0045)	0.0362(.0176)	-0.0054(.0026)	-0.0228(.0053)
DIS	-0.0137(.0038)	0.9815(.0052)	0.0303(.0045)	0.0232(.0178)	-0.0040(.0024)	-0.0160(.0043)
FNM*	-0.0112(.0050)	0.9696(.0102)	0.0324(.0049)	0.0206(.0206)	-0.0028(.0041)	-0.0243(.0070)
FNM	-0.0141(.0045)	0.9746(.0105)	0.0282(.0054)	-0.0215(.0217)	0.0002(.0040)	-0.0188(.0073)
GM*	-0.0224(.0031)	0.9923(.0017)	0.0260(.0029)	0.0371(.0166)	-0.0000(.0020)	-0.0048(.0017)
GM	-0.0214(.0031)	0.9929(.0017)	0.0238(.0028)	0.0090(.0170)	0.0003(.0016)	-0.0037(.0017)
BAC*	-0.0257(.0040)	0.9715(.0064)	0.0438(.0048)	0.0156(.0147)	-0.0000(.0018)	-0.0229(.0046)
BAC	-0.0269(.0043)	0.9724(.0070)	0.0402(.0051)	0.0206(.0154)	0.0025(.0020)	-0.0200(.0047)
MCD*	-0.0116(.0033)	0.9829(.0044)	0.0297(.0041)	0.0538(.0184)	-0.0039(.0025)	-0.0192(.0039)
MCD	-0.0112(.0033)	0.9839(.0045)	0.0271(.0042)	0.0280(.0199)	-0.0040(.0025)	-0.0162(.0037)
MTC*	-0.0161(.0026)	0.9904(.0017)	0.0304(.0031)	0.0521(.0162)	-0.0001(.0014)	-0.0195(.0032)
MTC	-0.0136(.0025)	0.9903(.0019)	0.0272(.0032)	0.0356(.0165)	0.0002(.0014)	-0.0188(.0033)
SLB*	-0.0174(.0025)	0.9932(.0016)	0.0248(.0028)	0.0523(.0154)	-0.0008(.0012)	-0.0084(.0020)
SLB	-0.0180(.0027)	0.9930(.0016)	0.0238(.0029)	0.0502(.0157)	-0.0006(.0013)	-0.0067(.0020)

Table 11 compares the estimates using the two-step approach and joint estimation of the trade equation defined as above using the thinned quote series containing only quotes with mid-quote changes. The rows marked with an asterix are associated with the two-step approach. Standard errors are reported in parentheses.

Table 12
Joint estimation. The quote equation, only mid-quote changes

$$\ln(\varphi_i) \quad = \quad \mu + \rho \ln(\varphi_{i-1}) + \delta_1 \frac{\tilde{y}_{i-1}}{\varphi_{i-1}} + \delta_2 \frac{\tilde{y}_{i-1}}{\varphi_{i-1}} d_{i-1} + \tau \frac{x_i}{\widehat{\psi}_i} + \eta_1 \frac{x_{i-1}}{\widehat{\psi}_{i-1}} + \eta_2 \ln(\widehat{\psi}_{i-1}) + \eta_4 \Delta \mathsf{Spr}_{i-1} + \eta_5 \mathsf{lev.Spr}_{i-1} + \eta_6 \sqrt{\mathsf{vol}}_{i-1}$$

FIRM	$\hat{\mu}$	$\hat{\rho}$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\tau}$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_4$	$\hat{\eta}_5$	$\hat{\eta}_6$
PG*	-0.2248(.0286)	0.8643(.0226)	0.1674(.0121)	0.0655(.0119)	0.8257(.0352)	-0.4569(.0590)	0.0311(.0113)	0.0408(.0121)	-0.0078(.0042)	0.0221(.0106)
PG	-0.2232(.0291)	0.8650(.0229)	0.1672(.0121)	0.0654(.0119)	0.8271(.0360)	-0.4563(.0610)	0.0297(.0118)	0.0414(.0122)	-0.0078(.0042)	0.0223(.0106)
DIS*	-0.1492(.0471)	0.7482(.0359)	0.1980(.0177)	0.1395(.0222)	1.5012(.0604)	-0.8983(.0947)	0.0761(.0247)	0.0430(.0186)	-0.0600(.0144)	-0.1454(.0297)
DIS	-0.1116(.0448)	0.7546(.0344)	0.2000(.0181)	0.1391(.0223)	1.4852(.0609)	-0.9165(.0950)	0.0636(.0223)	0.0455(.0189)	-0.0639(.0151)	-0.1523(.0306)
FNM*	-0.1541(.0647)	0.7317(.0388)	0.2418(.0284)	0.2245(.0323)	1.8156(.0764)	-0.8075(.1385)	0.0878(.0361)	0.2052(.0225)	-0.0640(.0184)	-0.2725(.0349)
FNM	-0.1841(.0774)	0.7275(.0407)	0.2438(.0294)	0.2254(.0330)	1.8452(.0791)	-0.8339(.1486)	0.0974(.0441)	0.1988(.0231)	-0.0601(.0182)	-0.2671(.0348)
GM*	-0.2949(.0443)	0.7439(.0333)	0.2562(.0260)	0.1282(.0288)	1.3617(.0518)	-0.6395(.0925)	0.1070(.0211)	0.1444(.0169)	-0.0110(.0110)	-0.1408(.0246)
GM	-0.2948(.0451)	0.7417(.0335)	0.2575(.0263)	0.1262(.0292)	1.3712(.0513)	-0.6409(.0933)	0.1073(.0219)	0.1412(.0171)	-0.0117(.0109)	-0.1415(.0247)
BAC*	-0.2309(.0394)	0.6925(.0406)	0.2277(.0218)	0.0607(.0215)	1.0617(.0376)	-0.3020(.0832)	0.1292(.0290)	0.0886(.0184)	-0.0592(.0146)	-0.1804(.0205)
BAC	-0.2349(.0393)	0.6903(.0401)	0.2286(.0218)	0.0603(.0216)	1.0649(.0377)	-0.3073(.0822)	0.1311(.0291)	0.0889(.0182)	-0.0586(.0145)	-0.1790(.0204)
MCD*	-0.0586(.0715)	0.7381(.0394)	0.2231(.0175)	0.2616(.0318)	2.0068(.0845)	-1.0801(.1444)	0.0990(.0356)	0.2112(.0215)	-0.0827(.0189)	-0.3249(.0409)
MCD	-0.0555(.0721)	0.7377(.0379)	0.2233(.0171)	0.2610(.0314)	2.0192(.0833)	-1.0918(.1375)	0.0970(.0359)	0.2070(.0219)	-0.0831(.0186)	-0.3226(.0403)
MTC*	-0.0216(.0360)	0.8347(.0281)	0.1964(.0169)	0.1532(.0220)	1.5846(.0759)	-1.0935(.1034)	-0.0037(.0142)	0.0639(.0284)	-0.0475(.0099)	-0.2071(.0266)
MTC	-0.0165(.0367)	0.8347(.0279)	0.1967(.0168)	0.1532(.0221)	1.5955(.0747)	-1.1067(.1013)	-0.0067(.0149)	0.0615(.0285)	-0.0476(.0099)	-0.2075(.0265)
SLB*	-0.3532(.0531)	0.6330(.0523)	0.1997(.0167)	0.1137(.0203)	1.1811(.0493)	-0.4224(.1190)	0.1617(.0305)	0.0245(.0157)	-0.0395(.0116)	-0.1158(.0268)
SLB	-0.3556(.0561)	0.6324(.0551)	0.1999(.0171)	0.1135(.0204)	1.1823(.0504)	-0.4257(.1278)	0.1633(.0325)	0.0242(.0159)	-0.0395(.0117)	-0.1149(.0274)

Table 12 compares the estimates using the two-step approach and joint estimation of the quote equation defined as above using the thinned quote series containing only quotes with mid-quote changes. The rows marked with an asterix are associated with the two-step approach. Standard errors are reported in parentheses.

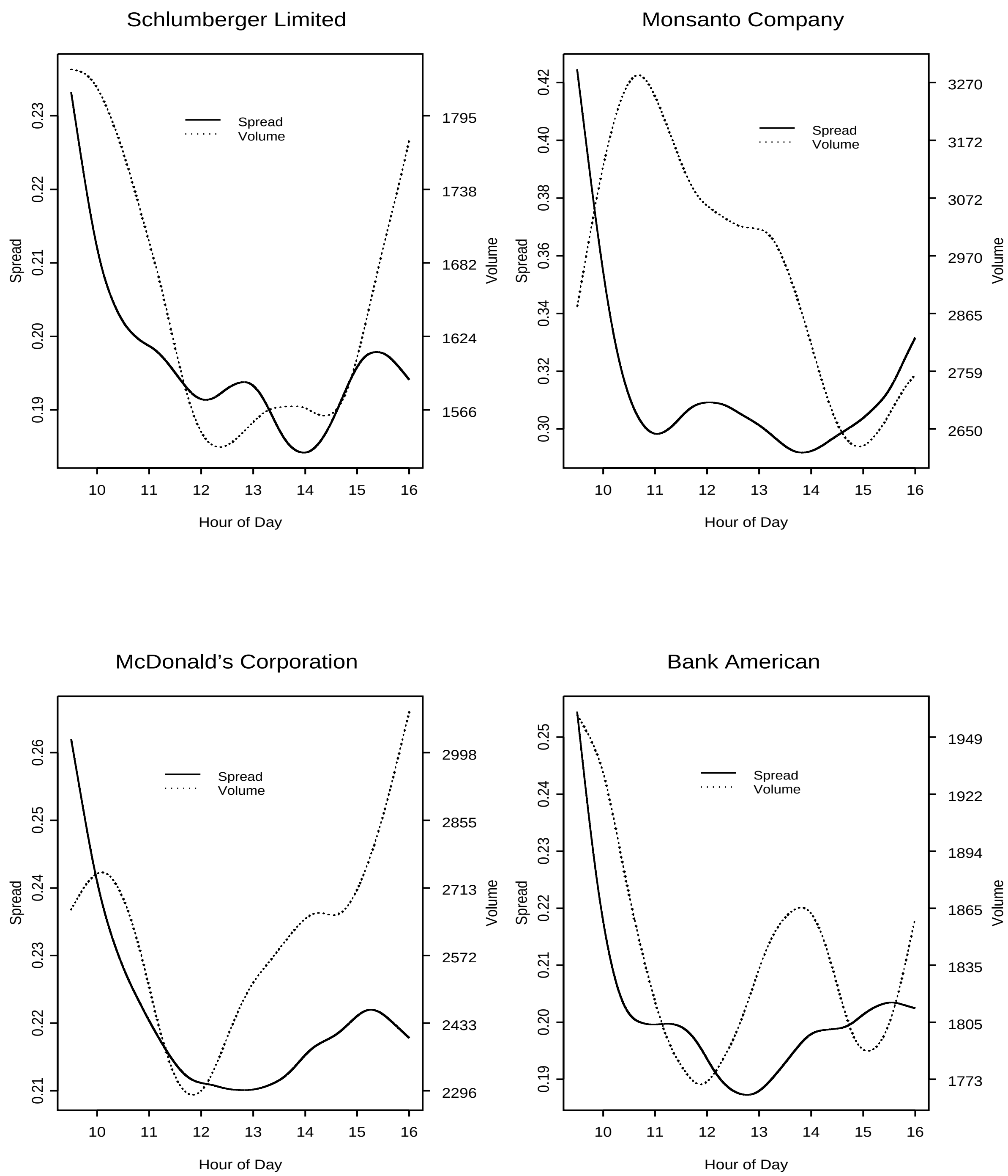

Figure 1a

Time-of-the-day splined mean of volume and spread

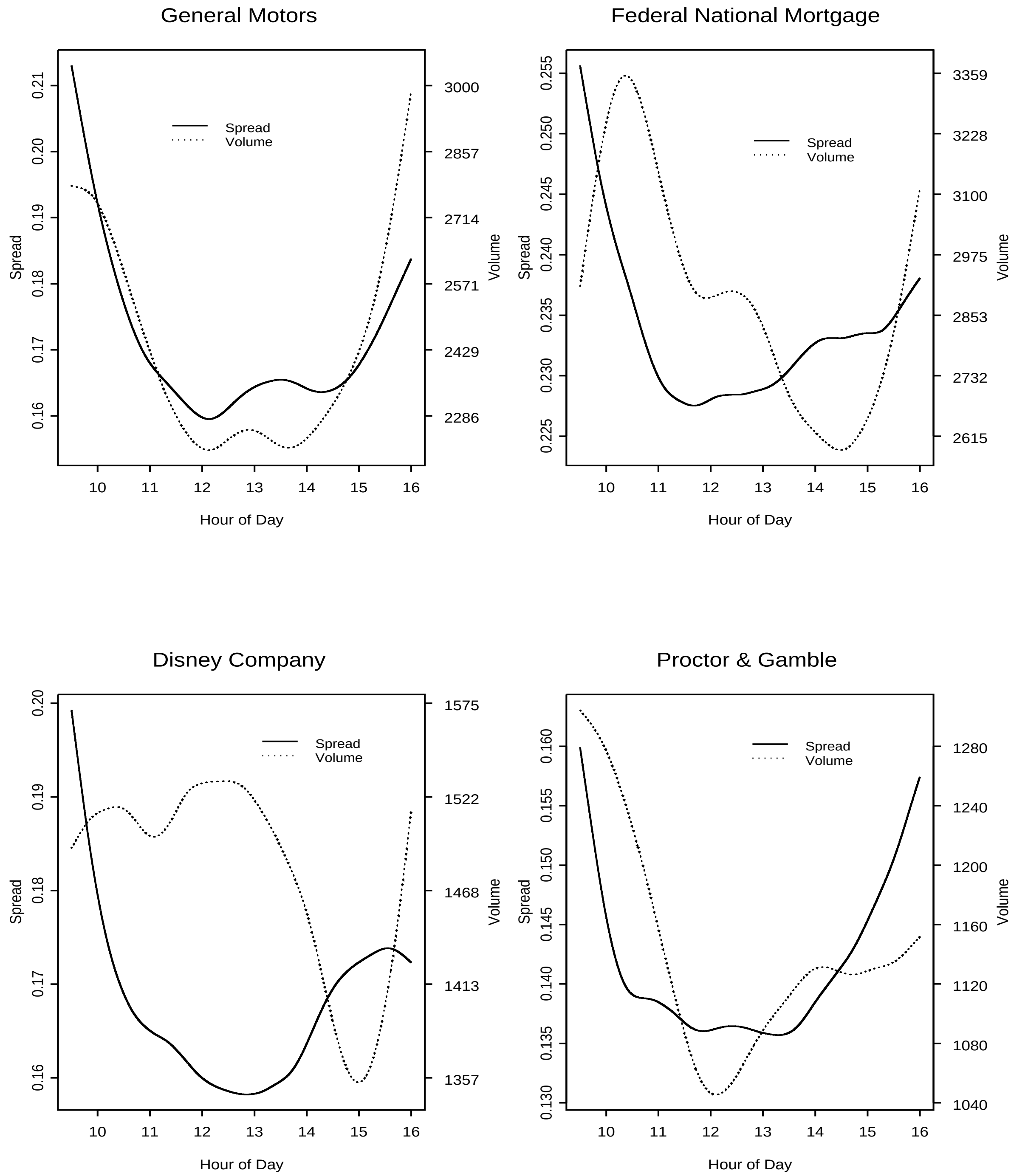

Figure 1b

Time-of-the-day splined mean of volume and spread