INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2002-17

Asymmetry of Information, Market Liquidity and the Activity of the Specialist on the NYSE

T. KAMIONKA[1]

This version : May 2002

[1] CNRS, CREST (paris) and IDEI (Toulouse) France. Email : kamionka@ensae.fr

Asymmetry of Information, Market Liquidity and the Activity of the Specialist on the NYSE

THIERRY KAMIONKA[†]

[†] *CNRS, CREST (Paris) and IDEI (Toulouse), France*
E-mail: kamionka@ensae.fr

This version : May 2002

Summary In this paper we model jointly the interval of time between financial events and the category of these events (trades or quotes). We study the validity of some empirical implications of market microstructure theories. The model we estimate allows to take into account the intraday activity of the markets and allows for serial correlations of the durations. The conditional distribution of the duration and the category of the financial event may depend on the whole history of the process and the characteristics of the market. This model is estimated for IBM on the NYSE using the TAQ data set. We find in particular that, in the case of a wide spread, the information content of the last trade is important. The quantity traded have a significant impact on the probability that the market maker increases the spread. We find a diagonal effect on the category of the trade indicating the possible existence of imitation behavior or orders splitting strategies. This result indicates that trades are successively positively correlated. On the other hand, a trade at the midquote is the signal that more likely no new information is arrived on the market.

Keywords: *Durations, High Frequency Data, Market Microstructure*

Electronic trading systems allow to record continuously the arrival times of trades and quotes. The availability of such high frequency data, permits to consider the arrival times of trades or quotes as random variables.

There are two motives to the placement of orders by traders: liquidity and information (cf. Glosten and Milgrom (1985), Biais et alii (1995) and Bisière and Kamionka (2000)). The literature relative to the microstructure of financial markets has examined the way the specialist could modify its believes on the existence of a new information relative to the traded asset. In particular, it has been shown that the market maker could make evolve these believes by observing the arrivals - or the lack - of transactions. For instance, Easley and O'Hara (1992) consider a model in which the potential arrival of a new information is a random event. These authors show that the absence or the arrival of transactions makes the specialist modify its quotes. They show that there exists a relation between the inter trades duration and the importance of the spread. They find that the spread and volumes will have an influence on the speed of the quotes revision by the specialist. Observing the activity of the market, the specialist is going to revise its believes relative to the potential arrival of a new information on the market and to the nature of this information. The market marker can react to the arrival of a new information modifying its quotes.

In the same way, some authors, as Glosten and Milgrom (1985), have noted that the importance of the spread was connected to the importance of the asymmetry of information existing among the informed and not informed agents on the market. However, the existence of a large spread makes the consumption of the liquidity more expensive and limits ipso facto the frequency of the transactions.

Consequently, the process of arrival of quotes is going to depend on the timing of the trades and on the characteristics of these trades. On the other hand, the process of the arrivals of the trades is going to depend on the quotes posted by the specialist. Finally, the processes of quotes and trades should be studied within a single join modeling.

Engle and Russell (1998) and Engle (2000) have proposed a modeling framework particularly adapted to the characteristics of high frequency data in finance. Bauwens and Giot (2000) model, on the New York Stock Exchange, the price durations relative to bid-ask quote process. They analyze the influence, on the bid-ask quote process, of some characteristics of the trades such as the average volume per trade, the average spread or the trading intensity. To do so, they propose an alternative class of models designed for the study of financial durations. In this new class of models, the logarithm of the conditional expectation of the price duration takes an autoregressive form. The ACD and Log-ACD models allow a dependence between successive durations (see Kamionka (2000) and Bauwens and Giot (2001), for a review of the literature relative to the econometrics of high frequency data).

In this paper, in order to consider the validity of some empirical implications of the market microstructure theories we take into account the information conveyed by intermediate trades arriving between two successive quotes on the market. In order to model the interactions between the arrival of trades and the arrival of quotes on the New York Stock Exchange, we use a transition model (cf. Russel (1999) or Kamionka (2000)). This type of model allows to take into account all the history of the process: characteristics of the market as the spread, previous trades and the sequence of the past quotes.

The model we consider allows to study jointly the durations between financial events and the categories of the events. In this model, every duration is distributed according to a Log-ACD model (cf. Bauwens and Giot, 2000). The successive durations are not considered as independent. The parameters of the conditional duration model are estimated by maximizing a quasi likelihood function (cf. Engle, 2000) and, alternatively, by using a parametric distribution. We take into account the intraday structure of the market activity doing a non parametric regression of the durations on the time of the day of the week.

The model is estimated using data from the "Trades and Quotes" (TAQ) data set of the New York Stock Exchange for IBM and a period going from September 1998 till November 1998. We find the usual intraday activity structure on this type of high frequency data: the activity is maximum at the beginning and at the end of the trading day, minimum in the middle of day (see Gouriéroux, Jasiak and Le Fol, (1999) and Bisière and Kamionka (2000) for the Paris Bourse). The join modeling of the duration between financial events and the category of these events allows to test some empirical implications of the theories of the microstructure of financial markets. We find, in particular, that a large spread is going to have the effect of reducing significantly the mean duration that it is necessary to wait for a new quote when the last event is a trade. A large spread associated to the arrival of a trade is the signal that, more likely, a new information is arrived on the market. The specialist will react more intensively to the arrival of such an event by increasing the spread. This result is compatible with the implications of the model of Easley and O'Hara (1992). However, we do not find the same effect when the last event is a quote associated to a spread decrease indicating that it is costly to consume the liquidity when the spread is wide. The arrival of a trade at the midquote is the signal that, more likely, no information is arrived on the market. The sum of the coefficients of the autoregressive part is close to 1. The estimations shows an important persistence of durations between financial events.

In the first section we present some empirical implications of market microstructure theories. The second section includes a description of the treatment of the intraday activity. The next section contains a presentation of a joint model of durations and categories of financial events. In the fourth section we consider the aspects related to the estimation of the model. The fifth section includes a discussion on the results. The last section concludes.

1. EMPIRICAL IMPLICATIONS OF MICROSTRUCTURE THEORIES

The Easley and O'Hara (1992) model contains several empirical implications.

Lack of trade : the quotes will be modified if there is no trade. Indeed, the absence of trade may signal the fact that no new information has arrived among the participants to the market.

Timing of trades : when the inter-trades duration increases then the spread decreases with time.

Market activity : has the arrival of a (several) trade(s) may signal the existence of a new information, the specialist will modify more intensively its quotes. Consequently, periods of intensive trades should be associated with frequent modifications of the quotes.

Information content of the volume : the behavior of the prices may depend on the volume traded. A high volume traded is a signal of new information has occurred on the market. Trades are arriving more frequently when the volume is important.

The successive trades are positively correlated : as a trade may signal the arrival of a new information the informed agents will continue to trade. Given a new information has occurred, trades are arriving more frequently.

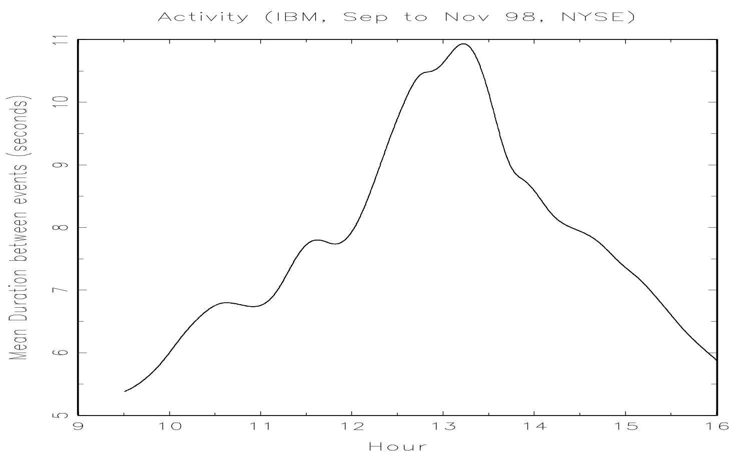

Figure 1. Activity on Monday, IBM.

2. INTRADAY ACTIVITY

Some authors have shown that there exists a typical intraday activity of the markets. Engel and Russel (1998), Engle (2000), Kamionka (2000) for the New-York Stock Exchange, Gourieroux, Jasiak and Le Fol (1999), Bisière and Kamionka (2000) for the Paris Bourse have shown that the activity is maximum at the beginning and at the end of the trading day and minimum in the middle of this day. In order to take into account this U-shape pattern of the activity, we are going to make a non parametric regression of the duration between financial events (trades or quotes) on the time of the day.

Let τ_ℓ denote the arrival time of the financial event ℓ ($\ell \in \mathbb{N}$). Let u_ℓ denote the duration between events $\ell - 1$ and ℓ ($u_\ell = \tau_\ell - \tau_{\ell-1}$). The period of time starting with the financial event $\ell{-}1$ and lasting with the next one is defined as the spell ℓ.

We are going to consider a non parametric regression of U_ℓ on the time of the day $\tau_{\ell-1}$:

$$U_\ell = \mathrm{E}[U_\ell \mid \tau_{\ell-1}] + v_\ell = \phi(\tau_{\ell-1}) + v_\ell,$$

where ϕ is a smooth univariate function and $v_1, \ldots, v_n$ are random variables that are identically distributed with a common density and have zero mean.

The function ϕ is estimated by

$$\hat{\phi}(\tau) = \frac{\displaystyle\sum_{\ell=1}^{n} u_\ell \; K_h(\tau - \tau_{\ell-1})}{\displaystyle\sum_{\ell=1}^{n} K_h(\tau - \tau_{\ell-1})},$$

where h is a bandwidth, $K_h(.) = K(./h)$ and K is a kernel function.

The approach used for the estimation of the regression function is similar to the one used for the smoothing in single index models (see Härdle et alii 1993, Horowitz 1993). The kernel function we use is a symmetric probability density function.

We propose to use the Triweight kernel:

$$K(x) = \frac{35}{32}\,(1 - x^2)^3\, \mathbb{1}_{\{|x|<1\}}.$$

This kernel is often used for kernel smoothing (see Wand and Jones, 1995).

The bandwidth we used is calculated using the normal scale rule for a Triweight kernel

$$h^* = 3.15 * \frac{\hat{\sigma}}{n^{1/5}},$$

where $\hat{\sigma}$ is the empirical standard error of the time of the financial event (in seconds since midnight).

The corresponding rule is usually used in order to choose a bandwidth for a density smoothing and, here, we found that the corresponding value performs well.

On IBM, on the New-York Stock Exchange, we find the usual U-shape pattern of the activity. Figure 1 depicts the activity of the market on Monday from 9h30 to 16h00.

Engel (2000) proposes to work on the transformed durations

$$u'_\ell = \frac{u_\ell}{\hat{\phi}(\tau_{\ell-1})},$$

where $u_\ell = \tau_\ell - \tau_{\ell-1}$.

A similar approach to treat these seasonal patterns is used by Veredas, Rodriguez-Poo and Espasa (2001). The authors consider a non-parametric estimation of the effect of the time of the day and of the day the week and a parametric modeling of the conditional distribution of the financial duration. The approach is similar to the one used by Engle and Russel (1998) who consider a joint model of the seasonality and the dynamics of the duration using cubic splines to model the effect of the time of the day.

3. MODELING JOINTLY TRADES AND QUOTES DURATIONS

We have to model jointly the durations between financial events and the categories of these events (trades or quotes) for several trading days on the New York Stock Exchange. In the empirical application, we are going to distinguish buys from sells, trades at the midpoint of the spread, quotes corresponding only to a modification of the depths, quotes corresponding to an increase (respectively, a decrease) of the spread. Let E denote the space of the categories of financial events.

What we observe is the arrival times of the financial events and the corresponding categories.

Let us consider a given trading day. The process we study is denoted X_t $(t \in \mathbb{R}^+)$. X_t is the category of the last event recorded at time t. Let x_t denote a realization of the process at time t $(x_t \in E)$. The process X_t is constant between two successive transitions. Let τ_0 denote the time of the first observation and τ^* denote the time the observation stops for a given trading day (τ^* = 16h00). It is a transition model.

Let us consider a typical realization of the process for a given trading day (see figure 2). The first even occurred at time τ_0, it is a spread increase. Then, we observe a buy at

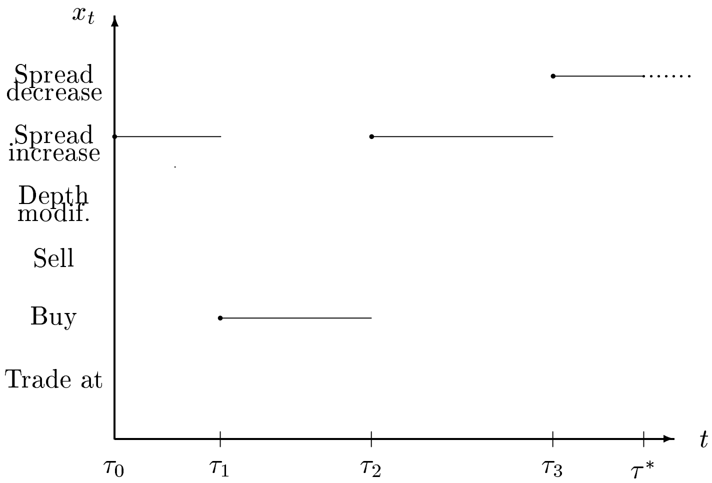

Figure 2. A realization of the process.

time τ_1. At time τ_2 there is a modification of the quotes such that the spread increases and, at time τ_3, the spread decreases.

A spell is a period of time delimited by two consecutive financial events (a buy and a sell, for instance). During a given spell, the value of the process is equal to the category of the event occurring at the starting time of the considered spell. For a given spell we can associate the length of this spell. The spell ℓ starts at time $\tau_{\ell-1}$ and stops at time τ_ℓ ($\tau_\ell > \tau_{\ell-1}$). Let u_ℓ be the duration of spell ℓ. Consequently, we have $u_\ell = \tau_\ell - \tau_{\ell-1}$ (see figure 3). The last observation of the day can be considered to be right-censored: we know that the duration of the last spell is at least equal to $\tau^* - \tau_n$ (see figure 2).

Let $x_\ell \equiv x_{\tau_\ell}$ denote the category of the event occurring at time τ_ℓ. As the process records the arrival times of all the events whatever their categories, we can consider that between two successive events, the explanatory variables are constant. Let z denote a vector of explanatory variables and let n ($n \in I\!N^*$) be the number of completed spells in the considered trading day.

Let us examine a realization of the process for a given day:

$$r(\tau_0, \tau^*) = (y_0, y_1, \ldots, y_\ell, \ldots, y_n, y_{n+1}),$$

$$\text{where} \quad y_\ell = \begin{cases} (\tau_0', x_{\tau_0}'), \text{ if } \ell = 0, \\ (u_\ell, x_{\tau_\ell}), \text{ if } 1 \leq \ell \leq n, \\ (\tau^* - \tau_n, 0), \text{ if } \ell = n+1, \end{cases}$$

and τ_0' is the arrival time of the first event during the period of observation ($\tau_0' \geq \tau_0$).

The first component of the realization includes the information relative to the arrival time of the first event during this trading day and the category of this event (an increase of the spread, for instance). The last component of the realization incorporates

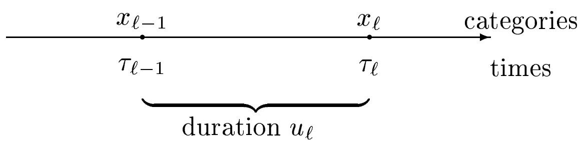

Figure 3. Spell ℓ.

the information that the duration of the last spell is at least equal to the observed value u_{n+1}.

We are going to write the density of a realization $(y_0, y_1, \ldots, y_{n+1})$ conditionally to the first observation y_0 and to the vector of the characteristics of the market z. The density of this realization can be written as a product of marginal and conditional probability density functions. Let $f(y_1, \ldots, y_{n+1} \mid y_0, z_{\ell-1}; \theta)$ denote this density. We obtain that

$$f(y_1, \ldots, y_{n+1} \mid y_0, z; \theta) = \prod_{\ell=1}^{n+1} f(y_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta),$$

where $y_\ell = (u_\ell, x_\ell)$ (for $\ell = 0, \ldots, n+1$) and $f(y_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)$ is the conditional density of Y_ℓ given $y_0, \ldots, y_{\ell-1}, z_{\ell-1}$.

We assume that the conditional distribution of Y_ℓ does not depend on the arrival order of this observation (the index ℓ) but only on the sequence of past realizations and on the vector z.

We have then to consider the joint distribution of $Y_\ell = (U_\ell, X_\ell)$ conditionally to the past history of the process and the explanatory variables.

3.1. Modeling within a given spell

Let $U^*_{\ell,k}$ be the duration we should wait in order to observe a transition towards the category k of financial events ($k \in E$) at the end of the spell ℓ. We assume that these latent variables are independent.

The duration of a given spell is the duration between two successive financial events and is the minimum of six latent durations[1]:

$$U_\ell = \min_{j \in E} U^*_{\ell,j},$$

where

$$X_\ell = \arg\min_{j \in E} U^*_{\ell,j},$$

is the category of the financial event occurring at time τ_ℓ.

[1] One for each category of financial events : buy, sell, trade at the midpoint of the spread, quote without modification of the spread, spread decrease, spread increase.

Let $h_j(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)$ be the hazard function associated to the latent duration $U^*_{\ell,j}$ and let $S_j(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)$ be the survival function of the same latent duration.

The conditional probability density function of the duration between two consecutive financial events when the next event category is j given the history of the process up to the starting time of spell ℓ (namely $\tau_{\ell-1}$) and given the characteristics of the market $z_{\ell-1}$ is

$$\begin{aligned} f(u, j \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) &= h_j(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) \\ &\quad \prod_{k \in E} S_k(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) \\ &= h_j(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) \\ &\quad S(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta), \end{aligned} \tag{1}$$

where S is the conditional survival function of the duration of spell ℓ.

The conditional density probability function of (U_ℓ, X_ℓ) can be rewritten

$$\begin{aligned} f(u, j \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) &= \pi_j(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) \\ &\quad f(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta), \end{aligned}$$

where

$$\pi_j(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta) = \frac{h_j(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)}{\sum_{k \in E} h_k(u \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)},$$

is the conditional transition probability and f is the conditional p.d.f. of the duration between successive financial events.

π_j is the conditional probability to observe a financial event belonging to category j (buy orders for instance) given the past realizations of the process, the observed duration u and the last characteristics of the market (spread, quantities at the quotes).

3.2. Likelihood function

The contribution to the likelihood function of a the realization for a given trading day is

$$L(\theta) = \prod_{\ell=1}^{n+1} f(y_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta),$$

where $y_\ell = (u_\ell, x_\ell)$.

Therefore, using expression (1), a contribution of a given trading day to the likelihood is given by

$$L(\theta) = \prod_{\ell=1}^{n+1} h_{x_\ell}(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)^{\delta_\ell} \prod_{j \in E} S_j(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)$$

where u_ℓ is the duration of spell ℓ and

$$\delta_\ell = \begin{cases} 1, & \text{if } \ell \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

is an indicator of the right-censoring.

This contribution can be re-written

$$L(\theta) = \prod_{j \in E} L_j(\theta)$$

where

$$L_j(\theta) = \prod_{\ell=1}^{n+1} h_j(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta)^{\delta_{\ell,j}} S_j(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta),$$

and

$$\delta_{\ell,j} = \begin{cases} 1, & \text{if } x_\ell = j, \\ 0, & \text{otherwise.} \end{cases}$$

$\ell = 1, \ldots, n+1$.

4. ESTIMATION

We have to consider the estimation of the joint distribution of the spell duration and the transition probabilities given the duration of the spell. Bawens and Giot (2000) have considered then a class of financial duration models taking explicitly into account the positivity of financial durations : the Log-ACD models. We are going first to model parametrically the distribution of the conditional transition probabilities taking into account potential clustering in successive categories of events. Then, we are going to model parametrically and semi-parametrically the conditional distribution of the spell duration considering a Log-ACD model.

4.1. Transition probabilities estimation

Let $p_1(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_1)$ denote the conditional probability of a transaction at the end of spell ℓ.

Let $p_2(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_2)$ denote the conditional probability of a transaction at the midpoint of the spread given that the financial event occurring at time τ_ℓ is a trade.

Let $p_3(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_3)$ denote the conditional probability of a buy given that the financial event occurring at time τ_ℓ is a trade with a price different from the midpoint of the spread.

Let $p_4(u, y_0, \ldots, y_{\ell-1}, z; \gamma_4)$ denote the conditional probability of a modification of the depths given that the financial event occurring at time τ_ℓ is a quote.

Let $p_5(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_5)$ denote the conditional probability for an increase of the spread given that the financial event occurring at time τ_ℓ is a quote and that the spread is modified.

We assume that

$$p_j(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_j) = \frac{\exp(w_j^0 + b_j^0\ u + z'_{\ell-1}\ c_j)}{1 + \exp(w_j^0 + b_j^0\ u + z'_{\ell-1}\ c_j)},$$

where $j = 1, \ldots, 5$. w_j^0, b_j^0 and c_j are parameters. Let γ_j denote the corresponding vector of parameters.

$z_{\ell-1}$ is a vector of characteristics of the market for spell ℓ and category j of events. $z_{\ell-1}$ can consist in the past volumes proposed or demanded, the past spreads, for instance[2].

Here, we assume that the conditional probability for an event of category k is given by the expression

$$\begin{aligned}
&\pi_k(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma) \equiv \pi_k(u_\ell, z_{\ell-1}; \gamma) \\
&= \prod_{j=1}^{5} p_j(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_j)^{d_{j,k}} \; (1 - p_j(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_j))^{d'_{j,k}},
\end{aligned}$$

where $d_{1,k} = \mathbb{1}[k \in \{1,2,3\}]$, $d'_{1,k} = \mathbb{1}[k \in \{4,5,6\}]$, $d_{2,k} = \mathbb{1}[k = 1]$, $d'_{2,k} = \mathbb{1}[k \in \{2,3\}]$, $d_{3,k} = \mathbb{1}[k = 2]$, $d'_{3,k} = \mathbb{1}[k = 3]$, $d_{4,k} = \mathbb{1}[k = 4]$, $d'_{4,k} = \mathbb{1}[k \in \{5,6\}]$, $d_{5,k} = \mathbb{1}[k = 5]$, $d'_{5,k} = \mathbb{1}[k = 6]$.

$k = 1$ represents a "trade at the midpoint"; $k = 2$ represents a "buy"; $k = 3$ represents a "sell"; $k = 4$ represents a "quote without modification of the spread"; $k = 5$ represents a "quote with an increase of the spread"; $k = 6$ represents a "quote with a decrease of the spread".

4.2. Parametric estimation of the duration distribution

The model :

$$U_\ell = \exp(\psi_\ell)\ \epsilon_\ell,$$

where

$$\psi_\ell = w + \sum_{k=1}^{q} a_k\ \psi_{\ell-k} + \sum_{i=1}^{p} b_i\ u_{\ell-i},$$

and ϵ_ℓ are i.i.d. Weibull(β,α) with $\beta = \Gamma(1 + \frac{1}{\alpha})^\alpha$, $\alpha > 0$.

This is the Log-ACD(p,q) model proposed by Bauwens and Giot (2000).

We have, $E[\epsilon_\ell] = 1$ and $E[u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}] = \exp(\psi_\ell)$.

The density of ϵ is given by the expression :

$$f(\epsilon) = \beta\ \alpha\ \epsilon^{\alpha-1} \exp\{-\beta\ \epsilon^\alpha\},$$

and the conditional density of the duration of spell ℓ is :

$$f(u) = \frac{\beta}{\exp(\psi_\ell)^\alpha}\ \alpha\ u^{\alpha-1} \exp\{-\frac{\beta}{\exp(\psi_\ell)^\alpha}\ u^\alpha\},$$

The conditional hazard function of the duration of spell ℓ is :

$$h(u) = \frac{\beta}{\exp(\psi_\ell)^\alpha}\ \alpha\ u^{\alpha-1},$$

If $0 < \alpha < 1$ the conditional hazard function is decreasing with u and, if $\alpha > 1$, the conditional hazard is increasing.

Let Q denote the total number of categories for financial events. In the empirical application, $Q = 6$. Let $\pi_k(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}, \gamma)$ denote the conditional probability to

[2]This specification is distinct from the one proposed by Russel and Engle (1998). Note that the conditional probabilities can depend on the actual - for spell ℓ - value of the duration.

observe a financial event belonging to category k ($k \in E$). We have

$$\pi_k(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}, \gamma) =$$
$$\prod_{j=1}^{5} p_j(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_j)^{d_{j,k}} \; (1 - p_j(u, y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \gamma_j))^{d'_{j,k}}.$$

The likelihood function can be written :

$$L(\theta) \;=\; \prod_{\ell=1}^{n+1} f(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta_1)^{\delta_\ell} \; S(u_\ell \mid y_0, \ldots, y_{\ell-1}, z_{\ell-1}; \theta_1)^{1-\delta_\ell}$$
$$\prod_{k=1}^{Q} \pi_k(u_\ell, y_0, \ldots, y_{\ell-1}, z_{\ell-1}, \gamma)^{x_{\ell,k}}$$

where $x_{\ell,k}$ is equal to 1 if and only if, at the end of spell ℓ, a financial event belonging to category k is observed.

4.3. Semi-parametric estimation of the duration distribution

Let us consider the conditional model where the duration is given by

$$U_\ell = \exp(\psi_\ell) \; \epsilon_\ell, \tag{2}$$

where

$$\psi_\ell = w + \sum_{k=1}^{q} a_k \; \psi_{\ell-k} + \sum_{i=1}^{p} b_i \; u_{\ell-i},$$

and the ϵ_ℓ are i.i.d. and $E[\epsilon_\ell] = 1$ (see Bauwens and Giot, 2000).

Let us assume that the distribution of ϵ_ℓ does not depend on the explanatory variables (in particular, past realizations of the durations $u_1, \ldots, u_{\ell-1}$).

The conditional expectation of the duration U_ℓ is

$$\mathrm{E}[U_\ell \mid u_1, \ldots, u_{\ell-1}] = \exp(\psi_\ell) = \exp(w + \sum_{k=1}^{q} a_k \, \psi_{\ell-k} + \sum_{i=1}^{p} b_i \, u_{\ell-i}),$$

where w, a_k and b_k are parameters.

The parameters of the model in the conditional mean can be estimated minimizing the sum of the squares of the differences between the observed duration and its conditional expectation.

We can minimize

$$\sum_{\ell=\ell_0}^{n} (u_\ell - \exp(\psi_\ell))^2,$$

with respect to the vector of parameters θ_1, where $\theta_1 \in \Theta_1$ and $\ell_0 = 1 + max\{p, q\}$.

Now, in order to introduce characteristics of the market, let us consider a Log-ACD(p,q) model for the duration (see Bauwens and Giot, 2000):

$$E(u_\ell) = \exp(\psi_\ell) = \exp(\varphi_\ell + z'_{\ell-1} \, c), \tag{3}$$

where $\psi_\ell = \varphi_\ell + z'_{\ell-1} \, c$ and

$$\varphi_\ell = w + a_1 \; \varphi_{\ell-1} + \ldots + a_q \; \varphi_{\ell-q} + b_1 \; u_{\ell-1} + \ldots + b_p \; u_{\ell-p},$$

where a_k, b_k are parameters. $z_{\ell-1}$ is a vector of characteristics of the market at time $\tau_{\ell-1}$ and c is a vector of parameters.

The parameters of the conditional expectation of the duration between financial events can be estimated maximizing a quasi-likelihood function constructed using the following contribution for a given trading day (the last observation of the trading day is omitted)[3]:

$$L_p(\theta_1) = \prod_{\ell=\ell_0}^{n} \exp(-\psi_\ell)\ \exp(-\exp(-\psi_\ell)\, u_\ell). \tag{4}$$

Let $\hat{\theta}_1$ denote the quasi-maximum likelihood estimator of θ_1 obtained by maximizing the conditional quasi likelihood function. We have $\sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{L} N(0, V)$ (see, Engle 2000). The asymptotic variance-covariance matrix can be estimated using the Bollerslev-Wolldridge (1992) estimator $\hat{V} = \hat{J}^{-1}\ \hat{I}\ \hat{J}^{-1}$ where

$$\hat{J} = -\frac{1}{n_0} \sum_{\ell=\ell_0}^{n} \frac{\partial^2 \ln(L_{p,\ell}(\hat{\theta}_1))}{\partial\theta_1 \partial\theta_1'},$$

$$\hat{I} = \frac{1}{n_0} \sum_{\ell=\ell_0}^{n} \frac{\partial \ln(L_{p,\ell}(\hat{\theta}_1))}{\partial\theta_1} \frac{\partial \ln(L_{p,\ell}(\hat{\theta}_1))}{\partial\theta_1'},$$

and $L_{p,\ell}(\theta_1)$ is the corresponding contribution to the quasi likelihood function of the spell ℓ. n, here, is the total number of completed spells and $n_0 = n - \ell_0 + 1$. Semi parametric estimation of the of duration models have been considered by Engle and Russell (1998), Engle (2000), Drost and Werker (2000, 2001), Gouriéroux and Jasiak (2001).

4.4. Conditional hazard functions

The semi-parametric estimation of the parameters of the conditional expectation of the duration between financial events allows to obtain consistent estimates of these parameters whatever the true conditional distribution of the duration. However, it is then interesting, for the interpretation of the results, to obtain a semi parametric estimation of the conditional hazard function of the duration between financial events using the estimated vector of parameters. Similarly, we can obtain a semi parametric estimation of the conditional hazard functions of latent durations corresponding to the categories of financial events.

4.4.1. Intertrades duration conditional hazard function As the latent durations are conditionally distributed as a Log-ACD model, then, the conditional expectation of the duration has the expression (see equation 2)

$$E[U_\ell \mid \psi_\ell] = q_\ell,$$

where $q_\ell = \exp(\psi_\ell) = \exp(\varphi_\ell + z_{\ell-1}' c)$ and $\varphi_\ell = w + a_1\ \varphi_{\ell-1} + \ldots + a_q\ \varphi_{\ell-q} + b_1\ u_{\ell-1} + \ldots + b_p\ u_{\ell-p}$.

Let $\hat{q}_\ell$ denote the corresponding estimated index $\hat{q}_\ell = \exp(\hat{\psi}_\ell) = \exp(\hat{\varphi}_\ell + z_{\ell-1}' \hat{c})$ and $\hat{\varphi}_\ell = \hat{w} + \hat{a}_1\ \hat{\varphi}_{\ell-1} + \ldots + \hat{a}_q\ \hat{\varphi}_{\ell-q} + \hat{b}_1\ u_{\ell-1} + \ldots + \hat{b}_p\ u_{\ell-p}$.

[3] Model (3) can be considered as a parametric model in the case where ϵ_ℓ are i..i.d. exponentially distributed with parameter equal to 1.

Let $\hat{q} = \exp(\psi)$ denote the estimated index for a fixed vector of characteristics of the market. We assume that $\psi = \hat{\varphi} + \bar{z}'\hat{c}$ where $\hat{\varphi} = \frac{1}{n_0}\sum_{\ell=\ell_0}^{n} \hat{\varphi}_\ell$ and $\bar{z}$ is a fixed value for the vector of characteristics of the market. In practice, the spread, the quantities at the quotes and the volume traded are fixed to their empirical mean values and the category of the last event is fixed.

The conditional probability density function of a duration U (of a given spell) can be estimated using a non parametric Kernel estimator

$$\hat{f}(u \mid q) = \frac{\frac{1}{h_1}\sum_{\ell=\ell_0}^{n} K(\frac{u-u_\ell}{h_1})\ K(\frac{q-\hat{q}_\ell}{h_2})}{\sum_{\ell=\ell_0}^{n} K(\frac{q-\hat{q}_\ell}{h_2})}, \tag{5}$$

where h_1 and h_2 are bandwidths.

The conditional survival function of the duration U can be estimated using the following Kernel estimator

$$\hat{S}(u \mid q) = \frac{\sum_{\ell=\ell_0}^{n} \mathbb{1}[u_\ell \geq u]\ K(\frac{q-\hat{q}_\ell}{h_2})}{\sum_{\ell=\ell_0}^{n} K(\frac{q-\hat{q}_\ell}{h_2})}, \tag{6}$$

where h_2 is a bandwidth.

As the conditional hazard function of the duration U is the ratio of the probability density function and the survival function, we can construct a Kernel estimator of the hazard function using equations (5) and (6). We obtain

$$\hat{h}(u \mid q) = \frac{\frac{1}{h_1}\sum_{\ell=\ell_0}^{n} K(\frac{u-u_\ell}{h_1})\ K(\frac{q-\hat{q}_\ell}{h_2})}{\sum_{\ell=\ell_0}^{n} \mathbb{1}[u_\ell \geq u]\ K(\frac{q-\hat{q}_\ell}{h_2})}. \tag{7}$$

Considering $q = \hat{q}$ it is then possible to draw the probability density function (5), the survival function (6) and hazard function (7) corresponding to the duration between financial events (see figures (5), (6) and (4)).

4.4.2. Conditional hazard functions of latent durations Let h_j denote the conditional hazard function of the latent duration necessary to observe the category j of financial events.

An estimate of $h_j(u \mid q)$ can be obtained using

$$\hat{h}_j(u \mid q) = \hat{h}(u \mid q)\ \ \pi_j(u; \bar{z}; \hat{\gamma}),$$

where the expression of $\hat{h}(u \mid q)$ is given by equation (7). π_j is a function of the characteristics of the market z, the duration between financial events u, and the estimated vector of parameters $\hat{\gamma}$ (see section 4.1).

We obtain,

$$\hat{h}_j(u \mid q) = \frac{\dfrac{1}{h_1}\displaystyle\sum_{k=\ell_0}^{n} K(\frac{u-u_k}{h_1})\ K(\frac{q-\hat{q}_k}{h_2})}{\displaystyle\sum_{k=\ell_0}^{n} 1\!\!1[u_k \geq u]\ K(\frac{q-\hat{q}_k}{h_2})}\ \pi_j(u;\bar{z};\hat{\gamma}). \quad (8)$$

In practice, the index $\hat{q}_k$ depends on the past realizations of the process and a set of characteristics of the market at the time of the last event. The bandwidths h_1 and h_2 are fixed in empirical applications using the normal scale rule (see section 3). q is fixed to $\hat{q}$ in equation (8) in order to draw the hazard function.

5. RESULTS

In standard microstructure models like Kyle (1985), Glosten and Milgrom (1985), O'Hara and Oldfield (1986), because the authors assume that a new information is arrived, the time plays no role. In these models, there is no uncertainty on the information. In the Easley and O'Hara (1992) model, the observed price will evolve because the specialist learn by the observation of the market. The arrival of trades reveals information to the market participants. In this model, the presence or absence of trade may provides some information on the value of the asset[4].

The consequence of such a modeling is that the arrival of trades on the market may convey information to the market participants. The importance of the interval of time between trades is informative on the possible arrival of a new information on the price of the asset. This has been underlined, in particular, by Engle and Russel (1998) who model the arrival times of the trades using an ACD model in order to take into account the existence of some clustering phenomenon. Indeed, when a sequence of trades is observed on a short period of time, the conditional expectation of the duration until the next trade should be significantly reduced. When we model the inter quotes duration we should consider the impact of the arrival of intermediate trades that can modify the expression of the remaining time until the next quote (see Kamionka, 2000).

In the Easley and O'Hara (1992) model, the spread and the volume traded are correlated with the inter trade durations. However, the spread and the quantities available at the best quotes may have an impact on the activity of the market. For instance, for the Paris Bourse, Biais, Hillion and Spatt (1995) and Bisière and Kamionka (2000) have found that the quantity proposed or demanded at the best quotes have an effect mainly on the process of arrival of the limit orders. Moreover, the importance of the spread should have an impact on the arrival rate of trades.

The data we use is extracted from the "Trades and Quotes" data set (TAQ) of the New York stock exchange (NYSE). This data base allows to obtain, on one hand, the trades and, on the other hand, the quotes posted by the specialists. We use the trades and the quotes recorded from 9:30 to 16:00 for IBM which is one of the most actively negotiated stock on this market. The data we use are relative to the period which goes from September till November 1998 and concerns 63 days. In our model we have to identify the arrival times of the events. However we can have groups of financial events with the

[4] For a study of market participants behavior in the presence of an asymmetry of information, see Biais et alii (1997), Easly et alii (1997), Hasbrouck and Sofianos (1993).

Table 1. Duration between financial events. Standard errors in parentheses.

	Log-ACD(1,1) Semi Para	Log-ACD(2,2) Semi Para	Log-ACD(2,2) Semi Para	Log-ACD(2,2) Parametric
w	-0.0151	-0.0178	-0.0228	-0.0223
	(0.0013)	(0.0012)	(0.0017)	(0.0016)
a1	0.9811	0.8960	0.8307	0.8238
	(0.0018)	(0.0717)	(0.0862)	(0.0795)
a2		0.0814	0.1460	0.1535
		(0.0726)	(0.0874)	(0.0805)
b1	0.0145	0.0021	0.0016	0.0007
	(0.0013)	(0.0026)	(0.0026)	(0.0023)
b2		0.0150	0.0161	0.0165
		(0.0028)	(0.0028)	(0.0026)
Spread			1.2224	1.2404
			(0.1628)	(0.1607)
Trade At			0.4055	0.4156
			(0.0299)	(0.0294)
Buy			0.0892	0.0947
			(0.0230)	(0.0224)
Sell			0.2249	0.2309
			(0.0232)	(0.0226)
Depth Revision			0.2498	0.2475
			(0.0294)	(0.0262)
Spread Increase			0.1222	0.1277
			(0.0301)	(0.0302)
Spread×Trade At			-2.0738	-2.1080
			(0.2689)	(0.2661)
Spread×Buy			-1.2012	-1.2176
			(0.1942)	(0.1909)
Spread×Sell			-1.9452	-1.9674
			(0.2019)	(0.1981)
Spread×Depth Rev.			-1.1384	-1.1149
			(0.2437)	(0.2262)
Spread×Spread Inc.			-1.5525	-1.5940
			(0.2330)	(0.2328)
QA			0.0027	0.0028
			(0.0007)	(0.0007)
QB			0.0051	0.0052
			(0.0008)	(0.0008)
QA×Trade			-0.0031	-0.0032
			(0.0009)	(0.0009)
QB×Trade			-0.0052	-0.0052
			(0.0010)	(0.0010)
Quantity			0.0016	0.0016
			(0.0006)	(0.0005)
α				-0.0588
				(0.0046)

same time stamp. We have chosen to select only one event in each group by randomly drawing one of it. When several trades occurs at the same time, the corresponding quantities traded are aggregated (the number of trades is reduced by 0.75%). When several quotes are observed with the same time stamp we retain those associated to the best ask or bid prices and the corresponding quantities are aggregated (the number of quotes is then reduced by 5.12%). As Bauwens and Giot (2000) the durations are put together as we had a single day and the overnight durations are not used. We have considered a total of 190788 durations between events and the corresponding categories (trades or quotes) such that the category of the last financial event is available (at least two quotes were observed previously for the corresponding trading day).

We have maximized the objective function $L_p(\theta_1)$ (see equation 4) with respect of the vector of parameters θ_1 in order to obtain the QML Estimator of the parameters of the conditional mean duration between financial events (trades or quotes). The estimation results are presented in the table 1. Each column corresponds to a given specification. Each row of the table is associated to a given parameter of the process.

For a specification, the parameters are associated, respectively, to

- a constant,
- the two auto-regressive components of the conditional expectation,
- the lagged values of the enter events durations,
- the spread times 100,
- an indicator that the last event is a trade at the midpoint of the spread,
- an indicator that the last event is a buy,
- an indicator that the last event is a sell,
- an indicator that the last event is a quote without modification of the spread,
- an indicator that the last event is a quote with an increase of spread,
- the spread times the indicator that the last event is a trade at the midpoint of the spread (times 100),
- the spread times the indicator that the last event is a buy (times 100),
- the spread times the indicator that the last event is a sell (times 100),
- the spread times the indicator that the last event is a quote without modification of the spread (times 100),
- the spread times the indicator that the last event is a quote with an increase of the spread (times 100),
- the quantity at the last ask quote,
- the quantity at the last bid quote,
- the quantity at the last ask quote times an indicator that the last event is a trade,
- the quantity at the last bid quote times an indicator that the last event is a trade,
- the last quantity traded.

The quantities at the bid and ask quotes are divided by 10. The quantities traded is divided by 1000.

Under our assumptions, the quasi maximum likelihood estimator of the parameter (θ_1) of the conditional expectation of the duration between financial events is consistent whatever the distribution of the "baseline" duration ϵ_ℓ in equation (2) (see columns 1 to 3, table 1).

We have estimated too a model for the duration between financial events assuming that the "baseline" durations ϵ_ℓ were i.i.d. Weibull(β, α), where $\beta = \Gamma(1 + \frac{1}{\alpha})^\alpha$ (see table 1, last column). We can obtain the value of the maximum likelihood estimator by maximizing separately the part of the likelihood function relative to the spell durations with respect to θ_1 and α.

Table 2. Conditional transition probabilities. Standard errors in parentheses. † *Given the arrival of a trade.* ‡ *Given the arrival of a trade at a price different from midquote.* § *Given the arrival of a new quote.* * *Given the arrival of a new quote with a spread modification.*

	Trade	Trade At Midpoint †	Buy ‡	Depth Revision §	Spread Increase *
	p_1	p_2	p_3	p_4	p_5
Constant	-0.5950	-1.7144	-0.4075	-0.2200	4.1140
	(0.0250)	(0.0435)	(0.0379)	(0.0350)	(0.0893)
Duration	0.1013	-0.0592	-0.0230	-0.0723	0.0072
	(0.0034)	(0.0065)	(0.0053)	(0.0053)	(0.0075)
Spread	0.4595	0.3998	0.4731	-0.0847	-3.1103
	(0.0248)	(0.0374)	(0.0354)	(0.0369)	(0.0926)
Trade At	-0.0358	2.9775	0.5980	1.3250	-3.2882
	(0.0478)	(0.0796)	(0.1029)	(0.0751)	(0.1303)
Buy	-0.1229	-0.9794	1.4505	0.5281	-1.3955
	(0.0335)	(0.0888)	(0.0589)	(0.0478)	(0.1101)
Sell	-0.1214	-1.2724	-0.8507	0.8743	-0.9672
	(0.0331)	(0.0798)	(0.0563)	(0.0492)	(0.1123)
Depth Revision	0.1583	0.0646	0.2991	1.7823	-1.0830
	(0.0311)	(0.0553)	(0.0478)	(0.0456)	(0.1200)
Spread Increase	0.7507	3.3678	0.5421	1.0371	-3.6414
	(0.0378)	(0.0716)	(0.0637)	(0.0626)	(0.1377)
Spread×Trade at	-0.2899	-0.9048	-0.4683	-1.1387	1.7512
	(0.0432)	(0.0667)	(0.0843)	(0.0728)	(0.1266)
Spread×Buy	-0.2820	-0.5247	-0.1417	-0.7870	0.8959
	(0.0295)	(0.0684)	(0.0489)	(0.0462)	(0.1072)
Spread×Sell	-0.3203	-0.2443	-0.4894	-1.3527	0.4620
	(0.0298)	(0.0582)	(0.0468)	(0.0508)	(0.1112)
Spread×Depth Rev.	-0.3627	0.0028	-0.0712	-1.0383	-0.3836
	(0.0316)	(0.0499)	(0.0471)	(0.0482)	(0.1390)
Spread×Spread Incr.	-0.4520	-2.1583	-0.4121	-0.7252	1.8983
	(0.0308)	(0.0578)	(0.0468)	(0.0516)	(0.1220)
QA	0.0013	-0.0002	-0.0213	0.0088	-0.0152
	(0.0009)	(0.0015)	(0.0015)	(0.0013)	(0.0029)
QB	0.0080	-0.0002	0.0122	0.0134	-0.0186
	(0.0012)	(0.0019)	(0.0018)	(0.0019)	(0.0045)
QA×Trade	0.0068	-0.0012	0.0032	0.0006	0.0011
	(0.0012)	(0.0027)	(0.0023)	(0.0019)	(0.0036)
QB×Trade	0.0054	-0.0023	0.0074	-0.0000	0.0047
	(0.0016)	(0.0033)	(0.0028)	(0.0025)	(0.0054)
Quantity	-0.0051	-0.0026	-0.0003	-0.0258	0.0312
	(0.0009)	(0.0018)	(0.0016)	(0.0017)	(0.0028)

5.1. Conditional expectation of the duration

According to market microstructure theories, the market marker try to infer on the motivations of traders (liquidity or information) by looking at the timing of the trades. When it becomes clear that some trader has a private information on the value of the asset, the specialist will react by modifying the quotes.

Consequently, the interval of time between past financial events should have an impact on the distribution of the next event (trade or quote). The results show that the estimates of the parameter b_2 associated to the lagged observed duration ($u_{\ell-2}$) is significantly different from zero (see columns 2 and 3, table 1). This result is consistent with previous empirical works on similar data sets (see Engle and Russel (1998), Engle (2000), Bauwens and Giot (2000), Kamionka (2000)). This reflects the existence of a clustering phenomenon in financial events. The presence of a lag can be explained by the fact that a trade is more likely to be followed by a quote and reciprocally. This result is consistent with the empirical implications of market microstructure theories. The sum of the parameters of the autoregressive part of the conditional expectation is close to 1 showing that the process of financial events has a strong persistence.

We have considered the effect of the spread, the spread times an indicator that the last event is a trade, distinguishing the different categories of trades (buy, sell, trade at the midpoint of the spread) and the effect of the spread times an indicator that the last event is a quote distinguishing the different categories of quotes (without modification of the spread, with a spread increase). The parameter of the spread when the last event was a trade is negative and significantly different form zero whatever the category of the trade (see column 3, table 1). The arrival of a trade can be a signal that a new information on the value of the asset is arrived, particularly when the spread is large. In this case, the conditional expectation of the duration is significantly reduced. This may correspond to a reaction of other traders (imitation strategies or orders splitting strategies) or a reaction of the specialist that infers the existence of an asymmetry of information. This can reflect a correlation of the successive trades or an increase of the intensities of the quotes when the trades are frequent. This is consistent with the results of Hasbrouck (1991) who finds that the effect of a trade that arrives when the spread is large is more important than the effect of a trade arriving when the spread is small.

When the last event recorded on the market is a modification of the quotes corresponding to a spread decrease and the spread is large, the conditional expectation of the duration of is relatively more important. This may indicate that the specialist, after a lack of trade following a period of intense trading, decreases the spread successively in order to observe the reactions of the traders[5]. He tries to infer on the motives of the traders by modifying the spread. In particular, when the spread decreases but stay relatively large, the arrival rate of trades does not increase because it is still costly, in this case, to consume the liquidity. As the spread is large, it is not urgent for the specialist to modify its quotes. The empirical implication 2 ("Timing of Trades") is verified but the results indicate that the specialist will consider continuously the reaction of traders to the modification of the quotes before continuing to decrease the spread.

[5] Indeed, the examination of the conditional probabilities let us conclude that the probability of a spread decrease, given the arrival of a new quote, is greater after a decrease of the spread associated to a large spread (see table 2).

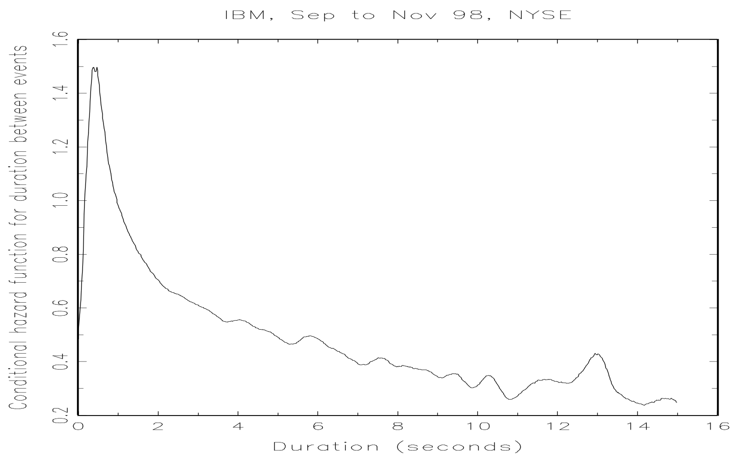

Figure 4. Conditional hazard function.

The estimated parameter of the spread times the indicator that the last event is a new quote associated to an increase of the spread is negative and significantly different from zero. When the specialist infer that a new information is more likely to be arrived on the market, he is going to increase the spread successively, observing each time the reactions of the participants to the market. He will react relatively quickly to the arrival of a new information.

The parameters associated to the presence of a trade are positive and significantly different from zero. This results is not consistent with the empirical implication 3 (in the presence of a trade the specialist will modify more frequently its quotes). Indeed, the positivity of the parameters associated to the trades is obtained because we control for the negative impact of the arrival of the corresponding categories of trades obtained when the spread is wide.

The parameter associated to the volume traded is positive and significantly different from zero. Consequently, the quantity traded alone has a significant effect on the mean waiting time between financial events. However the information content of the quantities is generally weaker than the one of the prices. An important volume traded is similar to an increase of the spread. Indeed, when the traded quantity is large, the probability to consume the liquidity at the best price on the other side of the market is important. When the spread becomes larger, there is a renewal of the competition among the providers of the liquidity. However, has the spread increases its becomes more costly to consume the liquidity. Finally, the arrival of a new quote will depend on the quantities available at the best prices after the last quantity was traded. This interpretation of the results is consistent with the estimated sign of the parameter of the indicator of the effect of the spread and the last event was a quote without modification of the spread. Indeed, the

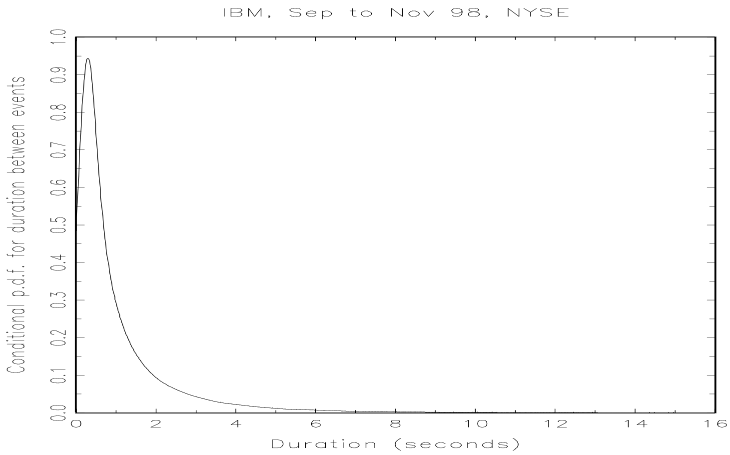

Figure 5. Conditional probability density function.

estimated parameter is negative indicating that the specialist is going to react quickly by decreasing the spread or that there is an increase of the competition among the suppliers of the liquidity (see section 5.2).

The effect of the quantities at the quotes is different according to the type of the last event recorded (see table 1). When the last event is a quote, the effect is positive and significantly different from zero. Given that the last event a trade, the effects of these quantities are negative and significantly different from zero. When the quantities proposed and demanded are large and the last event is a trade, the market maker will have to modify its quotes quickly in order to take the arrival of the last trade into account. This cannot be explained by an increase of the competition among the investors. Indeed, the results show that when the last event is a quote and the quantities proposed or demanded are important, the mean waiting time until the next event is increased. This may indicate a difference in the intensity of the reaction of the specialist face to the arrival of the two kinds of events (trades and quotes). The market marker reacts more intensively to the potential presence of an asymmetry of information on the market rather than to a more intensive competition among liquidity providers.

The estimated parameters of the duration model between financial events for a Weibull distribution are given in the table 1 (last column). The sign and the importance of each parameter is similar to what we have obtained using a semi parametric model without specifying the distribution of the duration. The parameter of the "baseline" duration is equal to $\hat{\alpha} = \exp(\hat{\xi}) = \exp(-0.0588) = 0.9429$. The estimated standard error of $\hat{\xi}$ is 0.0046 (table 1, column 4), so the corresponding parameter is significantly different from zero. This means that the hazard function of the duration is monotone decreasing. The characteristics of a given event will have an effect on the distribution of the duration

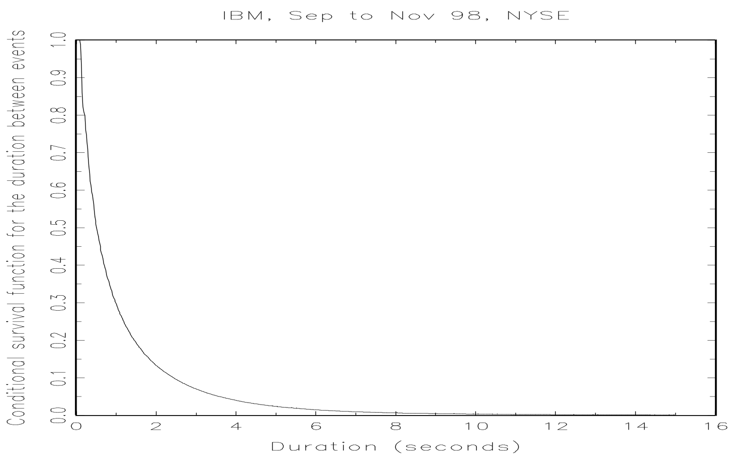

Figure 6. Conditional survival function.

necessary to observe, in particular, a reaction of the market maker. As the hazard function is decreasing, the information contained in the arrival of the last financial event becomes less and less relevant as time passes.

5.2. Conditional transition probabilities

The definition of the variables are the same as for the specifications of the inter events durations. The observed spread is, here, multiplied by 1000. We have considered 6 categories of events in the model : a trade at the midpoint (category 1)[6], a buy (category 2), a sell (category 3), a new quote without modification of the spread (category 4), a new quote with a spread increase (category 5), a new quote with a spread decrease (category 6).

Most of the parameters are statistically significant at the 5% level. The main exceptions concern the quantities at the ask and bid quotes, the volume traded or the impact of a revision of the quote sizes. Consequently, the effect of the quantities is generally smaller than the impact of the prices.

The estimated parameters of the conditional probability for a trade are given in table 2 (column 1). The parameter associated to the current duration (i.e. b_1^0) is positive and significantly different from zero. This result is consistent with the hypothesis that time plays an important role. Consequently, when the specialist wants to react he does it quickly.

[6]We do not have considered here the "tick test procedure" proposed by Lee and Ready (1991) in order to classify the trades inside the spread. Here, we are interested in considering the trades at the midquote separately in order to analyze their impact of the behaviour of the specialist.

The estimated parameters associated to the indicators that the last event is a trade at midpoint of the spread, a buy and a sell are negative. This is consistent with market microstructure theories and empirical implication 3 ("market activity"). For instance, Easley and O'Hara (1992) note that trades can cause the price quotes to change. As noted previously for the conditional expectation of the duration between financial events, when the spread is wide and the last event is a trade, the market maker is more likely to react because the arrival of such a trade reveals that a new information has occurred on the value of the asset. Indeed, the estimated parameters of the spread times the indicator of a trade at midpoint of the spread, the spread times the indicator of a buy and the spread times the indicator of a sell are negative and significantly different from zero. The specialist will react more intensively to the arrival of such events in such a context of the market : he will modify the quotes.

The estimated values of the parameters associated to the indicators that the last event corresponds to a modification of the quantities proposed or demanded or to an increase of the spread are positive and significantly different from zero. This result indicates that when the specialist modifies its quotes (increase of the spread or depths modification) the transactions becomes relatively more frequent. The estimated parameters associated to the spread is positive indicating that when the spread is large but decreasing, the conditional probability that a trade occurs increases.

When the last event is an increase of the spread and if the spread is large, then the conditional probability to observe a new quote is greater (see, table 2, column 1). In this case, the conditional probability to observe a modification of the spread is larger (column 4, table 2) and the conditional probability to observe a new quote associated to an increase of the spread is much more important given we have the arrival of a new quote associated to a modification of the spread (column 5, table 2). Moreover, we find the same kind of results when the spread is large and the last event is a trade at the midpoint of the quotes (indeed a spread increase favor the arrival of trade inside of the quotes). These results indicate that when the specialist decides to increase the spread, he is going post a sequence of successive modifications of the quotes.

The parameters associated to the quantity traded is negative and significantly different from zero (this is consistent with empirical implication 4 : "information content of the volume"). Consequently, after an important transaction the market maker will react modifying the quote more frequently but the information content of the volume is less important than the information content of the transaction price.

The quantities at the quotes have a positive effect on the conditional probability of a trade. Singularly, the parameters associated to the corresponding volumes proposed or demanded when the last event is a trade are positive and significantly different from zero. A wide spread associated to the arrival of a trade is an indicator of an asymmetry of information between traders and the market maker. Large quantities observed in the presence of a trade may indicate simply that the liquidity of the market is important. Once again, the information content of the quantities is less important than the information content of the prices.

The spread times the indicator that the last event is a trade (trade at the midpoint of the spread, buy or sell) has a negative and significant effect on the probability to observe a new quote without a modification of the spread given that the specialist has posted a new quote (see column 4, table 2). These variables have positive impacts on the conditional probability to observe a new quote with an increase of the spread (column 5). The arrival of a trade when the spread is large is the signal that a new information

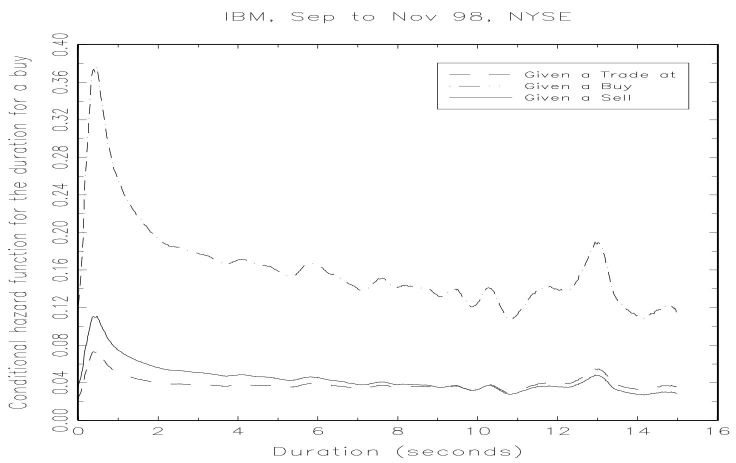

Figure 7. Conditional hazard functions.

is arrived on the market. The arrival of such an information is the signal for the market maker of the existence of an asymmetry of information among the participant to the market. The specialist will react to this signal by posting a new quote and by increasing the spread.

Conditionally to the arrival of a trade with a price different from the midpoint of the spread (see column 3, table 2), the probability to observe a buy is significantly greater (smaller) when the last event is a buy (a sell). This diagonal effect is similar to the one detected by Biais, Hillion and Spatt (1995) and Bisière and Kamionka (2000) for the Paris Bourse.

Given we observe the arrival of a new quote with a modification of the spread, the probability that the specialist increases the spread is significantly greater when the last quantity traded is large (table 2, last column). Finally, the conditional probability to observe an increase of the spread is significantly greater when the last quantity traded is large (see columns 1 and 4-5, table 2). These results are consistent with the empirical implication 4 ("Information content of the volume"). A large volume traded is the signal for the specialist that a new information is arrived on the market.

Conditionally to the arrival of a new quote, the probability to observe only a modification of the quantities at the quotes is significantly greater when the last event is a trade at the midpoint of the spread (see column 4, table 2). Then, this category of events may signal to the specialist that there is no asymmetry of information.

Given that we observe a new quote with a modification of the spread, the probability to observe a spread increase is significantly smaller when the last financial event is a trade at the midpoint of the spread (column 5, table 2). Consequently, a trade at the midpoint of the spread will be interpreted differently by the specialist. It is the signal

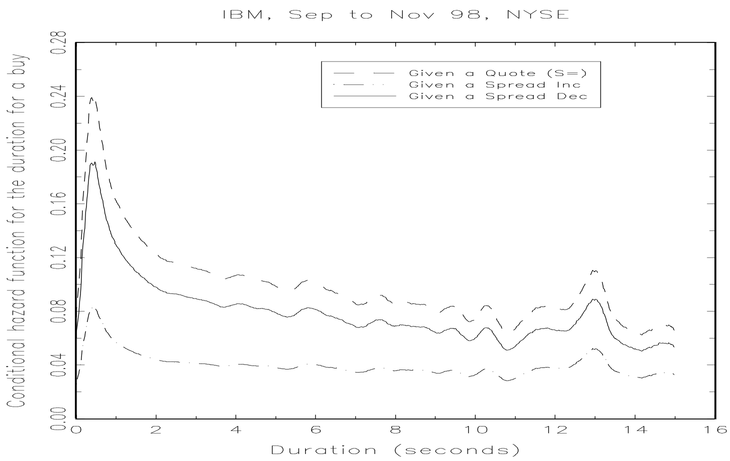

Figure 8. Conditional hazard functions.

that there is no informed trader and that he can decrease the spread or increase the quantities available at the best prices without taking an important risk.

5.3. Conditional hazard functions

We have calculated the semi parametric estimation of the conditional hazard function of the duration between financial events (see equation (7) for the expression of the estimator). The hazard function is depicted in figure 4. The estimated hazard function is calculated using sample mean value for the spread and the quantities, the empirical mean of the estimated auto-regressive part of the conditional expectation of the duration and assuming that the last event recorded is a buy. The estimated hazard function is increasing then decreasing. The first increase can correspond to a time delay necessary for the transmission of the information. The hazard function is then decreasing. Then, the arrival of the last financial event has been integrated by the individuals in their information sets.

We have calculated too the semi parametric estimation of the conditional probability density function (see equation (5)). This probability density function is drawn in figure 5. The corresponding conditional survival function (see equation (6)) is depicted in figure 6. These figures indicate that the distribution of the duration between financial events is characterized by a complex time dependence.

Using the estimated conditional hazard function of the duration between financial events, we can obtain an estimation of the conditional hazard functions of the latent durations corresponding to the categories of financial events (a trade at the midpoint of the spread, a buy, a sell, a new quote without modification of the spread, a new quote

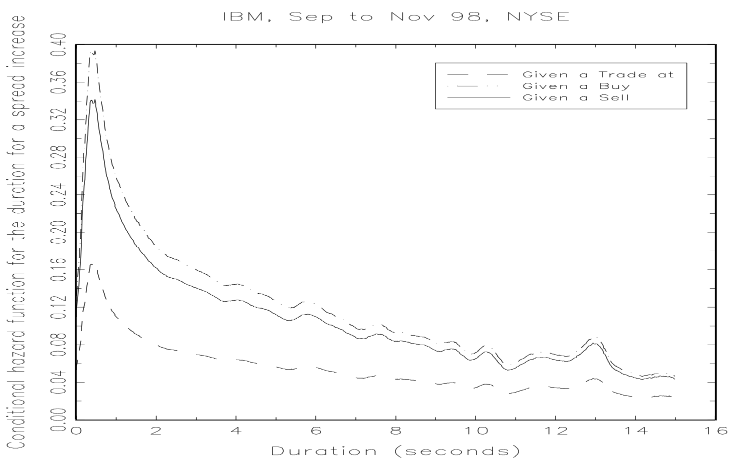

Figure 9. Conditional hazard functions.

with a spread increase and a new quote with a spread decrease). The expression of the conditional hazard function is given by the equation (8). For instance, the conditional hazard function for the duration for a new buy is depicted in figures 7 and 8 according to the category of the last financial event recorded. The hazard function is increasing then decreasing whatever the category of the last event. The arrival rate for a new buy is maximum when the last event is a buy. This can interpreted by the existence of imitation behaviors among traders or by orders splitting strategies. This result is consistent with the empirical implication 5 ("Successive trades are positively correlated"). The conditional hazard function for the duration for an increase of the spread is depicted in figures 9 and 10. This conditional hazard function is higher when the last event is a spread decrease rather than when the last event was a new quote associated to a spread increase. Consequently, the diagonal effect is not verified for quotes.

6. CONCLUSION

In this paper we study the process of the arrivals of trades and quotes for IBM on the NYSE. We use high frequency data from the TAQ database. We model jointly the inter events duration and the category of these events : trades or quotes taking into account six categories of the financial events. On such high frequency data, the arrival time of the events cannot be considered as exogeneous. In order to take into account the intraday activity of the market we realize a non parametric regression of the duration on the time of the day. We find the usual U-shape pattern of this activity : the activity is maximum at the beginning and at the end of the trading day and minimum in the middle of the day. Using a semi-parametric estimation of the conditional duration model, we find that the process of financial events has a strong persistence. We estimate the conditional

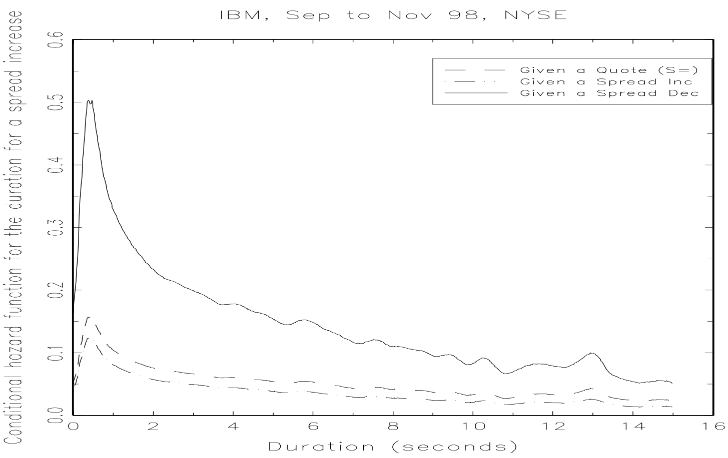

Figure 10. Conditional hazard functions.

probability for a trade given the past history of the process and the characteristics the market. We use the estimation results in order to consider the validity of some implications of market microstructure theories. For instance, we find that a wide spread, when the last event is a trade, favor the arrival of a new quote and reduce significantly the mean time necessary to observe a reaction of the market. This result is consistent with the model of Easley and O'Hara (1992) in the sense that trades make the specialist modify its quotes more intensively. However, the estimations indicates that the presence of a trade in the particular context of a large spread conveys a strong signal for the specialist. This provides evidence that the market maker will use the information on the category of the trade (buy, sell, trade at) jointly with the state of the market (spread) in order to determine the importance and the nature of its reaction. We do not find the same effect when the last event is a quote associated to a decrease of the spread. The market maker will react more intensively to the arrival of a trade at the midquote by decreasing the spread or increasing the quantities available at the quotes. So, the arrival of a trade at the midquote is the signal that, more likely, no new information is arrived on the market. Given that the next even is a trade, we find a diagonal effect with respect to the category of the trade (buy, sell, trade at). The specialist reacts generally faster than the other participants on the market when the arrival of a new information is likely. A parametric estimation of the model allows us to determine the shape of the conditional hazard function of the inter events duration. We find a negative dependence indicating that the information content of a given event is decreasing over time. We estimate then semi parametrically the conditional hazard functions of the latent durations corresponding to the different categories of financial events. These conditional hazard functions are closely related to the probabilities to observe the different categories of financial events in reaction to the realized activity of the market and the elapsed time since the last financial event.

ACKNOWLEDMENTS

I wish to thank the participants of the AFFI Conference in Paris and the Canadian Economics Association Meeting in Calgary for helpful comments. This paper was started when I was member of the GREMAQ, Université Toulouse I, France. The usual disclaimer applies.

REFERENCES

Bauwens, L., and P. Giot (2000). "The Logarithmic ACD Models : An Application to the Bid/Ask Quote Process of Free NYSE Stocks." *Annales d'Economie et de Statistique*, 60, 117–149.

Bauwens, L., and P. Giot (2001). *Econometric modelling of stock market intraday activities.* Kluwer Academic Publishers.

Biais, B., T. Foucault, and P. Hillion (1997). Microstructure des Marchés Financiers. Institutions, modèles et tests empiriques. *Collection Finance*, Puf.

Biais, B., P. Hillion, and C. Spatt (1995). "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse." *Journal of Finance*, 50, 1655–1689.

Bisière, C., and T. Kamionka (2000). "Timing of Orders, Order Agressiveness and the Order Book at the Paris Bourse." *Annales d'Economie et de Statistique*, 60, 43–72.

Bollerslev, T., and J. Wolldridge (1992). "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances." *Econometric Reviews*, 11, 143–172.

Drost, F. C., and B. J. M. Werker (2000). "Efficient Estimation in Semiparametric Time series: the ACD Model." *mimeo*, Tilburg University, The Netherlands.

Drost, F. C., and B. J. M.Werker (2001). "Semiparametric Duration Models." *mimeo*, Center, Tilburg University.

Easley, D., N. M. Kiefer, and M. O'Hara (1997). "One Day in the Life of a very Common Stock." *The Review of Financial Studies*, 3, 805–835.

Easley, D., and M. O'Hara (1992). "Time and the Process of security Price Adjustment." *The Journal of Finance*, 2, pp. 577–605.

Engle, R. F. (2000). "The Econometrics of Ultra-High Frequency Data." *Econometrica*, 1, pp. 1–22.

Engle, R. F., and J. R. Russell (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica*, pp. 1127–1162.

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.

Gouriéroux, C., and J. Jasiak (2001). Financial Econometrics. *Princeton series in Finance.*

Gouriéroux, C., J. Jasiak, and G. Le Fol (1999). "Intra-Day Market Activity." *Journal of Financial Markets*, 2, 193–226.

Härdle, W., P. Hall, and H. Ichimura (1993). "Optimal Smoothing in Single-Index Models." *The Annals of Statistics*, 1, 157–178.

Hasbrouck, J. (1991). "Measuring the Information Content of Stock Trades." *The Journal of Finance*, vol. XLVI, 1, 179–207.

Hasbrouck, J., and G. Sofianos (1993). "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists." *The Journal of Finance*, vol. XLVIII, 5, 1565–1593.

Horowitz, J. L. (1993). "Semiparametric Estimation of a Work-trip Mode Choice Model." *Journal of Econometrics*, 58, 49–70.

Kamionka, T. (2000). "La modélisation des données haute fréquence." *Le Journal de la SFdS*, 141, 167–211.

Kyle, A. P. (1985). "Continuous Auctions and Insider Trading." *Econometrica*, 6, 1315–1335.

Lee, C. M. C., and M. J. Ready (1991). "Inferring Trade Direction from Intraday Data." *Journal of Finance*, XLVI, 2, 733–746.

O'Hara, M., and G. S. Oldfield (1986). "The Microeconomics of Market Making." *Journal of Financial and Quantitative Analysis*, vol. 21, 4, 361–376.

Russell, J. R. (1999). "Econometric Modeling of Multivariate Irregurly Spaced High-Frequency Data." *mimeo*, University of Chicago.

Russell, J.R., and R.F. Engle (1998). "Econometric Analysis of Discrete-valued Irregularly-Spaced Financial Transactions Data Using A New Autoregressive Conditional Multinomial Model." *Discussion paper 98-10*, University of California, San Diego.

Veredas, D., J. Rodriguez-Poo, and A. Espasa (2001). "On The (Intradaily) Seasonality and Dynamics of the Financial Point Process: A Semiparametric Approach." *CORE Discussion Paper*, Université Catholique de Louvain.

Wand, M. P., and M.C. Jones (1995). Kernel Smoothing. *Chapman & Hall*.