MODELING HIGH-FREQUENCY
DYNAMICS OF FINANCIAL MARKETS
IN CONTINUOUS TIME: AN
EVENT-BASED APPROACH

Konstantin Tyurin
Department of Economics
Indiana University
November 2002

Abstract
This paper introduces the competing risks model as a useful and promising tool for empirical analysis of high-frequency financial data in continuous time. The emphasis is placed on the role of semiparametric estimation techniques as a flexible way to deal with observed and unobserved heterogeneity in the data. The competing risks method can be adjusted to accommodate a variety of problems in economics and finance, including optimal management of inventories in speculative markets, analysis of defaults and other qualitative risks, interactions of quotes and transactions in the markets for thinly or irregularly traded financial instruments and the emerging market securities. The developed technique is applied to the empirical analysis of timing and interactions between quotes and trades in the Reuters D2000-2 foreign exchange electronic brokerage system.

JEL Classification: C14, C41, D49, G15

Key Words and Phrases: high-frequency data, multivariate point process, hazard rate, competing risks, censoring, Cox proportional hazard model, foreign exchange, market microstructure, limit order book, market order, liquidity.
1 Introduction

This paper introduces an econometric framework that can be used to study short-run liquidity provision in limit order markets. This model can be applied with minor modifications to any sequence of events that occur at random, rather than predetermined, time intervals. In the context of electronic limit order market, such events can naturally be associated with changes of the information which is publicly available from the trading screens. More generally, the events can also include public “news” announcements which also can be thought of as a point process in continuous time.

The central example considered in this paper involves application of the popular statistical methodology of competing risks to the analysis of the order flow and price formation in the Reuters D2000-2 brokerage system, which has been one of the main liquidity providers in the electronic segment of the foreign exchange market. The basic assumption behind the general set-up of competing risks is that the state of a subject of study (in our case, it is primarily the publicly observable part of the electronic limit order book) can be changed by a finite number of causes (sometimes called notional risks), which can be potentially of entirely different nature and usually are assumed independent. At any moment of time the relative importance of notional risks is determined by the odds of their instantaneous realization which can be numerically characterized by the risk specific hazard functions. After a single risk is realized and the limit order book modifies, the remaining risks that competed with each other under the prior market conditions become irrelevant in the new state. Then the “internal clock” of the “race” is initialized, and a new “race” begins immediately among the participating notional risks under the new market conditions, proceeding up to the moment when the next “winner” triggers another change in the limit order book, and so forth.

From a broader perspective, the competing risks approach developed in this paper considers the limit order flow as a sequence of independent or quasi-independent realizations of a multivariate marked point process (Snyder and Miller, 1991), the number of its components being equal to the number of identified notional risks. The tradition of using univariate and bivariate marked point processes in the analysis of high-frequency financial data was started relatively recently with a series of papers by Engle and Russell (1997, 1998), Engle and Lunde (1999), Gouriéroux et al. (1999), Engle (2000), and Russell et al. (2001). In this literature the irregularly spaced arrivals of transactions or quotes are modeled by a self-exciting point process with memory, the autoregressive conditional duration (ACD) being its most popular specification. The markers are usually represented by quotes or transaction prices and volumes (when available) and modeled conditional on their arrival times.

One major impediment to broad applications of this approach to multivariate financial data has been the unresolved issue of how the markers of qualitatively distinct nature
should be treated. This problem is quite difficult since the intervals between irregularly spaced ticks (events) of one type can overlap with the intervals between irregularly spaced ticks of other types in a complicated way, and there is no natural ordering of such multiple duration intervals. For instance, the researcher extending the ACD framework to bivariate tick-by-tick data on Deutsche Mark/U.S. dollar and Japanese yen/U.S. dollar exchange rates needs to decide whether he is interested primarily in the durations between all consecutive quotes, between the quotes coming from the same market, or, probably, the duration intervals initiated by a Deutsche Mark quote which is followed by the Japanese yen quotes. A similar problem arises when the ACD model is applied to simultaneous modeling of quotes and trades that occur on the same market. If the durations between transactions and subsequent quote revisions represent the primary object of investigation (Engle and Lunde, 1999), then transaction ticks can be naturally assumed a forcing variable driving the quote revisions. Similarly, an extension of this approach to multivariate setting requires imposing a specific recursive structure describing interaction between different types of events. In this setting, modeling directly the distribution of durations is subject to *ad hoc* restrictions on the information arrival process which are usually hard to justify from the first principles.

The competing risks model developed in this paper treats events of each type symmetrically and thus avoids the need to select the “driving process.” It has at least two counterparts in the recent literature. Bisière and Kamionka (1999) apply a fully parametric competing risks model to the analysis of dynamics and sequencing of orders to trade the Alcatel shares at the Paris Bourse, gives a joint explanation of the duration between consecutive orders and their aggressiveness, and stresses the important role of information about the limit order book in the price discovery process. Unlike Bisière and Kamionka (1999), the present paper treats the hazard functions of competing risks semiparametrically, allowing more freedom in the specification of baseline hazards without increasing computational requirements.

Symmetric treatment of the components of multivariate point process makes the model of this paper similar in many respects to the multivariate version of the autoregressive conditional intensity model introduced by Russell (1999). The feature of the competing risks methodology that distinguishes it from the mainstream literature on the point processes is the inherently limited observability of the components of the underlying multivariate stochastic process due to a self-censoring mechanism directly built into the structure of the model. This leads to a natural behavioral interpretation of the hazard functions of competing risks as the instantaneous likelihoods of observing the action of trader of a particular type in the market populated by agents possessing heterogeneous information, using different models, and exercising different trading strategies. This representation appears to be especially relevant in the foreign exchange market, where the true model is
far from obvious and the information is widely dispersed across traders.

This paper begins with the review of the electronic limit order market environment and describes the Reuters D2000-2 trading system focusing specifically on the sources of public information available to the traders. Then the paper gives a stylized representation of trading history in the form of a sequence of marked random events and casts this idea into the full-fledged counting process framework. After presentation and brief discussion of the necessary statistical tools, the competing risks model is applied to the analysis of empirical data from the Reuters D2000-2 electronic limit order market.

2 The Anatomy of a Pure Limit Order Market

To provide motivation for the development of theoretical results in this paper, we introduce a stylized automated limit order book. While technical details of the workings of limit order markets might be extremely important in specific applications, our presentation will be deliberately simplified and focus attention on the salient features of this market structure. The presentation in this section has been developed bearing in mind a specific application to a real limit order market, the Reuters D2000-2, which is organized (essentially) as a pure electronic limit order book.

2.1 Introduction to the Reuters D2000-2 electronic dealing system

The Reuters D2000-2 dealing system, which is the source of our data, is one of the two largest providers of electronic brokerage services on the spot foreign exchange market in major currencies. It operates as an anonymous matching service for trading in major currencies.1 The D2000-2 host queues buy and sell orders on precise criteria of price and then time priority. The system incorporates a pre-screened pricing facility to ensure that two anonymous counterparties in a potential trade have mutual credit limits before a match can happen. Once a match between the two orders occurs, the parties are notified on the resulting transaction and provided with the information needed to settle the deal.2

Reuters clients for D2000-2 are the dealers and traders in foreign exchange trading rooms of major financial institutions around the globe. The network of active counterparties of D2000-2 was as large as 1,124 sites globally as of the middle of 1997.

The D2000-2 trading screen has the following sections (Figure 1):

---

1 Clients must subscribe to the voice-based dealing system of D2000-1 to access D2000-2. Users can move from one service to the other on the same keystation. The D2000-2 service has two components of its own, D2000-2 Spot and D2000-2 Forwards, the latter designed to facilitate foreign exchange forward trading.

2 Only the counterparty initiating the deal is charged for the transaction. The transaction fee as small as $25 is paid by the aggressor to Reuters Transaction Services Ltd.
• The multiple currency display allows to trade up to six currency pairs simultaneously and highlight the active pair.

• The market quote area shows the anonymous best bid and offer prices $P_{bid}$ and $P_{ask}$ for limit orders entered into the system.

• The market depth indicator shows the accumulated quantities $(Q_{bid}, Q_{ask})$ available at the market bid and offer.

• The best quote area shows the best bid and offer prices available to this trader based on mutual credit.

• The best depth indicates the quantities available at the best pre-screened prices.

• The trader quote and trader depth section indicates the best prices and quantities of active limit orders entered onto D2000-2 at this keystation.

• The last price indicator shows the last transaction price ($P^*$) and the direction of trade (Sell/Buy) that occurred on D2000-2 for each currency pair.

• The trader mailbox section contains immediate confirmation messages of the trader’s quoting activity and the deal tickets showing all details of her trades.

Additionally, the trading screen can be customized to highlight various aspects of the subscriber’s activity in single or multiple currencies.

2.2 Description of the Limit Order Market Structure

In the development below, we call by limit order an offer to sell or a bid to buy a specified number of units of financial asset at a specified price, which is called the limit order price. We call by market order an order to sell (or to buy) a specified number of units of financial asset at the best market price available. It follows from these definitions that market orders always execute completely, while limit orders may face only partial immediate execution, or may not execute at all, in which case the unexecuted portion of the limit order is put into one of the two queues (for sell and for buy limit orders) and remains there until explicitly cancelled by the trader. The market structure which consists of the two queues for buy and sell limit orders, along with the specified trading protocol describing the priority of limit order execution is called the limit order book. It should be clear from these definitions that the queues of sell and buy limit orders can be interpreted as approximate representations of excess supply and demand curves for the traded financial asset (Figure 2). The priority of limit orders in the book depends on the details and the trading protocol of the particular system. All orders are submitted
Figure 1: The customized display of the Reuters D2000-2 automated brokerage terminal. Part of the display in the center of the screen contains electronic communication messages for direct bilateral trades transmitted via D2000-1 system (which is not covered in this paper). The upper part of the display provides information about the state of the Reuters D2000-2 electronic limit order book. The dealer can choose up to five exchange rates or select just one pair. The exchange rates are displayed in two alternative formats: in the upper left side of the screen the best market quotes and quantities available at these quotes are displayed; to the upper right, the dealer can see the best bid and ask quotes and quantities available to him. In the upper right corner of the screen, the dealer observes the direction and price of the last trade through the Reuters D2000-2 automated brokerage. The image of the Reuter D2000-2 electronic trading screen used in this insert has been taken from the Reuters information webpage at http://about.reuters.com/transactions/d22s.htm which provides detailed information on the Reuters D2000 trading system.
anonymously as the identity of a foreign exchange trader is considered a strictly confidential information and never disclosed.

Only a handful of real financial markets are organized as pure electronic limit order books. Of course, in practice virtually any real limit order market represents a hybrid system, which makes it hard to analyze using the simple model presented below. For example, traders submitting market sell and buy orders in Reuters D2000-2 must provide not only the quantity but also the price (which does not have to be but usually is the best market price available on the buy or sell side of the market). After the market order is submitted it is matched only with the limit orders submitted at the prices equal to the arriving market order price. The unexecuted portion of the market order is cancelled automatically. Unlike the market orders, aggressive limit orders submitted at the prices that are different from the current best price can obtain price improvement, as the outstanding limit orders in the book submitted at the better prices receive priority in execution. The traders demanding early execution must be willing to submit a market order which implies paying the transaction cost equal to the difference between best bid and ask prices (the bid-ask spread), or submit a limit order which gives them a chance of price improvement at the expense of some uncertainty of execution, especially in the event of a sudden cancellation of outstanding limit orders at the better prices. The seller who is willing to receive a better execution price may submit an ask limit order with a relatively high ask price and be prepared to wait. Such trader should deal with the uncertainty of execution in the event of the market price sliding down in the opposite direction, and with the risk of being “picked off” in the event of a sudden increase in the market price beyond the level considered reasonable by this limit order trader, unless he promptly cancels his order.

The dynamics of a limit order market are illustrated on the graphs below. Figure 3 represents the situation when a limit order trader submits a bid to buy one million dollars at the price 1.7510 Deutsche Marks per dollar, which exceeds the previously available best market price by 0.01 Pfennigs but still 0.02 Pfennigs short of the price which (almost) guarantees immediate execution.

Alternatively, the trader can submit a less aggressive limit buy order at the current best market price (Figure 4). In such event the arriving limit order receives a lower priority in comparison to the limit order that has been previously submitted to the limit order book at the same price. The traders who are closely watching the monitors of their trading screens should notice increase of the quantity available at the best market bid price. Whether and how this public information arrival would affect the behavior of other traders remains one of the central issues in the empirical and theoretical market microstructure literature.

The trader can avoid revealing any information to the public about his intentions
Figure 2: This diagram shows the state of the Reuters D2000-2 electronic limit order book at a particular moment in time. The erzatz supply and demand curves in the market for US dollars are represented by limit sell and buy orders waiting their execution. Two limit orders, for one and two million dollars, are available at the best market sell price of DEM 1.7512 per dollar. Additionally, there is one limit order for three millions at the ask price of DEM 1.7513, one limit order for one million at the ask price of DEM 1.7514, and two limit orders for one million each at the ask price of DEM 1.7515 per US dollar. On the bid side, there is one limit order to purchase two millions at the best market buy price of DEM 1.7509 per dollar, which is followed (in the order of priority) by three limit orders for two million, one million, and two million dollars at the bid price of DEM 1.7507, a limit order for two million at the bid price of DEM 1.7506, a limit order for one million at the bid of DEM 1.7505, and another large limit order at the same price (the size of this buy order is unclear from the graph). Note that traders observe only the best market buy and sell prices DEM 1.7509 and DEM 1.7512, along with the quantities $2 mln. and $3 mln., respectively, on their trading screens.
Figure 3: New limit order to purchase one million dollars arrives at the bid price of DEM 1.7510 per dollar, leading to change of the information on the trading screens.

Figure 4: New limit order to purchase one million dollars arrives at the current best bid price of DEM 1.7509 per dollar, leading to change of the information about the quantity available on the buy side.
Figure 5: New limit order to purchase one million dollars is submitted at DEM 1.7508 per dollar, just below the current best bid market price. No information about the intention to trade is revealed to the public.

and willingness to trade if he submits a subsidiary limit order, which can be any limit buy order at the price just below the current best market price (Figure 5) or two ticks (minimal increments) below the current bid market price (Figure 6). In both cases the best market prices and the quantities available to sellers and buyers at these prices remain unchanged. Therefore, the information on the trading screen is not affected by the arrivals of subsidiary limit orders.

The last three diagrams provide illustrations of the effects of a subsidiary limit order cancellation (Figure 7), a market order-initiated transaction followed by immediate cancellation of the portion of order which cannot be matched at the best sell market price (Figure 8), and a similar transaction initiated by an aggressive limit order buyer (Figure 9). Note the difference between the effect of market versus aggressive limit order arrival on the market liquidity. While the market order trader only consumes the liquidity available at the best market sell price of DEM 1.7512 per dollar, the limit order buyer also provides liquidity at this price, once it becomes clear that the quantity available for sale at this price is insufficient to satisfy his demand for liquidity. Since the information content and the role played by aggressive limit orders in liquidity provision can be different from the role played by market orders, it is important to differentiate between these two types of events as they might be used by traders possessing distinct information or having different beliefs and risk attitudes.
Figure 6: The limit order to purchase one million dollars is submitted at DEM 1.7507 per dollar, which is two ticks below the current best bid market price of DEM 1.7509 per dollar. The arriving order receives lower priority of execution relative to the three outstanding bid limit orders submitted earlier at the same price.

3 An Econometric Model of the Limit Order Book

3.1 Events, epochs, sub-epochs, and the internal clock

Suppose we observe $N$ epochs $(T_{n-1}; T_n), n = 1, 2, ..., N$, which are represented by the time intervals between consecutive occurrences of publicly observable events. Such events are assumed to have one of the $S$ types $r = 1, 2, ..., S$, and can be interpreted as observations generated by the underlying economic process. In the context of the pure limit order market, any such event is associated with a separate epoch. At the beginning of each epoch the internal clock is re-started and run until the first event associated with arrival of publicly available information which is represented by vector $Z_n$ of discrete covariates.

The covariates can represent the information available on the electronic trading screens, public news announcements, etc. By design, vector $Z_n$ is predetermined by the previous history and by the initial conditions and thus remains constant throughout the epoch.

In addition to publicly observable events, there are events of types $r = S + 1, ..., R$ that cannot be observed by the general public. More precisely, occurrences of unobservable events of types $r = S + 1, ..., R$ remain private information of agents initiating such events.

Any epoch $n$ contains a non-negative number of such unobservable events, which naturally identify the sub-epochs within a given time interval $(T_{n-1}; T_n]$. The internal clock of the
Figure 7: This diagram illustrates the effect of cancellation of a limit order to purchase two million dollars at DEM 1.7506 per dollar, which was three ticks below the current best bid market price of DEM 1.7509 per dollar. The information on the trading screens is unaffected by this event, which occurs deep on the bid side of the limit order book.
Figure 8: The market order to purchase five million dollars is submitted at the best sell market price of DEM 1.7512 per dollar. Since the quantity available at this price is only three million dollars, part of liquidity demand created by the market order trader is not satisfied. The unmatched portion of the market order is cancelled immediately, while the best sell price goes up one tick to DEM 1.7513 per dollar.
Figure 9: The limit order to buy five million dollars is submitted at the best sell market price of DEM 1.7512 per dollar. The situation is analogous to the market order submission, except that the unmatched portion of the arriving aggressive limit order remains in the limit order book, leading to the shift of the best ask price by three ticks to DEM 1.7512 per dollar.
The system is not re-started at the beginning of sub-epoch within a given epoch. The covariate vector \( Z_n \) representing the publicly available information remains unchanged until the occurrence of next observable event at the random time \( T_n \).

The general structure of an epoch is shown on Figure 10. To fix the notation, we will assume throughout this section that the sample contains the total number \( N^* \) of observable and unobservable events. We also assume \( R \ll N \), which means that the number of event types is much smaller than the number of interarrival epochs in the sample. The terms “event”, “failure”, and “risk” will be used interchangeably. The term “previous history” will refer to the records of durations and events prior to the beginning of a given epoch.

### 3.2 A counting process representation of event history

Let the random process \( N_{nr}(t) \) count the events of type \( r \) that occur between zero and \( t \) seconds after the beginning of the \( n \)th epoch \( (T_{n-1}; T_n] \). Thus \( N_{nr}(. ) \) is a non-decreasing univariate counting process taking values \( \{0, 1\} \) for the observable events of types \( r = 1, ..., S \), and taking any integer values for unobservable events of types \( r = S + 1, ..., R \). Consider the random process

\[
N_r(t) = N_{1r}(t) + N_{2r}(t) + \ldots + N_{Nr}(t)
\]

counting the total number of events of type \( r \) in the duration interval \( (0; t] \) since the beginning of all epochs \( n = 1, ..., N \) in the sample. Assume that all processes \( N_{nr}(t) \) are well-defined with respect to a right-continuous, increasing, and complete filtration \( \mathcal{F}_n(t)_{t=0}^\tau \), where \( \tau \) is a fixed positive number or infinity, which means that they obey the usual technical conditions imposed in the statistical literature on counting processes (see Andersen et al., 1993, p.60).

The dynamic properties of the analyzed collection of counting processes

\[
\mathbf{N} = (N_{nr}, \ r = 1, ..., R; \ n = 1, ..., N)
\]

are characterized by their compensators

\[
\Lambda(\theta) = (\Lambda_{nr}(\theta), \ r = 1, ..., R; \ n = 1, ..., N).
\]

The compensators are absolutely continuous and given by

\[
\Lambda_{nr}(t; \theta) = \int_0^t \lambda_{nr}(u; \theta) du, \quad r = 1, ..., R, \ n = 1, ..., N,
\]

where \( \theta \) is the parameter of interest. The individual intensity processes \( \lambda_{nr}(\cdot; \theta) \) have a multiplicative structure

\[
\lambda_{nr}(t; \theta) = Y_{nr}(t)h_{nr}(t|Z_n; \theta), \quad r = 1, ..., R, \ n = 1, ..., N.
\]
The indicators $Y_{nr}(t)$ of being at risk of type $r$ at time $t$ of epoch $n$ are predictable (but not necessarily publicly observable) binary processes which are independent of $\theta$. Moreover, we assume for convenience that for any type of notional risk $r$ in epoch $n$ the corresponding hazard rate $h_{nr}()$ satisfies the Cox’s proportional hazard model

$$h_{nr}(t|Z_n; \theta) = h_{0r}(t; \gamma) \exp(Z_n'\beta_r), \quad r = 1, \ldots, R, \quad n = 1, \ldots, N,$$

where $Z_n = (Z_{n1}, \ldots, Z_{np})'$ is the vector of covariates observed at the beginning of the $n$th epoch. The parameter $\theta$ has a semiparametric structure,

$$\theta = (\beta_1', \ldots, \beta_R; \gamma'),$$

where each $\beta_r, \quad r = 1, \ldots, R$, has dimension $p$, whereas $\gamma$ describing the shape of baseline hazard $h_{0r}(; \gamma)$ can be, in general, infinite dimensional. The only assumption about the baseline hazards $h_{0r}(; \gamma)$ that we need at this point is that $h_{0r}(; \gamma)$ are non-negative with

$$\int_0^t h_{0r}(u)du < \infty, \quad r = 1, \ldots, R,$$

for all $t \in [0; \tau]$.

Components of the joint process $N_n(t) = (N_{nr}(t))_{r=1}^R$ are assumed to be independent with respect to the filtration $(\mathcal{F}_n(t))_{t=0}^\tau$. This means that conditional on $(N_n(u))_{u=0}^t$ and the covariates $Z_n$, we postulate the existence of the intensity

$${\lambda}_n(t; \theta) = \lim_{\delta \to 0} \frac{1}{\delta} \Pr\{N_n(t+\delta) - N_n(t) > 0|Z_n, (Y_{nr}(u))_{r=1}^R, 0 \leq u \leq t\}$$

of the joint process and the relationship

$${\lambda}_n(t; \theta) = \sum_{r=1}^R \lambda_{nr}(t; \theta)$$

between the intensities of the joint process and its components, which holds for all $n = 1, \ldots, N$ and $0 \leq t < \tau$.

To formulate the asymptotic results for the estimates of the hazard functions of individual events, we have to make some technical assumptions about the covariate process $(Z_n)_{n=1}^N$. In particular, assume that it is stationary, ergodic, and supported by a compact set in $\mathbb{R}^p$. Denote by $\Phi_r(z, t)$ the conditional distribution functions of the covariate vector $Z_n$, provided that $t$ seconds after the beginning of a randomly selected epoch $(T_{n-1}; T_n]$ the system is still at risk of type $r$, that is,

$$\Phi_r(z, t) = \Pr\{Z_n \leq z|Y_{nr}(t-1) = 1\}.$$
If there exists an absolutely continuous non-increasing function

\[ S_r(t) = \Pr\{Y_{nr}(t^-) = 1\}, \]

which can be interpreted as the survival probability by time \( t \) while being under the risk of type \( r \) at time \( t \), then the distribution function \( \Phi_r(z) \) of the covariate \( Z_n \), conditional on the information that the system is under the risk of type \( r \), can be defined as follows

\[ \Phi_r(z) = \Pr\{Z_n \leq z | Y_{nr}(t^-) = 1 \text{ for some } t \in [0; \tau)\} \]

\[ = - \int_0^\tau \Phi_r(z, t) dS_r(t). \]

Finally, consider

\[ \mathcal{F}_n(t) = \sigma\{(N_n(u))_{u=0}^{t}, (Y_n(u))_{u=0}^{t}, Z_n\} \]

and define the filtration

\[ \mathcal{F}(t) = \bigvee_{n=1}^N \mathcal{F}_n(t). \]

With the above definitions, \( \lambda_n(t) = (\lambda_{nr}(t))_{r=1}^{R} \) is a predictable process having independent components with respect to \( \mathcal{F}_n(t) \), and hence with respect to \( \mathcal{F}(t) \), while the processes

\[ M_n(t) = (M_{nr}(t))_{r=1}^{R} = (N_{nr}(t) - \Lambda_{nr}(t))_{r=1}^{R} \]

with compensators

\[ \Lambda_{nr}(t) = \int_0^t \lambda_{nr}(u) du \]

are square-integrable local martingales with respect to \( \mathcal{F}(t) \), which also have independent components on the interval of durations \([0; \tau]\). More precisely,

\[ \Lambda_n(t) = (\Lambda_{nr}(t))_{r=1}^{R} \]

is the compensator of vector process

\[ N_n(t) = (N_{nr}(t))_{r=1}^{R} \]

with respect to the filtration \( \mathcal{F}_n(t) \lor \mathcal{H}_r(t) \), where \( \mathcal{H}_r(t) \) is the \( \sigma \)-field generated by the entire future of

\[ Y_n(t) = (Y_{nr}(t))_{r=1}^{R} \]

within the given epoch \( n \).

Before describing the methods which can be use for estimation of this rather abstract general model of event history, it is worthwhile to recast the key concepts in the context of the two simple examples.
3.3 Example 1: Two competing risks

Suppose we observe a sample \((Z_n, T_n, \delta_n)_{n=1}^N\), where \(Z_n\) is a vector of covariates, \(\delta_n = 1\{T_{Sell,n} < T_{Buy,n}\}\) is the indicator of whether the \(n\)th epoch ends with a seller-initiated (vs. buyer-initiated) event, and \(T_n = \min(T_{Buy,n}, T_{Sell,n})\). In the prototypical example of two competing risks without covariates, the only observables are the outcomes \((T_n, \delta_n)\) of the two latent durations \(T_{Buy,n}\) and \(T_{Sell,n}\) (Figure 11).

3.4 Example 2: Multiple risks in a dynamic limit order environment

The main application of the competing risks theory developed in this paper concerns with the analysis of the trading history in an electronic limit order book, which can be completely characterized by the sample \((Z_n, Y_n(t), T_n, r_n)_{n=1}^N\), where \(Z_n\) is a set of covariates that includes all publicly observable characteristics of the limit order book at the beginning of epoch \(n\) (such as the information on a trading screen at the outset of epoch \(n\)) as well as a publicly available recent history of the book. As usual in a competing risks environment,

\[
T_n = \min_{r=1,...,R} T_{nr}
\]
Consider interval $n = 2$

Type of risk, i.e., whether next event is S or B

- $r = 2$ (B, buyer-initiated) Process $N_{2n}(t_n)$
- $r = 1$ (S, seller-initiated) Process $N_{1n}(t_n)$

$0 \leq t_n \leq \min\{t_{1n}, t_{2n}\}$ (time interval since last observed event)

Figure 11: Stylized representation of the competing risks within epoch $n = 2$. This epoch was initiated by an event of type S and ends with an event of type B. The event of type S that could have occurred $t_{1n}$ seconds after the beginning of epoch is censored by the early arrival of the competing risk B at time $t_{2n} < t_{1n}$.
Add an unobservable type of event, \( r = 3 \)

**Figure 12:** Stylized representation of epoch \( n = 2 \) after the third (unobservable) type of event has been added.

is defined as the duration of epoch \( n \),

\[
    r_n = \arg\min_{r=1, \ldots, R} T_{nr}
\]

is the type of event causing the end of epoch \( n \), and \( Y_n(t) = (Y_{nr}(t))_{r=1}^R \) is the vector of at-risk indicator processes, which are usually functions of covariates \( Z_n \) and sometimes also depend on the state of the limit order book, which may be hidden and unavailable from the trading screens. The simplified illustration of three alternative types of risk, of which only two are assumed to be publicly observable, is shown on Figure 12.

We conclude this section with three alternative examples of at-risk indicators \( Y_{nr}(t) \).

The first example is characterized by \( Y_{nr}(0+) = 1 \), which corresponds to the typical situation where the market conditions at the beginning of epoch \( n \) do not prevent the occurrence of event of type \( r \) at any moment within the given epoch.

The second example corresponds to the case \( Y_{nr}(t) \equiv 0 \) for all \( t \in [0; \tau) \). For instance, if \( r \) is the risk that the next event will be submission of a limit order within the bid-ask spread, and we know that the bid-ask spread currently takes the minimum value of one tick, then risk \( r \) can be effectively eliminated from the set of the risks competing in the current epoch \( n \) since it can never occur.

\[\text{For the time being we assume that the tied durations are not allowed, so the type } r \text{ of event is determined uniquely.}\]
The third example involves the risk $r$ that the next event will be cancellation of a subsidiary limit order one tick above the best ask price, whereas no limit orders at this price level are currently available. Therefore, risk $r$ is not among those competing for time priority in the beginning of the current epoch. However, if prior to any observable event that would end the current epoch there will be submission of a subsidiary sell limit order one tick above the ask price, risk $r$ can re-emerge and therefore should be included the set of competitors. This happens, for example, when the at-risk indicator $Y_{nr}(t)$, which equals zero at the beginning of the epoch, switches its value before the epoch ends. Of course, any switch in the indicator would mean the beginning of a new sub-epoch, but not necessarily the new epoch. (See the definition of sub-epoch in the beginning of this section.)

4 Semiparametric Estimation of the Competing Risk Model

This section presents the likelihood-based approach to the semiparametric estimation of hazard rates for competing risks. The methods frequently used for estimating hazard rates are either fully parametric or nonparametric. While the theory behind parametric methods is rather straightforward, the estimation results can be biased if the parametric model is misspecified. On the other hand, purely nonparametric techniques often suffer from the “curse of dimensionality” leading to unacceptably wide confidence bands. Therefore one can expect to hit the middle ground with the semiparametric approach which takes advantage of both parametric and nonparametric techniques.

In this section we will focus on a semiparametric likelihood-based estimation procedure which is robust and computationally efficient. Robustness can be a serious concern in applications since economic and financial theories usually leave at least part of the model unspecified. Moreover, frequently there is a logical gap between the specification of highly stylized theoretical models of market microstructure delivering crisp predictions under the purist assumptions and the loose data-driven empirical models used to test implications of the theory. Taking into consideration multiple sources of noise and errors in the empirical data, successful implementation of empirical methods often requires large data sets that might involve tens and even hundreds of thousand observations. Therefore, the issue of computational efficiency should also be taken into consideration, since the methods that can be very successful in applications to the samples of moderate size often become excessively cumbersome and impractical when applied to large databases, which is often the case with high-frequency financial econometrics.

The following review will begin with the description of the profile likelihood method for hazard rate estimation. We illustrate this method using the Cox proportional hazard
specification which leads to an extremely simple form of the partial likelihood function and has a certain robustness property. Moreover, the likelihood function can be efficiently maximized using the standard statistical software. The partial likelihood estimation naturally leads to the classical Nelson–Aalen estimator for the cumulative hazard functions.

Next we review the idea of smoothing the increments of this estimator which was originally applied by Ramlau-Hansen (1983) to estimate the baseline hazard functions. After a brief and heuristic discussion of the deficiencies of the proposed “internal” estimator we present an alternative “external” estimator, which has the same first-order asymptotics as the Ramlau-Hansen estimator but is expected to perform better in the presence of heavily censored observations, which is a typical situation in the models with multiple competing risks. Finally we provide an extension of the locally constant “external” estimator to the locally linear estimator. Preliminary results of this section motivate the analysis of asymptotic properties of the proposed estimators.

4.1 The Nelson–Aalen estimator for the cumulative hazards and the incidence rates

Assume that the observed dataset available on the triplets of censored data \((T_n, r_n, Z_n)\), \(n = 1, ..., N\), consist of the times and types of observed events

\[
T_n = \min_{r=1,...,R} T_{nr} \quad \text{and} \quad r_n = \arg\min_{r=1,...,R} T_{nr},
\]

where the stopping times \(T_{nr}\) of counting processes \(N_{nr}\) are defined by

\[
t_{nr} = \inf\{t \in [0; \tau) : \Delta N_{nr}(t) > 0\}, \quad r = 1, ..., R; \quad n = 1, ..., N.
\]

The partial likelihood function of the data is proportional to the product integral

\[
\prod_{n=1}^{N} \prod_{t \in [0; \tau]} \left\{ \prod_{r=1}^{R} (dH_{0r}(t) \exp(Z_n^{0} \beta_r))^\Delta N_{nr}(t) \left( 1 - \sum_{r=1}^{R} dH_{0r}(t) S_r^{(0)}(\beta_r, t) \right)^{1-\Delta N_{nr}(t)} \right\}
\]

defined on the complete dataset \((N_{nr}, r = 1, ..., R; \ n = 1, ..., N)\), which are only partially observable. In the previous formula we use the notation

\[
S_r^{(0)}(\beta_r, t) = \sum_{n=1}^{N} Y_{nr}(t) \exp(Z_n^{0} \beta_r), \quad r = 1, ..., R,
\]

for the cumulative risk index of type \(r\) at time \(t-\), and

\[
H_{0r}(t) = \int_{0}^{t} h_{0r}(u) du, \quad r = 1, ..., R,
\]

for the cumulative baseline hazard functions of type \(r\).
After rewriting the partial likelihood in the form
\[
\prod_{t \in [0; \tau]} \left\{ \prod_{n=1}^{N} \prod_{r=1}^{R} (dH_{0r}(t) \exp(\mathbf{Z}_n' \beta_r))^{\Delta N_{nr}(t)} \right\} \exp \left[ - \sum_{r=1}^{R} dH_{0r}(t) S_r^{(0)}(\beta_r, t) \right]
\]
and maximizing it with respect to \( \Delta H_{0r}(t) \) for a fixed value of \( \beta = (\beta_1', ..., \beta_R')' \), we obtain
\[
\Delta \tilde{H}_r(t, \beta_r) = \frac{\Delta N_{nr}(t)}{S_r^{(0)}(\beta_r, t)}
\]
and, therefore, for a fixed value of \( \beta \), \( H_{0r}(t) \) is estimated by the Nelson–Aalen estimator
\[
\tilde{H}_{0r}(t, \beta_r) = \int_0^t \frac{J_r(u) dN_{nr}(u)}{S_r^{(0)}(\beta_r, u)}
\]
where \( J_r(u) \equiv 1(Y_r(u) > 0), Y_r = Y_{1r} + \ldots + Y_{Nr} \), and \( N_r = N_{1r} + \ldots + N_{Nr} \).

Inserting the Nelson–Aalen estimator into the partial likelihood expression yields the profile likelihood which only depends on \( \beta \):
\[
L_{\text{part}}(\beta) = L_{\text{Cox}}(\beta) \prod_{t \in [0; \tau]} \prod_{r=1}^{R} \Delta N_{nr}(t)^{\Delta N_{nr}(t)} (1 - \Delta N_{nr}(t))^{1 - \Delta N_{nr}(t)}
\]
\[
= L_{\text{Cox}}(\beta) \prod_{t \in [0; \tau]} \left\{ \prod_{r=1}^{R} \Delta N_{nr}(t)^{\Delta N_{nr}(t)} \right\} \exp(-N_r(\tau)),
\]
where \( N_r = N_{1r} + \ldots + N_{Nr} \), and
\[
L_{\text{Cox}}(\beta) = \prod_{t \in [0; \tau]} \prod_{n=1}^{N} \prod_{r=1}^{R} \left( \frac{\exp(\mathbf{Z}_n' \beta_r)}{S_r^{(0)}(\beta_r, t)} \right)^{\Delta N_{nr}(t)}
\]
is the Cox partial likelihood. The value of \( \beta \) maximizing the log Cox partial likelihood
\[
l_{\text{Cox}}(\beta) = \log L_{\text{Cox}}(\beta)
\]
\[
= \sum_{r=1}^{R} \left[ \sum_{n=1}^{N} \int_0^\tau \mathbf{Z}_n' \beta_r dN_{nr}(t) - \int_0^\tau \log S_r^{(0)}(\beta_r, t) dN_{nr}(t) \right]
\]
will be denoted \( \hat{\beta} \). Then the cumulative baseline hazard function \( H_{0r}(t), r = 1, ..., R \), is estimated by \( \tilde{H}_{0r}(s, \hat{\beta}_r) \), which is the Breslow estimator of the baseline hazard of competing risk \( r \).

Finally, the conditional probability that event of type \( r \) will be next to occur less than \( t \leq \tau \) seconds after the previous event is
\[
P_r(0, t; \mathbf{Z}_0) = \int_0^t P_r(0, u; \mathbf{Z}_0) \exp(\mathbf{Z}_0' \beta_r) dH_{0r}(u), \quad r = 1, ..., R,
\]
and can be estimated by

\[ \hat{P}_r(t; \mathbf{z}_0) = \int_0^t \tilde{P}_0(0, w; \mathbf{z}_0) \exp(\mathbf{z}_0' \tilde{\mathbf{\beta}}_r) d\tilde{H}_0r(u, \tilde{\mathbf{\beta}}_r), \]

where \( P_0(0, t; \mathbf{z}_0) \) and \( \tilde{P}_0(0, t; \mathbf{z}_0) \) are respectively the conditional probability of survival at time \( t \) since last event,

\[ P_0(0, t; \mathbf{z}_0) = \prod_{u \in [0; t)} \left( 1 - \sum_{r=1}^R \exp(\mathbf{z}_0' \mathbf{\beta}_r) dH_0r(u) \right) \]

and its estimator

\[ \hat{P}_0(0, t; \mathbf{z}_0) = \exp \left[ - \sum_{r=1}^R \exp(\mathbf{z}_0' \hat{\mathbf{\beta}}_r) \hat{H}_0r(t) \right], \]

and its estimator

\[ \hat{P}_0(0, t; \mathbf{z}_0) = \exp \left[ - \sum_{r=1}^R \exp(\mathbf{z}_0' \hat{\mathbf{\beta}}_r) \hat{H}_0r(t, \hat{\mathbf{\beta}}_r) \right]. \]

4.2 Invariance of the maximum partial likelihood estimator

The partial likelihood function \( L_{\text{Cox}}(\mathbf{\beta}) \) has a remarkable robustness property which makes the maximum likelihood estimator \( \hat{\mathbf{\beta}} \) an attractive choice in the following dynamic context. Specifically, it is invariant with respect to monotonic deformations of the time scale within the duration period \([0; \tau]\). This invariance property holds by virtue of the fact that the numerical value of the log Cox partial likelihood function

\[ \ell_{\text{Cox}}(\mathbf{\beta}) = \log L_{\text{Cox}}(\mathbf{\beta}) = \sum_{r=1}^R \left[ \sum_{n=1}^N \int_0^\tau Z_n' \mathbf{\beta}_r dN_{nr}(t) - \int_0^\tau \log S_r^{(0)}(\mathbf{\beta}_r, t) dN_r(t) \right] \]

is unaffected after the parameter \( t \in [0; \tau] \) is transformed monotonically into \( \tilde{t} = g(t) \in [0; g(\tau)] \). Indeed, the ordering of occurrences of the failure events and their censorings is unaffected by the deterministic time scale deformation \( g(\cdot) \), which means invariance of the second integral in the above formula. Invariance of the first integral with respect to the time deformation is also obvious.

4.3 Alternative kernel estimators of baseline hazard rates

Historically the first kernel estimator for \( h_{0r}(t) \) was proposed in Ramlau-Hansen (1983). It is defined as

\[ \tilde{h}_{0r}(t|\mathcal{K}_r, b_r) = \frac{1}{b_r} \int_0^\tau \mathcal{K}_r \left( \frac{t - u}{b_r} \right) d\tilde{H}_0r(u, \tilde{\mathbf{\beta}}_r), \]
where \( \hat{H}_0r(u, \beta_r) \) is the Nelson–Aalen martingale estimator of the cumulative hazard function at point \( u \), with the kernel function \( K_r \) that satisfies the usual conditions \( K \). It is easy to see that that Ramlau-Hansen estimator is obtained by application of the kernel smoother to the increments of the Nelson–Aalen estimator. The “internal” property of this estimator comes from the fact that since its properties depend only on the realized durations of risk \( r \), and are independent of the realized durations of the competing risks \( r' \neq r \).

The heuristic idea behind the second method of baseline hazard estimation is that the value of \( h_0r(\cdot) \) at the point \( t \) is likely to affect the process \( N_{nr}(\cdot) \) only in a neighborhood of \( t \). Thus in order to estimate \( h_0r(t) \) it may be sufficient to consider a portion of the likelihood function which emphasizes the behavior of the process around \( t \). This is accomplished by computing the kernel-weighted partial log-likelihood

\[
(\ell_{\text{part},N}(\hat{\beta}) * K_b)(t) = \sum_{r=1}^{R} \int_{-\infty}^{\infty} \left[ \int_{0}^{\tau} \sum_{n=1}^{N} \Delta N_{nr}(u)(\log(h_0r(u)) + Z_u^r \beta_r) \right. \\
- \int_{0}^{\tau} S^{(0)}(\hat{\beta}_r, u)dH_0r(u) \frac{K_{r,b}(t-u)}{t} \\
= \sum_{r=1}^{R} (\ell_{\text{part},r,N}(\hat{\beta}_r) * K_{r,b_r})(t)
\]

instead of the ordinary likelihood, where the kernel function \( K_{r,b_r} \) is defined as \( K_{r,b_r}(t) = \frac{1}{b_r} K_r(t/b_r) \), and \( K_r(\cdot) \) satisfies the following standard condition for kernel functions.

**Condition K.** The kernel functions \( K_r, r = 1, ..., R, \) satisfy

\[
\int_{-1}^{1} K_r(t)dt = 1, \quad \int_{-1}^{1} tK_r(t)dt = 0, \quad \int_{-1}^{1} t^2K_r(t)dt = \kappa_{2r} > 0.
\]

Let \( \hat{h}_0r(t) \) maximize the kernel-weighted partial log-likelihood

\[
(\ell_{\text{part},r,N}(\hat{\beta}_r) * K_{r,b_r})(t) \equiv \int_{-1}^{1} \left[ \int_{0}^{\tau} \sum_{n=1}^{N} \Delta N_{nr}(u)(\log(h_0r(u)) + Z_u^r \beta_r) \right. \\
- \int_{0}^{\tau} S^{(0)}(\hat{\beta}_r, u)dH_0r(u) \frac{K_{r,b_r}(t-u)}{t} \\
\]
\begin{align*}
\int_0^\tau \int_0^1 \sum_{n=1}^N \Delta N_{nr}(u) (\log(h_{0r}(u)) K_{r,b_r}(t-u) du \\
+ \int_0^\tau \sum_{n=1}^N Z'_{n}\hat{\beta}_r \int_{-1}^1 \Delta N_{nr}(u) K_{r,b_r}(t-u) du \\
- \int_0^\tau \int_{-1}^1 S_r^{(0)}(\hat{\beta}_r, u) h_{0r}(u) K_{r,b_r}(t-u) du
\end{align*}

\begin{align*}
\int_0^\tau \int_0^1 \sum_{n=1}^N \Delta N_{nr}(u) (\log(h_{0r}(u)) K_{r,b_r}(t-u) du \\
+ \int_0^\tau \sum_{n=1}^N Z'_{n}\hat{\beta}_r Y_{nr}(u) K_{r,b_r}(t-u) du \\
- \int_0^\tau \int_{-1}^1 S_r^{(0)}(\hat{\beta}_r, u) h_{0r}(u) K_{r,b_r}(t-u) du
\end{align*}

where $K_b \equiv (K_{1,b_1}, K_{2,b_2}, ..., K_{R,b_R})'$ is an $R$-vector of non-negative smooth symmetric kernels $K_{r,b_r}$ with support $[-1; 1]$ and vector of bandwidth parameters $b = (b_1, b_2, ..., b_R)'$. Note that the kernel-weighted partial log-likelihood can still be re-written so that the Cox partial log-likelihood is factored out yielding the so-called “external” estimator

\[
\tilde{h}_{0r}(t | K_{r,b_r}) = \frac{\int_0^\tau K_r \left( \frac{t-u}{b_r} \right) J_r(u) dN_r(u)}{\int_0^\tau K_r \left( \frac{t-u}{b_r} \right) S_r^{(0)}(\hat{\beta}_r, u) du}
\]

This estimator was proposed by Hjort (1992) and studied extensively by Jones et al. (1998) and Nielsen and Linton (1995), among others. The key feature of estimator $\tilde{h}_{0r}$ is its external property. Indeed it is easy to see that $\tilde{h}_{0r}$ incorporates the information about the realized durations of the notional risk $r$ under consideration both directly and indirectly (via the risk index $S_r^{(0)}(\hat{\beta}_r, u)$) by eliciting the information about censoring of risk $r$ by its competitors $r' \neq r$.

Both Ramlau-Hansen and Hjort estimators are consistent and asymptotically normal under the mild regularity conditions (Andersen et al. (1993), Jones et al. (1998)). Moreover, the asymptotic distributions of both proposed estimators are identical to each other (Hjort, 1992).\footnote{This problem was discussed by Nielsen et al. (1998). They considered a closely related problem of estimating a semiparametric hazard rate with a parametric duration part and nonparametric dependence of hazard rates on the marker (covariate) process.} However, despite its simplicity, the Ramlau-Hansen estimator uses the information inefficiently, especially when the realized durations of risk $r$ are subject to heavy censoring by a large number of “competitors”, which is true in the main example.
of this paper. Therefore, the external estimator will be our primary choice throughout the next section.

4.4 Bias reduction by local linear smoothing

In local linear kernel smoothing, the baseline hazard rate \( h_0(t) \) is estimated from the solution to the minimization problem

\[
\hat{\theta}^*(t) \equiv \arg \min_{\theta} \sum_{n=1}^{N} \frac{1}{b_r} \int_{0}^{\tau} \left[ \Delta N_{nr}(u) - X_{t-u} \theta \right]^2 K_r \left( \frac{t-u}{b_r} \right) Y_{nr}(u) \exp(Z_u \hat{\beta}_r) du,
\]

where \( X_{t-u} \theta = \theta_0 + \theta_1 (t-u) \) denotes the local linear trend with parameter \( \theta = (\theta_0, \theta_1)' \) and the regressors \( X_{t-u} = (1, t-u) \). The optimal parameters \( \hat{\theta}^*(t) = (\theta^*_0(t), \theta^*_1(t))' \) solve the system of equations

\[
\sum_{n=1}^{N} \frac{1}{b_r} \int_{0}^{\tau} X_{t-u} \left( \Delta N_{nr}(u) - X_{t-u} \theta \right) \exp(Z_u \hat{\beta}_r) K_r \left( \frac{t-u}{b_r} \right) Y_{nr}(u) du = 0,
\]

which can be written as

\[
\sum_{n=1}^{N} \frac{1}{b_r} \int_{0}^{\tau} X_{t-u} X_{t-u} K_r \left( \frac{t-u}{b_r} \right) Y_{nr}(u) \exp(Z_u \hat{\beta}_r) du \times \theta
\]

or, equivalently,

\[
c^*_0 \theta_0 + c^*_1 \theta_1 = \frac{1}{b_r} \int_{0}^{\tau} K_r \left( \frac{t-u}{b_r} \right) J_r(u) dN_r(u),
\]

\[
c^*_1 \theta_1 + c^*_2 \theta_2 = \frac{1}{b_r} \int_{0}^{\tau} (t-u) K_r \left( \frac{t-u}{b_r} \right) J_r(u) dN_r(u),
\]

with

\[
c^*_m = \frac{1}{b_r} \int_{0}^{\tau} (t-u)^m K_r \left( \frac{t-u}{b_r} \right) S_r^{(0)}(\hat{\beta}_r, u) du, \quad m = 0, 1, 2.
\]

The solution is the estimator

\[
\hat{h}_{0r}(t) = \frac{1}{b_r} \int_{0}^{\tau} X_{t-u} \theta^*(u) K_r \left( \frac{t-u}{b_r} \right) du
\]

\[
= \frac{1}{b_r} \int_{0}^{\tau} \hat{\theta}_0(u) K_r \left( \frac{t-u}{b_r} \right) du + \frac{1}{b_r} \int_{0}^{\tau} \hat{\theta}_1(u)(t-u) K_r \left( \frac{t-u}{b_r} \right) du,
\]
which can be represented in the standard kernel form

\[ \hat{h}_{0r}(t) = \frac{1}{b_r} \int_0^t K^* \left( \frac{t-u}{b_r} \right) J_r(u) dN_r(u), \]

with the linearly adjusted kernel

\[ K^*_r(s) = \frac{c_2^* - c_1^* s}{c_0^* c_2^* - c_1^* c_2^*} K_r(s). \]

Note the stochastic character of the corrected kernel that makes automatic adjustment near the boundaries.

4.5 Adaptive nearest neighbor estimation of the baseline hazard functions

In empirical applications the durations used for estimation of the hazard functions are rarely uniformly distributed. When the sample of observed durations is heavily skewed, the usual kernel nonparametric estimates of baseline hazards with global bandwidth provide poor approximations for the true baseline functions in the low density domain of observed durations. Since the researchers in economics and finance are often interested in the behavior of a system in the regions of high duration, as the extremely high durations represent rare events which are frequently poorly modeled by conventional econometric techniques. Therefore an optimal estimation of the functional forms in these regions may also be crucial in simulation experiments and in the global performance evaluation of such econometric models.

Two practically relevant problems with estimation of hazard rates under the random censoring that has been extensively in the review paper by Müller and Wang (1994) are the boundary effects near the endpoints of the support of the hazard rates, in particular, around the origin, and the substantial, often explosive, increase in the variance in the range of extremely high durations. In this section we argue that the use of an adaptive external kernel smoother defined as the ratio of the smoothed numerator to the smoothed denominator in the original formula substantially reduces the bias in the range of extremely short durations without increase in the asymptotic variance. The external smoother has also some advantages in the range of extremely high durations, since its denominator accommodates the dynamics of the risk index at the censored durations which dominate the right tail of the duration range for any notional risk.

With regard to the second problem we advocate the use of the nearest-neighbor smoother which engenders many attractive properties of the fixed-bandwidth kernel smoothers but is more flexible with regard to the choice of bandwidth parameter. The flexible choice of bandwidth parameter in combination with the enhanced efficiency of the “external” estimator allow substantial reduction in the variance of the estimator in the regions of
sparse sample design. However, one should be aware of the fact that in the presence of
large number of competing risks this improved efficiency is still insufficient to warrant
precise estimation of the baseline hazard functions.

As has been mentioned in the beginning of this section, successful estimation with high-
frequency databases requires efficient computational methods. We estimate an adaptive
version of local kernel regression, which is a version of the kernel smoother procedure
\texttt{ksm.ado} in \textit{Stata}. The estimator is a weighted average of the \( k \) nearest neighbors to
the point of interest \( t \). The neighbors are selected in such a way that \( \lfloor k/2 \rfloor \) of them lie
to the left of point \( t \) and \( \lceil k/2 \rceil \) lie to the right of \( t \). Unlike the regular nearest neighbor
estimator, the weights are selected on the basis of distance (rather than rank ordering)
between the neighbor and the target point \( t \). This method assigns heaviest weights to the
closest neighbors of the target point, without regard to their rankings in the left and right
orderings. If the number \( j^* \) of available durations to the left (right) of the target is less
than \( \lceil k/2 \rceil \), the estimation is performed using all \( j^* \) observations to the left (right) and
the usual number of \( \lceil k/2 \rceil \) durations to the right (left) of the estimation point. Finally,
the weights can be determined by the tricube kernel (Cleveland, 1979)

\[
K(u) = (1 - |u|^3)^3 \cdot 1_{[-1,1]}(u),
\]

which is twice continuously differentiable and has the mass which is concentrated more
heavily around the center of bandwidth spectrum.

5 An Empirical Application: Liquidity and Order Flow in the Electronic Forex Limit Order Market

5.1 Description of events and covariates

We specify \( R = 46 \) types of buyer- and seller-initiated events and select \( S = 14 \) types
responding to the events that can be observed on the trading screens. The definitions
of sell side events are given in Table 1. Buy side events can be defined similarly.

The limit order arrivals are denoted by “A” followed by a numerical index that depends
on the order aggressiveness. Similarly, the cancellations of sell limit orders are denoted
by “AC” followed by a numerical index that depends on the distance between the limit
order price and the best market ask quote. Stars and double stars in the left column of
the table indicate the types of events corresponding to publicly observable events that
affect information on the screen about the state of the limit order book. Double stars
indicate the events that can (and usually will) trigger immediate trade execution. Note
that sometimes trades will not be executed automatically following order crossings because
of the lack of mutual credit among the counterparties, communication delays, etc.
<table>
<thead>
<tr>
<th>Risk (r)</th>
<th>Limit order price $P$</th>
<th>Price change</th>
<th>Quantity change</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1**</td>
<td>$P &lt; P_{bid}$</td>
<td>$P_{bid} ↓ P_{ask} ↓$</td>
<td>$Q_{bid} ↓ \uparrow Q_{ask} ↓ \uparrow$</td>
</tr>
<tr>
<td>A2**</td>
<td>Market sell order</td>
<td>$P_{bid} ↓$</td>
<td>$Q_{bid} ↓ \text{if } \Delta P_{bid} = 0$</td>
</tr>
<tr>
<td>A3**</td>
<td>$P = P_{bid}$</td>
<td>$P_{bid} ↓ P_{ask} ↓$</td>
<td>$Q_{bid} ↓ \uparrow Q_{ask} ↓ \uparrow$</td>
</tr>
<tr>
<td>A4*</td>
<td>$P - P_{ask} &lt; -1$</td>
<td>$P_{ask} ↓$</td>
<td>$Q_{ask} = Q$ (new ask size)</td>
</tr>
<tr>
<td>A5*</td>
<td>$P - P_{ask} = 1$</td>
<td>$P_{ask} ↓$</td>
<td>$Q_{ask} = Q$ (new ask size)</td>
</tr>
<tr>
<td>A6*</td>
<td>$P - P_{ask} = 0$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A7</td>
<td>$P - P_{ask} = 1$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A8</td>
<td>$P - P_{ask} = 2$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A9</td>
<td>$P - P_{ask} = 3$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A10</td>
<td>$P - P_{ask} = 4$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A11</td>
<td>$P - P_{ask} = 5$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A12</td>
<td>$5 &lt; P - P_{ask} \leq 10$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A13</td>
<td>$10 &lt; P - P_{ask} \leq 20$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>A14</td>
<td>$P - P_{ask} &gt; 20$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC6*</td>
<td>$P - P_{ask} = 0$</td>
<td>$P_{ask} ↓$</td>
<td>$Q_{ask} ↓ \text{if } \Delta P_{ask} = 0$</td>
</tr>
<tr>
<td>AC7</td>
<td>$P - P_{ask} = 1$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC8</td>
<td>$P - P_{ask} = 2$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC9</td>
<td>$P - P_{ask} = 3$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC10</td>
<td>$P - P_{ask} = 4$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC11</td>
<td>$P - P_{ask} = 5$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC12</td>
<td>$5 &lt; P - P_{ask} \leq 10$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC13</td>
<td>$10 &lt; P - P_{ask} \leq 20$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
<tr>
<td>AC14</td>
<td>$P - P_{ask} &gt; 20$</td>
<td>No price effect</td>
<td>No quantity effect</td>
</tr>
</tbody>
</table>

Since the events associated with changes of subsidiary quotes and the quantities available at these quotes are not included into the public information domain, in spite of the fact that some of those events can be potentially observed by some market participants. So we essentially follow the logic of the model developed in the previous sections according to which the subsidiary events do not restart the “internal clock” of the “race” between competing risks. Thus all types of events except A1 through A6, AC6, B1 through B6, and BC6 are assumed unobservable.

The number and types of events are chosen to provide a sufficiently fine partition across the events types on the one hand and to allow flexible estimation of a sufficiently rich covariate structure on the other hand. Given such limitations, for each type of event we select the components of vector $\mathbf{Z}_n \ (n = 1, \ldots, N_r)$ associated with the characteristics of the limit order book and the events of the recent trading history shown in Table 2.
## Definitions of the Covariates Characterizing the State of Limit Order Book and Trading History

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positioning bias</td>
<td>Current midquote minus price of last trade</td>
</tr>
<tr>
<td>Spread</td>
<td>$\max(P_{ask} - P_{bid}, 0)$</td>
</tr>
<tr>
<td>$\Delta P_{ask}$</td>
<td>Change of ask quote between last two events</td>
</tr>
<tr>
<td>$\Delta P_{bid}$</td>
<td>Change of bid quote between last two events</td>
</tr>
<tr>
<td>$\log(Q_{ask})$</td>
<td>Log depth at $P_{ask}$, or log $10^M$, whatever is smaller</td>
</tr>
<tr>
<td>$\Delta \log(Q_{ask})$</td>
<td>Change of $\log(Q_{ask})$ if $\Delta P_{ask} = 0$, zero otherwise</td>
</tr>
<tr>
<td>$\log(Q_{bid})$</td>
<td>Log depth at $P_{bid}$, or log $10^M$, whatever is smaller</td>
</tr>
<tr>
<td>$\Delta \log(Q_{bid})$</td>
<td>Change of $\log(Q_{bid})$ if $\Delta P_{bid} = 0$, zero otherwise</td>
</tr>
<tr>
<td>Sell/Buy</td>
<td>1/-1 if last trade is seller-/buyer-initiated</td>
</tr>
<tr>
<td>$F_{0.5^*}$</td>
<td>Signed number of trades 0 to 5 sec. prior to last event</td>
</tr>
<tr>
<td>$F_{5.10^*}$</td>
<td>Signed number of trades 5 to 10 sec. prior to last event</td>
</tr>
<tr>
<td>$F_{10.15^*}$</td>
<td>Signed number of trades 10 to 15 sec. prior to last event</td>
</tr>
<tr>
<td>$F_{15.30^*}$</td>
<td>Signed number of trades 15 to 30 sec. prior to last event</td>
</tr>
<tr>
<td>$F_{30.60^*}$</td>
<td>Signed number of trades 30 to 60 sec. prior to last event</td>
</tr>
<tr>
<td>$F_{1.2^*}$</td>
<td>Signed number of trades 1 to 2 min. prior to last event</td>
</tr>
<tr>
<td>$F_{2.5^*}$</td>
<td>Signed number of trades 2 to 5 min. prior to last event</td>
</tr>
<tr>
<td>$T_{0.5^*}$</td>
<td>Number of trades 0 to 5 sec. prior to last event</td>
</tr>
<tr>
<td>$T_{5.10^*}$</td>
<td>Number of trades 5 to 10 sec. prior to last event</td>
</tr>
<tr>
<td>$T_{10.15^*}$</td>
<td>Number of trades 10 to 15 sec. prior to last event</td>
</tr>
<tr>
<td>$T_{15.30^*}$</td>
<td>Number of trades 15 to 30 sec. prior to last event</td>
</tr>
<tr>
<td>$T_{30.60^*}$</td>
<td>Number of trades 30 sec. to 1 min. prior to last event</td>
</tr>
<tr>
<td>$T_{1.2^*}$</td>
<td>Number of trades 1 to 2 min. prior to last event</td>
</tr>
<tr>
<td>$T_{2.5^*}$</td>
<td>Number of trades 2 to 5 min. prior to last event</td>
</tr>
<tr>
<td>$T_{5.15^*}$</td>
<td>Number of trades 5 to 15 min. prior to last event</td>
</tr>
</tbody>
</table>

The components of vector $\mathbf{Z}_n$ incorporate the information which is available on the screen (at least, in principle) and closely monitored by traders (c.f. Figure 1). The *positioning bias* of the midquote relative to the last trade can be interpreted as the shift of the market quote identified with the midpoint of the bid-ask spread that has not been profit accrued to the trader participating in the last transaction if she liquidates her last trade position at the mid-point of the current bid-ask spread.\(^7\) The size of the market bid-ask spread is often identified with an intuitive notion of illiquidity in the market microstructure literature, and is expected to have a strong impact on the types of submitted orders. The two market depth variables represent the second dimension of liquidity, viz. how many units of asset can be bought (or sold) at the current ask (or bid) market prices. The depth variables are also expected to be significant for the risks of cancellations since the likelihood of a cancellation event can be positively related to

---

\(^7\) This interpretation of positioning bias disregards the transaction cost, which is always incurred by aggressor (the counterparty initiating the trade) according to the trading protocol of the D2000-2 trading system.
the total number of active limit orders, and the latter number should be correlated with
the quoted depth at the best market price.\(^8\) The changes of quoted prices and quantities
capture indirectly the traders' reaction to changes of the publicly available limit order
book information. \textit{Side}, which is the indicator of direction for the previous transaction,
characterizes the asymmetry in the impact of \textit{completed transactions} on the hazard rates as
opposed to the asymmetry in the impact of aggressive quotes captured by other variables,
since the quotes that occur without transactions only indicate the \textit{intention to trade},
not the actual trades. There is a strong evidence that the buy-sell indicator has a high
predictive power for the direction of future transactions on the foreign exchange market
(Goodhart \textit{et al.}, 1996) and on the stock markets (Hausman \textit{et al.}, 1992, Lo \textit{et al.}, 2002,
Huang and Stoll, 1997).

Finally, the choice of the lagged order flow and trade covariates is an attempt to
incorporate influence of the common factors contributing to the unobserved heterogeneity,
which cannot be captured by the current state of the limit order book. The presence
of common unobserved heterogeneity can invalidate the conditional independence which
is one of the crucial assumptions in the competing risks framework. Inclusion of the
covariates representing the trade and activity history captures a substantial part of the
lower-frequency serial dependence and also strikes some balance between identification
and empirical tractability of the model.\(^9\)

5.2 Estimation results

Tables 3, 4, and 5 report the signs of the estimated coefficients of the Cox proportional
hazard covariates for the competing risks of arrivals and cancellations of sell-side limit
orders and for the sell market orders. Only the events recorded between 6 a.m. and 5
p.m. GMT on October 6–8, 1997, which are the most liquid trading hours in the Reuters
D2000-2 trading system have been used for the analysis. Significance of the signs of
covariate coefficients is determined by the robust \textit{t}-statistics (Lin and Wei, 1989) at the
95\% level. The qualitative effects of covariates on the buy-side events are determined
analogously and almost symmetrically to the covariate effects on the sell-side events.

---

\(^8\)The determinants of limit order submission strategies are briefly discussed in O'Hara (1995) and
Goodhart and O'Hara (1997). The main theoretical contributions on the topic are Chakravarty and
(1996), and Brown \textit{et al.} (1999) contain interesting empirical results.

\(^9\)An alternative lagged activity measure represented by amplitude of the transaction price over the
period leads to qualitatively similar estimation results. Extension of the model to incorporate the dynamic
error correction terms in the spirit of the ACD model (Engle and Russell, 1997) is currently under
investigation. Results of that study will be reported in a separate paper.
TABLE 3

Estimated Coefficients for Price and Quantity Covariates in Cox Regressions for Competing Risks of Seller-Initiated Events

<table>
<thead>
<tr>
<th>Risk (r)</th>
<th>Pos.bias</th>
<th>Spread</th>
<th>∆Pask</th>
<th>∆Pbid</th>
<th>Qask</th>
<th>∆Qask</th>
<th>Qbid</th>
<th>∆Qbid</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1**</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A2**</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3**</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A4*</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5*</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6*</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A7</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A8</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A9</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A12</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A13</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AC6*</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>AC7</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC8</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>AC9</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC11</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC12</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC13</td>
<td>0</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

5.3 Empirical regularities associated with changes in the best bid and ask quotes and transaction prices (Table 3)

The arrivals of market sell orders and aggressive limit sell orders, as well as the ask price (but not quantity!) improvement events are more likely to occur after recent increases in the best ask prices. A similar regularity is observed on the buy side of the market with regard to recent decreases in the best bid prices.

The arrival and cancellation rates of limit orders a fixed number of ticks above current ask and below current bid market prices are negatively affected by recent deteriorations of best quotes on the same side and by recent improvements of best quotes on the opposite side. A similar negative reaction to recent improvements of best quotes on the opposite side is observed for the rates of price and quantity improvement. In fact, price improvements on one side of the market are more likely to be followed by reduced liquidity provision on the other side of the market.

There is a strong negative relationship between the intensity of transactions and the bid-ask spread, a strong positive relationship between the rate of large price improvement
and the bid-ask spread, and a fairly strong negative relationship between the subsidiary limit order arrival and cancellation activity and the size of the bid-ask spread.

Consistent with the theoretical prediction of Foucault (1999) and an empirical finding of Danielsson and Payne (2001, 2002), the hazard rates of aggressive limit order arrival tend to be more sensitive to quote and transaction returns than the hazard rates of market order arrival. Even though the reaction of the hazard rates to price changes is qualitatively similar for the market and aggressive limit orders, the sensitivity is larger for limit orders. The stronger is the upward price trend, the smaller is the proportion of market orders in the total aggressive sell order flow and the larger is the proportion of market orders in the total aggressive buy order flow.

Market sell orders are discouraged by buy price improvements; market buy orders are discouraged by sell price improvements. This “contrarian” property of market orders stands out as an opposite to aggressive limit orders that appear to be more frequent following improvements of prices on the opposite side. However one should be aware of the possibility of mechanical misclassification of limit orders as being “aggressive” based on the fact that more limit sell orders submitted at the same prices will overlap with higher bid quotes, and more limit buy orders will overlap with lower ask prices, even though the latter might be completely transitory price swings lacking an information content.

The fresh supply of liquidity (in the form of new non-aggressive limit orders) and demand for liquidity (in the form of aggressive limit and market orders) are negatively related to the size of the bid-ask spread. However, since cancellation rates of outstanding non-aggressive limit orders are also negatively affected by the size of the bid-ask spread, the effect of the spread on the net non-aggressive supply of liquidity is close to neutral. The only source of liquidity provision which is positively affected by the size of the bid-ask spread is large price improvements that occur when the spread exceeds two ticks. Other than that, supply of liquidity appears to be fairly steady and driven primarily by short-term price trends and by temporary fluctuations of quotes and depth of the limit order book.

The positive signs of the coefficients on Pos.bias for submission and cancellation rates of subsidiary sell limit orders and negative signs of these coefficients for submissions and cancellations of subsidiary buy limit orders provide evidence that the general level of subsidiary limit order activity declines when the recent transaction occurs at the price in their direction relative to the price of the previous transaction. However, the overall effect of the recent transaction price becomes not so obvious when the trade is accompanied by changes in quoted prices and quantities. Note that there is also migration of submission and cancellation events from one class to another every time the reference quotes (P_{ask} or P_{bid}) are modified. We decided to keep the Pos.bias covariate since it significantly improves the fit of the Cox regression.
5.4 Empirical regularities associated with depth of the limit order book at the best bid and ask quotes (Table 3)

The rates of arrival for market sell orders and aggressive limit sell orders, as well as the rates of ask price and quantity improvements are higher when the depth on the ask side of the market is high. A similar effect is observed on the buy side of the limit order book. This points to the competition among limit order traders for priority of trade execution as one of the driving forces behind price improvements and higher rates of transactions.

The rates of arrival for subsidiary limit orders at the prices that are close but inferior to the market bid and ask quotes are negatively affected by large depth at those prices. This is also an evidence of competition among limit order traders as the large amount of liquidity available on the market discourages limit order submission at inferior prices.

The rates of quantity improvement at the best market quotes increase immediately after the depth improvement at the same quote. This provides evidence that accumulation of additional liquidity at the best market quotes might be interpreted as a cautionary signal to “stay with the crowd”. Curiously enough, a similar reaction to quantity improvements is observed for the rates of market order arrival submitted by traders competing for liquidity with quantity improving limit order traders but not for the rates of price improvement on the same side of the market. In fact, this mechanism might be at work only at the instances when the possibilities of price improvement have been already exhausted (i.e., when the market spread is one tick or less).

There is a strong evidence that market depth improvements not accompanied by price improvement generally encourage cancellations of subsidiary limit orders on the same side and discourage demand for liquidity that comes in the form of aggressive limit and market orders from the opposite side of the market. In fact, the deteriorating depth level encourages traders on the opposite side of the market to chase more aggressively the remaining liquidity at the same price.

The rates of market bid and ask price deterioration are smaller when positively related to the depth available on the same side of limit order book on the same side of the limit order book.

The cancellation rates of limit orders at the best bid and ask prices decrease following increases in the observed depth on the opposite side of the limit order book. The increasing depth encourages a limit order trader on the opposite side of the book to keep his order at the best price but only if he knows that he will be the first (and the only one) to trade at that price.
**Table 4**

Estimated Coefficients for Lagged Order Flow in Cox Regressions for Competing Risks of Seller-Initiated Events

<table>
<thead>
<tr>
<th>Risk (r)</th>
<th>Sell/Buy</th>
<th>$F_{0.5}^+$</th>
<th>$F_{5.10}^+$</th>
<th>$F_{10.15}^+$</th>
<th>$F_{15.30}^+$</th>
<th>$F_{30.60}^+$</th>
<th>$F_{1-2'}$</th>
<th>$F_{2.5'}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1**</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A2**</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3**</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A4*</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5*</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>A6*</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A7</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A8</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A9</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A10</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>A12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A15*</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A16*</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A17*</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A18*</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A19*</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A20*</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
</tbody>
</table>

5.5 **Empirical regularities associated with the lagged signed order flow (Table 4) and the level of transaction activity (Table 5)**

The arrival rates of market sell orders, aggressive limit sell orders, and the market ask price improvements increase following the seller-initiated transactions and decrease following the buyer-initiated transactions. A similar effect takes place on the buy side of the limit order book. This can be viewed as a manifestation of persistence in the buyer and seller pressure which is consistent with the evidence of strong high-frequency directional momentum in foreign exchange reported in previous studies.

There is some evidence of higher cancellation rates for subsidiary sell limit orders up to ten seconds after seller-initiated transactions and lower cancellation rates for subsidiary sell limit orders up to ten seconds after buyer-initiated transactions. A similar effect is observed for subsidiary buy limit orders. This can be generally interpreted as an evidence of traders updating the limit order book within ten seconds after receiving the signal that the market is moving in the opposite direction.
### TABLE 5

Estimated Coefficients for Lagged Trade Activity in Cox Regressions for Competing Risks of Seller-Initiated Events

<table>
<thead>
<tr>
<th>Risk (r)</th>
<th>T0-5⁺</th>
<th>T5-10⁺</th>
<th>T10-15⁺</th>
<th>T15-30⁺</th>
<th>T30-60⁺</th>
<th>T1-2⁺</th>
<th>T2-5⁺</th>
<th>T5-15⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1⁺⁺⁺</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A2⁺⁺⁺</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A3⁺⁺⁺</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A4⁺⁺⁺</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A5⁺⁺⁺</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A6⁺⁺⁺</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A7</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A10</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A11</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>A12</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A13</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>A14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>AC6⁺⁺⁺</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC7</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>AC8</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>AC9</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>AC10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>AC11</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>AC12</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC13</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>AC14</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

There is some evidence that the prolonged periods of lower submission rates for subsidiary sell limit orders are more likely to occur after the periods dominated by seller-initiated transactions. On the other hand, there is only a weak evidence that similar effects are at work for subsidiary buy limit orders. Even though this asymmetry is likely to be period-specific, the general perception of the US dollar as a currency with stronger fundamentals than the Deutsche Mark translated into the longer memory of limit order positions after a strong “buy” signal can provide an alternative explanation.

There is some evidence that the order submission and cancellation rates increase after the periods of high transaction activity. The effect of transaction activity is persistent at all price levels of the limit order book. Finally, the effect and persistence of activity is stronger for the rates of subsidiary limit order arrivals and cancellations than for any other events. This points toward the subsidiary limit orders and, more generally, to stop-loss orders (Osler, 2001) as one possible transmission mechanism and the potential source of long memory in foreign exchange markets.
6 Conclusion

This paper proposes a unified competing risks framework the analysis of multivariate marked point processes in continuous time. The competing risks in this model have attractive behavioral interpretation as the salient characteristics of traders possessing diverse information and beliefs which is the salient feature of a typical foreign exchange market environment. The model can be applied with minor modifications to other irregularly spaced observations that are frequently encountered in applied finance and economics. The asymptotic theory developed by Andersen et al. (1993) for counting processes can be adopted with minimal adjustments to conduct the inference for competing risks. The model of this paper represents a semiparametric alternative to the fully parametric model of limit order trading by Bisière and Kamionka (2000). The method of this paper can be applied more generally without imposing unjustified restrictions on the form of the baseline hazards, the practice leading to potential biases in the estimated covariate effects and incorrect inferences.

The competing risk model is used to analyze the timing and interactions between quotes and trades in the Reuters D2000-2 electronic brokerage system. The major stylized facts about the foreign exchange trading that have been reaffirmed by the empirical application of the semiparametric competing risks model of this paper include the clustering of market activity on the directional characteristics of last trade (“buyer or seller pressure”) and the considerable sensitivity of the order submission strategies employed by traders to the state of the limit order book and the quoting and trading history.

The model developed in this paper can be potentially useful in applications to a variety of high- and low-frequency financial data. One natural example involves the empirical analysis of thinly or irregularly traded financial instruments, corporate bonds, and emerging market securities. In combination with more conventional pricing models, the methodology of the paper may be extended to the analysis of financial instruments (such as mortgage-backed securities and credit derivatives) that might be simultaneously affected by complex combinations of qualitative risks.
References


