High-Frequency Principal Components and Evolution of Liquidity in a Limit Order Market*

Konstantin Tyurin†
Department of Economics
Indiana University

November 12, 2003

Abstract

The paper applies a popular methodology of competing risks to the analysis of the timing and interaction between the Deutsche Mark/U.S. dollar transactions, quotes, and cancellations in the Reuters D2000-2 electronic brokerage system. Consistently with previous stock market studies, the bid-ask spread and market depth at the best bid and ask quotes are found to be major determinants of limit order market dynamics at ultra-high frequencies. Consistently with the microstructure approach to exchange rate determination, the signed transaction activity appears to be the main factor behind the limit order market dynamics at lower frequencies. Application of principal component analysis to the covariate indices of competing risks identifies five pervasive factors that capture 85% of the Reuters D2000-2 limit order book activity. The multifactor competing risks model substantially improves the quality of short-term probability forecasts for buyer- and seller initiated transactions, relative to popular moving average-type forecasting rules.

---

*I thank participants of seminars in Chicago Graduate School of Business, Michigan State University, and University of Michigan, and participants of the New Economic School Tenth Anniversary Conference in Moscow, Russia, for discussions and helpful comments on the earlier versions of this paper. Thanks also go to Reuters for permission to access the data set used in this paper.

†Correspondence to: Konstantin Tyurin, Department of Economics, Wylie Hall 105, Indiana University, 100 S. Woodlawn, Bloomington, IN 47405. E-mail: *ktyurin@indiana.edu*

JEL Classification: C4, G14, D44

Key Words and Phrases: foreign exchange, limit order, market order, order flow, liquidity, competing risks, principal component, probability forecast.

1 Introduction

The present paper is aimed to enhance our understanding of the short-run dynamic links between liquidity provision and order flow in the order-driven segment of the foreign exchange market. It studies dynamic interactions between quotes and transactions in the Reuters D2000-2 foreign exchange electronic trading system on the one hand, and the publicly available proxies of liquidity such as market depth and size of the bid-ask spread on the other hand. The competing risks structure of the model facilitates decomposition of the order flow into a number of distinct components identified by the direction and aggressiveness of trading activity. Therefore, it is particularly well suited to study the relative strength and interaction of passive and aggressive order flow on the buy and sell side of the limit order book. It is found that the limit and market order trading activity is very sensitive to the publicly available information on the state of the limit order book. Another major vehicle behind the dynamics of the price discovery process is the signed order flow, defined in this paper as the difference between the number of buyer-initiated and seller-initiated transactions.

The competing risks model developed in this paper can be viewed more generally as a comprehensive partially disaggregated description of trading activity, where the level of disaggregation is ultimately determined by the objectives of study and the data limitations. This methodology can be used as an agnostic behavioral framework for learning about activity levels in financial markets populated by multiple interacting agents with heterogeneous information and diverse trading strategies. It will be shown how the model can generate probability forecasts of short-term market dynamics that might facilitate the development of short-term trading strategies by dealers and their sponsors.

The paper proceeds by the following plan. Section 2 gives the literature review on liquidity and the order flow. Section 3 describes a stylized model of trading activity in an ideal limit order market reminiscent of the electronic limit order book studied in the empirical application and outlines the objectives of empirical analysis. Section 4 presents the competing risks model and shows how its structure can incorporate observable and unobservable limit order book characteristics. Section 5 reports the estimates of covariate indices that capture a large portion of cross-sectional and serial correlation between various types of trading activity in the Reuters D2000-2 electronic limit order book. Section 6 applies the principal component analysis (PCA) to the covariate indices of competing risks estimated in the previous section and identifies five pervasive factors (principal components) that capture 85% of the observed activity in the Reuters D2000-2 electronic limit order market and have distinct characteristics that facilitate their interpretation. Section 7 applies graphical diagnostics to evaluation of goodness-of-fit and out-of-sample forecasting power of the principal components. Section 8 concludes. The technical appendix describes semiparametric estimation procedure of the competing risks model and gives the formulas for probability forecasts.

2 Literature Review

A large body of empirical and theoretical literature on market microstructure is motivated by the central question of modern finance: How long does it take for information to be fully incorporated into prices? According to the microstructure approach, private information in financial markets is ultimately transmitted via continuous interaction of quotes and transactions. Therefore, it is natural to expect that the frequency and complexity of arrival patterns for quotes and transactions, as well as the choice of the trading mode, should be among the key ingredients of any transition mechanism to the efficient price.

It has been long recognized in theoretical microstructure literature (Diamond and Verrecchia [11], Admati and Pfleiderer [1], [2], Easley and O'Hara [12], O'Hara [41], Chapter 6) that at least some non-trading and quote delays can be purposeful and informationally motivated. For example, a Bayesian market maker in Easley and O'Hara [12] infers about the presence of informed traders on the market from the intertrade durations, with shorter durations signalling the informed trader activity. In applications, a purely statistical model for duration process is almost always augmented by some proxy characterizing the amount of private information on the market, such as the market bid-ask spread, trading volume, etc. The choice of such proxy is justified either theoretically by market microstructure models of price determination or empirically by availability of appropriate data at high frequencies.

In early theoretical models, the choice of trading mode is usually restricted to market orders that could be filled at the prices provided continuously by the market maker, or, alternatively, to limit orders for a fixed quantity that remained valid for a single time period and expired automatically if the order was unfilled. This obvious simplification was a price to pay for analytical tractability and crisp empirical implications of the models. For example, in static models of the automatic limit order book by Glosten [20] and Chakravarty and Holden [8], the focus is on the optimal bidding strategies of limit order traders who are unwilling, or unable, to use market orders. A more recent strand of dynamic trade execution models (Parlour [44], Foucault [16]) emphasize the importance of the risk of non-execution and the risk of being picked off by informed traders for the order placement strategies employed by uninformed market participants. The first analytical results on the non-trivial interplay between the limit order price and the time-to-execution in complex dynamic environments began to emerge only recently. One such example is the paper by Foucault, Kadan and

Kandel [17], where the interactions between the trading decisions of patient and impatient traders play central role in the determination of dynamic equilibrium quotes and the bid-ask spread.

Even though the information processing lags, random delays, and occasional congestions in communication networks can provide partial explanation for the high level of noise and irregularity of intertrade and interquote durations, their persistence and high correlation with economically relevant variables such as the bid-ask spread and price volatility warrant the closer look at the duration processes. At least, it would be fair to say that discounting intertrade durations as pure noise cannot be justified in many microstructure applications, whereas the attempts to fully explain the duration dynamics by purely technical factors are likely to tell only part of the story.

It should be clear from the above discussion that the heterogeneity of investors emerges as one of the driving forces behind the nontrivial properties of duration dynamics. At the same time it must be emphasized that the timing of transaction needs is not always synchronized across traders, even in the absence of traders with superior information. Traders always have a choice between the submission of a market order that will be filled immediately at the best price available at the moment of submission, and the submission of a limit order that can improve the execution price for the submitting counterparty at the cost of delay and uncertainty of execution. As a result, the variations in traders' demand for immediacy of execution affects simultaneously the bid-ask spread and the depth of the market, driving the dynamics of the limit order book and the market liquidity. Investigation of the non-trivial dynamic relationship between three dimensions of liquidity, such as *immediacy* (the ability to trade a given quantity at a given cost), *breadth* (the cost of doing a trade of a given size quickly represented by the bid-ask spread), and *depth* (the size of trade that can be dealt quickly at a given cost) then naturally becomes the central object of research.[1]

The complete theoretical analysis of market liquidity in a realistic general equilibrium framework remains the major challenge for the market microstructure literature. The problem of solving and analyzing the multiple equilibria can be extremely complicated since traders' choices are not restricted to one parameter such as the price or quantity, but also include the decisions between the limit and market orders, sell and buy orders, as well as about the timing of order execution. Moreover, in real markets the traders can cancel and resubmit strategically their orders at any moment of time. Since full analysis of such a dynamic game remains excessively complex and just impractical to implement, the theoretical literature so far has been focusing on one or two dimensions of traders' decision, holding all other variables fixed or making other simplifying assumptions about traders' behavior.

In an attempt to better understand the decisions made by traders in the real world, many researchers concentrated their efforts on the search for stylized facts and empirical regularities

[1] See the discussion of different aspects of liquidity in Chapter 19 of Harris [26].

with regard to different aspects of supply and demand for liquidity in the real markets. This effort was facilitated by broader availability of transaction and quote data at intradaily frequencies. Since this search was originally restricted to the markets that make such data readily available, most of empirical papers in this area focused on the stock markets across the globe where systematic collection of data on quote and transaction activity was part of the institutional design enforced by financial regulators.[2] Sometimes the data were also collected by private enterprises and provided to researchers on proprietary basis. Examples of stock market analyses include Biais, Hillion and Spatt [4], Handa and Schwartz [25], Harris and Hasbrouck [27], Lo, MacKinlay and Zhang [37], Griffith *et al.* [23], and Hollifield, Miller and Sandås [31]. For example, Biais, Hillion and Spatt [4] found that market orders in the Paris Bourse consume the major portion of liquidity available on the opposite side of the bid-ask spread, which then reverts to its original level as the limit order traders place new orders within the best bid and ask quotes. Even though most of transactions in the Paris Bourse occur at the small values of spread, the authors observed high-frequency negative autocorrelation between the quotes, as the spread showed a tendency to alternate between small and large values.

All of these studies either avoid the analysis of cancellation events altogether, or make very simplistic assumptions about traders' cancellation policy for the existing limit orders. The results of empirical studies taking this problem more seriously suggest that going beyond the trivial assumption might be a challenge. For example, no systematic studies presently available explain rigorously the empirical finding of Hasbrouck and Saar [29] who find that the majority of limit orders submitted on Island ECN that are cancelled, get cancelled within the first couple of seconds after submission. Needless to say, incorporating nontrivial order cancellation strategies, even though crucial for understanding the dynamics of liquidity, is likely to make the theoretical analysis far more complex, if not infeasible. At present the most promising approach appears to be the accumulation of additional stylized facts about alternative aspects of liquidity in financial markets and interpretation of their properties and interaction within an appropriate statistical framework.

3 The Anatomy of a Pure Limit Order Book

To motivate the modeling strategy and provide the background for the empirical results of this paper, we give a stylized description of the automated limit order book. While technical details and the peculiarities of actual limit order markets must always be accounted for

[2] It must be emphasized that empirical research of liquidity patterns reveals a surprising degree of similarity across alternative market instruments, trading organizations, and locations. In particular, we believe that implications of our research will be mostly valid for a broad range of markets organized as electronic limit order books.

in specific applications, our presentation here is deliberately simplified. Even though it has been developed bearing in mind the application to the electronic segment of foreign exchange market in section 5, in this informal review we pay special attention to the salient features shared by all markets organized as pure electronic limit order books.

In the forthcoming development, the *limit order* is defined as an instruction to sell (or buy) a certain number of units of financial asset at a certain price, which is called the *limit order price*. We define *market order* as an instruction to sell (or buy) a certain number of units of financial asset immediately at the best available market price. While market orders always face full and immediate execution, the limit orders may face only partial execution, or may not be filled at all, in which case the untraded portion of the limit order is placed into one of the two queues (separate for sell and for buy limit orders) and kept in the queue until the order is explicitly cancelled by the trader or hit by the arriving market or limit order.

A bank dealer can enter a buy or sell limit order into the system at any moment of time, indicating the limit order price and the quantity of foreign exchange (usually an integer number, in millions of US dollars) that he wants to trade. After an attempt to match the incoming order with outstanding orders submitted by other traders on the opposite side of the limit order book, the new order is entered into the system. Additionally, traders have an option to submit a buy or sell market order indicating the price and the quantity. In this case, after an attempt to match the incoming market order with outstanding limit orders at the price *equal* to the incoming market order, an unexecuted portion of the market order is cancelled automatically and a confirmation message to the trader is sent. The system allows traders to monitor their limit orders that can be removed by a hit of a "Cancel" key at any moment of time.

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*. 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 1). The priority of limit orders in the book depends on the details and the trading protocol of the particular system. All orders are submitted 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, such as the one described above. In practice virtually any real limit order market represents a hybrid system, which would be hard to confront directly with the simple model presented in this paper. For example, the trader submitting a market order 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 at the time of submission. 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 unfilled 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 may be willing to submit a market order and sacrifice the difference between the best bid and ask quotes (the *bid-ask spread*), or submit a limit order which gives them a chance of getting a better price at the expense of execution uncertainty. The seller who is willing to receive a better execution price may submit a limit order with a relatively high price and be prepared to wait longer and deal with the risk of execution uncertainty in the event that the market price moves 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.

The dynamics of a limit order market are illustrated on a series of graphs (Figures 2–8). Figure 2 provides graphical representation of the situation when a limit order trader submits a bid to buy one million dollars at the price 1.7510 DEM per US dollar, which exceeds by one tick[3] the previous best market price 1.7509 DEM, but falls two ticks short of the price that would guarantee immediate execution. Alternatively, the trader can submit a less aggressive limit buy order at the current best market price (Figure 3). 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. In the first case the traders watching monitors of the trading screens notice immediately the increase in the best bid market price, while in the second case they will only observe the increase in the quantity available at the unchanged best bid market price. Whether and how this public information arrival affects the behavior of other traders remains one of the central issues in the empirical and theoretical market microstructure literature.

The limit order buyer has yet another option. He can avoid revealing any information to the market about his intentions and willingness to trade if he submits a *subsidiary limit order*, which can be an order to buy at the price just below the current bid market price (Figure 4) or two ticks (minimal increments) below the current bid market price (Figure 5). In both cases the best market prices and quantities available to sellers and buyers at these prices remain unchanged. Therefore, the information on the screens is not affected by arrivals of subsidiary limit orders.

The last three diagrams provide illustrations of the effects of a subsidiary limit order cancellation (Figure 6), 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 7), and a similar transaction initiated by an aggressive limit order buyer (Figure 8). Note the difference between the effect of market versus aggressive limit order arrival on the market liquidity. While the market order buyer only consumes the liquidity available at the best

[3] One tick = 0.01 Prennig.

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 attitude.

In the liquid financial markets such as the segment of foreign exchange market studied in this work, interactions between the limit order arrivals, cancellations, and transactions, similar to those described above, occur virtually every second. Figure 9 shows a representative small subsample, which was the result of such interaction in continuous time. The time period on Figure 9 covers ten minutes of fairly active morning trading on Monday, October 6, 1997. The light solid curves show evolution of the best bid and ask prices available on the market. The crosses and circles mark the times and price levels of buyer- and seller-initiated transactions. The chart on Figure 9 illustrates the empirical fact that substantial fraction of activity in the electronic segment of foreign exchange market is not accompanied by any transactions, but rather represents the reaction of traders who submit and cancel subsidiary limit orders in response to prior market events. The analysis of dynamic interaction of limit order traders using the screen information on the limit order book will be the main object of investigation in this paper.

3.1 Data

Empirical research on the limit order markets has long been hampered by the lack of detailed order-level data. Until recently, most order-level data came from the stock exchanges organized as electronic limit order books. Examples are Biais, Hillion, and Spatt [4], Hamao and Hasbrouck [24], Harris and Hasbrouck [27], Hollifield, Miller, and Sandas [31], who investigate empirical properties of limit order markets in Paris, Tokyo, New York, and Stockholm. Even though there are now several electronic systems trading currencies, corporate and government bonds, and other financial instruments, the information disclosure restrictions on the providers of such systems in the foreign exchange markets usually make the detailed order-level data unavailable for academic researchers.

The data set made available by Reuters and provided by the Financial Markets Group at LSE covers the trading days from October 6 to October 10, 1997 and also contains a few orders originating late on October 5. As mentioned in the introduction, the only similar data previously available to academics is a short compilation of quotes from a seven-hour videotape of the D2000-2 screen dated by June 16, 1993 (Goodhart *et al.* [22]). The data contain information about 130,535 limit and market orders made on the bid or ask side of the market. Each line of the file represents one limit or market order and contains exact entry and exit times, price and quantity ordered, quantity dealt, and information on whether the order came on the bid or offer side of the market. In addition there are a few other entries on each line, most of them redundant, which are used to validate the information on the reported order characteristics.

The data set does not include confidential information on the identity of dealers submitting orders and completing transactions. Another highly relevant piece of information which is observed by individual subscribers on the D2000-2 screens but cannot be inferred from the data set are the best bid and ask quotes and the quantities of foreign exchange available at these prices to any of individual traders. At every point in time these best quotes and quantities are based on the existence of mutual credit lines between any given subscriber and her potential trade partners submitting limit orders at this time. Although the quotes and quantities available to individual subscribers would coincide with the best quotes and quantities of the market much of the time, on average the effective bid-ask spreads encountered by individual traders are slightly larger than the market bid-ask spread. For the same reason, the quantities available for trade to individual subscribers might be different from the quantities on the market at any given point in time.

3.2 Empirical Microstructure Approach

In its pure form, the microstructure approach to foreign exchange determination[4] postulates that the order flow (broadly defined as the sequence of buyer- and seller-initiated transactions) is the only variable explaining the long-range dynamics of the foreign exchange. The microstructure approach has found a solid support in the recent theoretical and empirical literature uncovering the ability of the order flow to explain a large share of movements in the major floating exchange rates[5]. It is commonly assumed that the beliefs of market participants formed on the basis of differential information find their outlet in the form of order flow. The ability of order flow to capture prevailing motives of buyers and sellers who exploit private information on the future price dynamics is a well documented empirical fact. Even though financial price fluctuations ultimately impound the traders' interpretation of private and public information about the fundamentals, one can take a stand that at the intraday frequencies the order flow represents the dominant mechanism transforming traders' beliefs into the dynamic patterns of buyer- and seller-initiated transactions. For a number of reasons, given below, no attempt to model this mechanism explicitly will be made in this paper.

First, the focus of this paper will be on the *short-run dynamics* of market activity. Therefore, we ignore the dynamic effects of past activity at the lags longer than 15 minutes. The key order flow-related variables driving the dynamics of notional hazard rates will be the counts of buyer- and seller initiated trades in the five-second periods immediately preceding

[4] This approach was popularized by Lyons [38]; see also a brief review of policy implications in Lyons [39].

[5] See, for example, Evans and Lyons [15].

the sub-epoch, and the similar counts from the earlier history. We deliberately bypass macroeconomic fundamentals as there is remarkably little evidence that macroeconomic variables have any consistent effect at the ultra-short horizons which are the focus of this study.[6]

Second, the price dependence emphasized in the literature on technical analysis of foreign exchange markets will be restricted to the price bounces between the last three transactions that occur prior to the beginning of the analyzed duration period. It is important to stress that no *a priori* assumptions will be made about the valuation of foreign exchange. In our opinion, the ultimate value of foreign exchange is impossible to define without the reference to the "efficient price" and strong assumptions about its time series properties. In the financial markets literature it is often assumed that the "efficient price" is a continuous martingale driven by a Brownian motion with constant volatility.[7] This assumption is very intuitive and appears well justified when observation intervals are monthly, weekly, or even, with some reservations, daily, but appears to be at odds with the main objective of market microstructure literature which aims to study the process of price formation.[8] It is especially important to avoid making such a strong assumption at the high and ultra-high frequencies when even the very notion of "efficient price" becomes less transparent because of the price discreteness, illiquidity, bid-ask bounce, asynchronous trading,[9] and other effects which are at the center of market microstructure research agenda. The broadly recognized and accepted empirical fact that even the most sophisticated agents may have differential beliefs makes virtually any *a priori* assumption about the underlying fundamental value difficult to defend. However, the agents' information and beliefs, which are intrinsically unobservable, cannot be structured without making additional identifying assumptions about the process of information discovery.

In our agnostic approach, we focus attention on the hazard rates of alternative events and formulate the model using only a handful of observable characteristics that can be inferred from a subset of publicly available data. Thus, we postulate that the limit order book is the only medium for the price dynamics. No matter what foreign exchange rates – bid, ask, or the actual transaction prices – are considered, their dynamics are thought to be driven by interaction of supply and demand of multiple agents with differential information, horizons, beliefs, and trading strategies.

---

[6] The interest rate news announcements appear to be the only exception but even those take usually several hours to be absorbed in the market price.

[7] In the modern literature, more realistic assumptions about dynamic properties of the volatility and the drift term are usually made.

[8] Hasbrouck [28] discusses the role of the "efficient price" assumption in the market microstructure literature.

[9] Chapter 3 in Campbell, Lo, and MacKinlay [7] summarizes the early econometric studies that made attempts to resolve these problems.

As the exact moments of information arrivals are unavailable from the data, and rarely can be identified in practice, difficulties arise in determining a good proxy for information. Trading volume has been by far the most frequently used proxy in the empirical studies of stock returns. However, Jones, Kaul, and Lipson [35] in a comparative study of different information proxies demonstrate that trading volume has the same informational content as the number of trades. Similarly, Marsh and Rock [40] show that the net order flow (the number of seller-initiated minus the number of buyer-initiated transactions) explains as much of the price variation as does the signed volume of trade. Geman and Ané [19] show that the moments of increments of the time deformation process that makes returns on a stock market index normal closely match the moments of the number of trades for that index per minute. All this and other evidence indicates that the signed number of trades could be a better proxy for the information arrival than trading volume and may be an important factor behind the market volatility, and its persistence. Moreover, the occurrence and direction of trades are readily observable on Reuters D2000-2 trading screens, to the extent that a trader can distinguish between the flashes on the screen that accompany any new transaction that occur in the system. In view of this positive evidence and to the extent that other information proxies such as the number of quote changes, price changes, and so on, used in the empirical literature produced mixed results, we take the general premise of the order flow approach to exchange rate determination (Lyons [38]) and accept the signed number of trades as a primary vehicle behind the information arrival process.

In this paper we focus on the dynamic links and interaction between the order flow and various types of market activity. Specifically, we investigate the effect of the following publicly observable variables on the order submission and cancellation decision:

1. bid-ask spread, defined as the difference between the best ask and bid quotes available on the market;

2. price improvements and deteriorations on the same side and on the opposite side of the limit order book;

3. depth improvements and deteriorations on the same side and on the opposite side of the limit order book;

4. dynamic links between the order flow and various types of market activity;

5. trading intensity (trading volume);

6. order flow, defined as the difference between the number of buyer-initiated and seller-initiated transactions;

7. the out-of-sample forecasting power of the order flow and other variables associated with limit order book trading.

To keep this work focused on a relatively limited set of issues and avoid additional methodological and practical complications, we do not model seasonal effects and clustering of orders at the round numbers (multiples of five ticks, in the case of DEM/USD exchange rate). The intraday seasonality is undoubtedly an important empirical feature of virtually any financial data. Even though, in principle, the mechanical introduction of time-of-the-day diurnal effects or a simple deterministic trend in the specification of hazard rates may lead to a slightly improved fit of the model, this will involve additional methodological and practical problems. A substantial part of the detected seasonality is likely to be spurious, since the trading history covered by our data set includes only five full trading days. In turn, this may lead to incorrect inferences about the effects of other covariates, which are the main focus of the present paper.

We also refrain from modeling explicitly the empirically relevant features of real trading process such as random communication delays and failures, lack of mutual credit agreement among counterparties, occasional violations of order priority, potential implications of the complex architecture of communication networks, and so on. For most of this study we do not distinguish between the screen information available to the market and the screen information of individual traders that appears in a separate section of the Reuters D2000-2 trading screen, or comes from alternative sources. Clearly, ignoring the peculiarities of actual trading process may play a crucial role for the success or failure of model's predictive performance and for the quality of fit between the simulated trading histories and the real market data. However, we accept the lack of realistic representation of some aspects of actual limit order trading as a price to pay for the relative simplicity, analytical tractability, and methodological generality of the competing risks specification developed in this paper.

4 Statistical Model

4.1 Competing Risks

The competing risks model belongs to the wide class of semi-Markov models with a finite number of state variables z_1, z_2, ..., z_k. The trading history is divided into a finite number of intervals ("epochs") with random durations t_1, t_2, ..., t_N. In the beginning of every period $n = 1, 2, ..., N$, the market is assumed to be in a transient state $\mathbf{z}_n = (z_{1n}, z_{2n}, ..., z_{kn})'$ characterizing by the current limit order book and the recent trading history. It is assumed that any epoch can be terminated by occurrence of one (and only one) of the R distinct types of event ("risks"). Each risk should be easily identified by an action of market participants who can submit new orders or cancel outstanding limit orders. Occurrence of any event induces a change in the limit order book, but only some of these changes are publicly observable.

In the competing risks framework, we postulate the existence of latent durations associated with R risks that are simultaneously and independently drawn, conditional on the transient state $\mathbf{z}$. Only the smallest of these R durations is observed while the remaining durations are right-censored. The competing risks are characterized by transition intensities $h_r(t|\mathbf{z})$, $r = 1, ..., R$, which are termed "cause specific", or "notional" hazard functions, and interpreted as the arrival rates of type r events given that the current state of the market is $\mathbf{z}$ and no observable event occurred for t units of time.[10]

The model can be immersed in the framework of multivariate counting processes (Andersen *et al.* [3]) and estimated in continuous time (Appendix A).

4.2 Events

The choice of event types that has been discussed in the previous subsection is somewhat arbitrary and depends on the questions to be answered, data limitations, and the prior theoretical considerations about the data generating process. We specify $R = 46$ types of buyer- and seller-initiated events and select $S = 14$ "observable" types of events that can be identified on the trading screens by all market participants. Table 1 provides the definitions and summary of sell-side events, which occurs on the *ask side* of the limit order book and denoted by codes "A" and "AC" followed by numerical indices. Buy-side events, which occur on the *bid side* of the book and denoted by codes "B" and "BC" followed by numerical indices, can be defined similarly.

The type of event in Table 1 is defined as a combination of order type (market or limit order), character of activity (submission or cancellation of limit order), and the distance between the price P^* of the limit order and the prevailing best bid and ask quotes P_{bid} and P_{ask} prior to the event (column 2). The consequences for the best bid and ask prices and for the liquidity of the limit order book (market depth) at these prices are indicated in columns 3 and 4 of Table 1. The events associated with arrivals of market and limit *sell* orders are denoted by letter "A" (which means the activity occurs on the *ask side* of the limit order book) followed by a numerical index corresponding to the sell 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. The events marked in the first column of the table by single and double stars can be observed by all market participants. The events marked by double stars, which are also observable by all market participants, typically trigger immediate trade executions.[11] The unmarked types of events are associated with limit order arrivals or cancellations at suboptimal prices

[10] An adjustment of the model to the more realistic situation when the hazards of almost contemporaneous events depend on the state of the market before the earliest of those events occurs is straightforward. Estimation of the modified model leads to similar results.

[11] Occasionally the trades will not be executed automatically following order crossings because of the lack of mutual credit among the counterparties, communication delays, etc.

that cannot be observed on the D2000-2 trading screens and therefore constitute private information of traders.

The events associated with changes of subsidiary quotes and the quantities available at these quotes are not included into the public information domain, even though some of these events can be potentially observed by market participants. Therefore, we follow the general logic of the approach according to which the subsidiary events do not restart the "internal clock" of the "race" between competing risks and assume that all types of events except A1 through A6, AC6, B1 through B6, and BC6 are unobservable.

4.3 Covariates

In the competing risks framework, the choice of the state space depends on the range of prices, quantities, and other market characteristics, whose effect on the point processes is investigated. The range of possible specifications of the Markov space is ultimately determined by the objectives of study and often is severely restricted because of the data limitations. However, in loosely structured problems it is desirable to start with as broad set of variables as could be reasonably possible. It is desirable that the variables in the state space were linked to the market factors and identified from theoretical considerations. These variables can be discrete or continuous, and the state space might have a fairly complicated topological structure.[12]

The full covariate vector $\mathbf{z}$ selected for the analysis in this paper is described in Table 2. Along with the "usual suspects" such as the bid-ask spread and the depth of limit order book on the bid and ask sides, components of the covariate vector include the dynamic characteristics of market liquidity and order flow, such as the side of the limit order book where the recent transaction occur, several types of price and quantity changes, and so on. This choice of covariates is not unique could be easily modified depending on the structure of the available data set, for instance, if traders' identity or relevant information from other markets were available.

The components of the covariate vector $\mathbf{z}$ are divided into the three categories.

4.3.1 Price covariates

- *Slippage* $\equiv$ difference between the current midquote $\frac{1}{2}(P_{\text{bid}} + P_{\text{ask}})$ and the last transaction price (measured in 0.0001 DEM);
- $Spread_{>0}$ $\equiv$ quoted size of the bid-ask spread (measured in 0.0001 DEM) when it is positive, zero otherwise;

[12] For instance, if the purpose of research is to study the effect of order clustering at the multiples of five ticks, the most natural specification of a state variable distinguishes the orders that arrive at the multiples of 5 ticks, one plus multiples of 5 ticks, two plus multiples of 5 ticks, and so on, like it was done in Osler [43].

- $\Delta P_{\text{ask}} = P_{\text{ask}} - P_{\text{ask,-1}} \equiv$ change of the best ask price (in 0.0001 DEM) between the last and second-to-last observable events;

- $\Delta P_{\text{ask,-1}} = P_{\text{ask,-1}} - P_{\text{ask,-2}} \equiv$ change of the best ask price (in 0.0001 DEM) between the second-to-last and third-to-last observable events;

- $\Delta P_{\text{bid}} = P_{\text{bid}} - P_{\text{bid,-1}} \equiv$ change of the best bid price (in 0.0001 DEM) between the last and second-to-last observable events;

- $\Delta P_{\text{bid,-1}} = P_{\text{bid,-1}} - P_{\text{bid,-2}} \equiv$ change of the best bid price (in 0.0001 DEM) between the second-to-last and third-to-last observable events.

The price covariates characterize the short-term dynamics of the best bid and ask quotes and their interactions with the most recent transaction prices. These covariates accommodate short-term deviations of the quoted bid and ask quotes from the long-run equilibria as well as the potential errors that could be committed in the reconstruction of the trading history. For instance, the *Slippage* variable defined as the shift in the market price given by the midpoint of bid-ask spread relative to the last transaction price may be interpreted as the midquote positioning bias. One can think of it as a 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.[13] Since the absolute value of positioning bias is expected to be larger during the periods of changes in the bid and ask quotes without transaction activity, the *Slippage* variable can capture the tradeless price discovery mechanism that might prevail around the public news announcement. The size of the market bid-ask spread is often associated with the intuitive notion of illiquidity in the market microstructure literature, and is expected to have a strong impact on the types of submitted orders, as has been emphasized in the empirical microstructure literature.

4.3.2 Depth covariates

- $\log(Q_{\text{ask}}) \equiv$ natural logarithm of the market depth quoted on the ask side, i.e., the value of currency (expressed in \$ mln.) available at the best ask price;[14]

- $Q^{+}_{\text{ask}} \equiv$ indicator of large depth on the ask side, equals to unity if and only if the ask market depth is at least \$10 mln.;

[13] 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.

[14] During the time period before 1998, the exact value of market depth on the ask and bid sides in the Reuters trading system was unobservable to market participants when it was in double digits (\$10 mln. or larger). Traders could see only the "R" indicator in the depth part of the screen. Therefore, we set $Q_{\text{ask}} = 10$ and $Q_{\text{bid}} = 10$ every time when the actual market depth is at least 10 million US dollars.

- $\Delta \log(Q_{\text{ask}}) = \log(Q_{\text{ask}}) - \log(Q_{\text{ask,-1}}) \equiv$ last change of logarithm of the market depth quoted on the ask side if the best ask price did not change between the last and second-to-last observable events, zero otherwise;

- $\Delta \log(Q_{\text{ask,-1}}) = \log(Q_{\text{ask,-1}}) - \log(Q_{\text{ask,-2}}) \equiv$ second-to-last change of logarithm of the market depth quoted on the ask side if the best ask price did not change between the last and third-to-last observable events, zero otherwise;

- $\log(Q_{\text{bid}}) \equiv$ natural logarithm of the market depth quoted on the bid side, i.e., the value of currency (expressed in \$ mln.) available at the best bid price;

- $Q^{+}_{\text{bid}} \equiv$ indicator of large depth on the bid side, equals to unity if and only if the bid market depth is at least \$10 mln.;

- $\Delta \log(Q_{\text{bid}}) = \log(Q_{\text{bid}}) - \log(Q_{\text{bid,-1}}) \equiv$ second-to-last change of logarithm of the market depth quoted on the bid side if the best bid price did not change between the last and third-to-last observable events, zero otherwise;

- $\Delta \log(Q_{\text{bid,-1}}) = \log(Q_{\text{bid,-1}}) - \log(Q_{\text{bid,-2}}) \equiv$ second-to-last change of logarithm of the market depth quoted on the bid side if the best bid price did not change between the last and third-to-last observable events, zero otherwise.

The two market depth variables represent the second dimension of liquidity identified in the introduction to this paper, specifically, how many units of asset can be bought (or sold) at the current ask (or bid) market prices. The depth covariates are also expected to be significant for the risks of cancellations since the likelihood of a cancellation event is expected to be positively related to the total number of active limit orders, and the latter number is correlated with the quoted depth at the best market price. Similarly to the quoted prices, the changes of quoted quantities capture the more subtle traders' reaction to changes in the publicly available information on the limit order book and private information from the customer orders. The indicators of large depth Q^{+}_{ask} and Q^{+}_{bid} accommodate potential nonlinearities in the dependence of market activity on the depth variables (which might be partially justified by unobservable exact levels of depth when the market depth exceeds \$10 mln.)

4.3.3 Order flow and transaction level covariates

- $Side \equiv$ directional indicator of the last transaction ($+1$ for seller-initiated trades, -1 for buyer-initiated trades);

- $Side_{-1} \equiv$ directional indicator of the second-to-last transaction ($+1$ for seller-initiated trades, -1 for buyer-initiated trades);

- $F_{0-5''} \equiv$ the signed order flow (measured as the difference between the number of seller- and buyer-initiated transactions) in the five-second period prior to the last observable event;

- $F_{5-10''}$, $F_{10-15''}$, $F_{15-30''}$, $F_{30-60''}$, $F_{1-2'}$, $F_{2-5'}$ and $F_{5-15'}$ are similarly defined as the signed order flow over the time periods five to ten seconds, ten to 15 seconds, and so on, prior to the last observable event;

- $T_{0-5''} \equiv$ the trade (measured as the total number of transactions) in the five-second period prior to the last observable event;

- $T_{5-10''}$, $T_{10-15''}$, $T_{15-30''}$, $T_{30-60''}$, $T_{1-2'}$, $T_{2-5'}$ and $T_{5-15'}$ are similarly defined as the number of transactions in the electronic system in the periods five to ten seconds, ten to 15 seconds, and so on, prior to the last observable event.

Side, which is the indicator of aggressor in the most recent transaction, characterizes the asymmetry in the impact of *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 *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 *et al.* [22]) and on the stock markets (Hausman *et al.* [30], Lo *et al.* [37], Huang and Stoll [32]).

The additional activity and order flow covariates denoted, respectively, by $F_{(\tau_0;\tau_1]}$ and $T_{(\tau_0;\tau_1]}$, attempt to capture some of the lower-frequency serial dependence in the market dynamics. Linear combinations of such variables for various time lags τ_0 and τ_1 take the form of differences between the long and short-run moving averages of signed order flow and trading activity, they should incorporate the influence of common factors contributing to the unobserved heterogeneity that cannot be captured by the current state of the limit order book. Since the failure to account for the unobserved common factors may invalidate the conditional independence assumption which is one of the foundations of the competing risks framework, the variables representing the trade and activity history are chosen to capture a substantial part of the lower-frequency serial dependence, at the same time striking a balance between the correct specification and empirical tractability of the model.[15]

[15] An alternative lagged activity measure given by the amplitude of transaction price fluctuations in a given time interval 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 [14]) is currently under investigation. Results of that study will be reported in a separate paper.

5 Empirical Results

Tables 3, 4, and 5 report 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 events recorded during the first three days of the week, October 6–8, 1997, between 6 a.m. and 5 p.m. GMT, which are the most liquid trading hours in the Reuters D2000-2 trading system, have been used for the analysis.[16] Significance of the covariate coefficients is determined from the robust t-statistics (Lin and Wei [36]) at the 99% level, and the statistically insignificant coefficients are shown in small script. The qualitative effects of covariates for buy-side events are determined analogously. They closely mirror the covariate effects for sell-side events (Tables 6, 7, and 8).

The empirical regularities uncovered by the Cox proportional hazard regressions can be assigned to one of the three groups depending on whether they are associated with shifts in the price levels, changes of market depth, or changes in the signed order flow. Accordingly, they are discussed in subsections 5.1, 5.2, and 5.3 below. The location of coefficients in Tables 3–8 providing evidence in support of these empirical facts is indicated in parentheses.

5.1 Change of the Best Bid and Ask Quotes and Transaction Returns

1. (*The top three entries of column 2 in Tables 3 and 6.*) The hazard rates of aggressive limit order arrivals tend to be more sensitive than the hazard rates of market order arrivals to the shifts in midquote levels, even though the dependence of hazard rates is qualitatively similar for market orders and aggressive limit orders. This is consistent with the theoretical prediction of Foucault [16] and the empirical finding of Daníelsson and Payne [10]. In other words, the stronger upward price adjustment without trades implies smaller proportion of market orders in the seller-initiated flow of transactions and larger proportion of market orders in the buyer-initiated flow of transactions.

2. (*Column 2 in Tables 3 and 6.*) Negative signs of *Slippage* coefficients for the submission and cancellation rates of subsidiary sell limit orders and positive signs of *Slippage* coefficients for the submissions and cancellation rates of subsidiary buy limit orders support the hypothesis that the general level of subsidiary limit order submission and cancellation activity declines when the midpoint of bid-ask spread moves closer to the limit order and *this move is not accompanied by any transactions*. In this case, the previously submitted orders are less likely to be cancelled, and the new orders are less likely to arrive. The overall effect of the *Slippage* variable is less obvious when the last

[16] The last two trading days of the sample are reserved for out-of-sample evaluation of the forecasts.

transaction leads to asymmetric changes in the best quoted prices and the depth at these prices. In this case, the interpretation becomes complicated because of migration of submission and cancellation events between alternative classes of risk every time the reference quotes (P_{ask} or P_{bid}) are modified.

3. (*The top section of column 3 in Tables 3 and 6.*) There is a strong negative association between the transaction intensity and bid-ask spread, and a strong positive association between the rate of large price improvements and bid-ask spread. Also we found a fairly strong negative association between the subsidiary order arrival activity and the size of the bid-ask spread. On surface, this last result appears to contradict the empirical evidence from Biais *et al.* [4] and Hollifield *et al.* [31]. Notice, however, that our results are obtained for the spread sensitivity of *hazard rates* of subsidiary limit order arrivals, while the results in the previous literature were established for the spread sensitivity of *probability* that the next event will be subsidiary order arrival.

4. (*Column 3 in Tables 3 and 6.*) Even though the fresh supply of liquidity (in the form of subsidiary limit order arrivals) and fresh demand for liquidity (in the form of aggressive limit and market order arrivals) are negatively related to the size of bid-ask spread, there is also negative association between the cancellation rates of previously submitted subsidiary limit orders and the size of bid-ask spread. The effect of spread on the *net supply* of subsidiary liquidity is close to neutral. The only type of liquidity that is unambiguously positively associated with the size of bid-ask spread comes from large price improvements that occur much more often when the spread exceeds two ticks. Otherwise, the liquidity supply appears to be fairly steady and driven primarily by short-term price fluctuations and trends and changes in the depth of the limit order book.

5. (*The top three entries of column 5 in Tables 3 and 6.*) Sell market orders are discouraged by bid price improvement; buy market orders are discouraged by ask price improvement. This "contrarian" property of market orders is in contrast with the properties of aggressive limit orders that appear to be more frequent following price improvements on the opposite side of the book. However one should be aware of the possibility of mechanical misclassification of limit orders as being "aggressive" since a larger proportion of sell limit orders submitted at the same prices overlap with higher bid quotes, and a larger proportion of buy limit orders overlap with lower ask quotes, even though the quote changes might be transitory price swings without any information content.

6. (*The top five entries of column 4 in Tables 3 and 6.*) Arrivals of market sell orders and aggressive limit sell orders, as well as ask price improvements (but not the ask

depth improvements) are more likely to occur after recent ask quote deteriorations (increases). A similar regularity is observed on the buy side of the market with regard to recent bid quote deteriorations (decreases).

7. (*The mid-section of columns 4 and 5 in Tables 3 and 6.*) The arrival and cancellation rates of subsidiary limit orders positioned several ticks above the current ask and several ticks below current bid market prices are negatively affected by recent deteriorations of best quoted prices on the same side and recent improvements of best quoted prices on the opposite side of the limit order book. A similar negative reaction to recent quoted price improvements on the opposite side of the book is observed for the arrival rates of price and quantity improving orders. In summary, 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.

5.2 Depth of the Limit Order Book at the Best Bid and Ask Quotes

1. (*The top six entries of column 6 in Tables 3 and 6.*) The arrival rates for sell market orders and aggressive sell limit orders, as well as the rates of ask price and quantity improvements are higher when depth on the ask side of the limit order book is high. A similar effect is observed on the buy side of the limit order book. This points to the competition among aggressive limit order traders for time priority in trade execution as one of the driving forces behind the price improvement and transaction activities.

2. (*The seventh and eighth entries of column 6 in Tables 3 and 6.*) The arrival rates for subsidiary limit orders at the price levels next to the best market bid and ask quotes are negatively affected by large depth at the best market quotes. This points to the competition among less aggressive limit order traders as the abundant liquidity on the market discourages limit order submission at inferior prices.

3. (*The third and seventh entries of column 7 in Tables 3 and 6.*) Depth improvements at the best market quote become more likely after another depth improvement event at the same price market quote. This provides an evidence that liquidity tends to accumulate at the best market quotes once the market confirms these quotes as the new price levels. Depth improvements at the ask market price also encourage sell market order submission but do not affect the rates of ask price improvement. In fact, this mechanism is predominantly at work when the possibilities of price improvement have been already exhausted (and the market spread is one tick or smaller).[17]

[17] Additional Cox regressions for these events run separately for the periods of small and large spread

4. (*The top three entries of column 8 in Tables 3 and 6.*) A tentative evidence that buy market orders arrive more actively during the periods of intermediate depth (less than \$10 mln.) and less actively during the periods of large depth (at least \$10 mln. or more) at the best ask price was not found for the arrivals of aggressive buy limit orders. Similar effects are observed for sell market orders and aggressive sell limit orders. On the other hand, the arrival intensity of very aggressive limit orders (submitted at the prices above the best ask and below the best bid quotes) appears to decline as the level of depth on the opposite side of the limit order book increases. This can be easily justified from the observation that very aggressive limit orders are more likely to be used during the periods of scarce liquidity (low depth) at the best market bid and ask quotes.

5. (*Column 9 in Tables 3 and 6.*) There is strong evidence that bid and ask depth improvement events without price improvements encourage cancellations of subsidiary limit orders on the opposite side of the book and discourage liquidity demand coming in the form of aggressive limit and market orders from the opposite side of the market. Moreover, the deteriorating level of depth at the best market quotes encourages traders on the opposite side of the market to chase more aggressively the remaining liquidity at these quotes and avoid cancellations of subsidiary limit orders on the opposite side of the market.

6. (*The bottom section of column 9 in Tables 3 and 6.*) The limit order cancellation rates at the best bid and ask prices tend to decrease after increases in the observed depth on the opposite side of the limit order book. The increasing depth encourages limit order traders on the opposite side of the book to keep his order at the best market price only if he knows that he will be the first trader to receive this price.[18]

5.3 Lagged Order Flow and Transaction Activity

1. (*The top section of columns 2–4 in Tables 4 and 7.*) The arrival rates of sell market orders, aggressive sell limit orders, and market ask price improvements increase after seller-initiated transactions and decrease after buyer-initiated transactions. A similar effect is detected on the opposite side of the limit order book. This clustering of the buyer and seller pressure is consistent with the ample evidence of strong high-frequency

confirm this hypothesis.

[18] This is confirmed by additional Cox regressions (not shown in Tables 3 and 6) run separately for cancellation events leading to price deterioration and leading to depth deterioration without change of the best quoted price.

directional momentum observed on foreign exchange and stock markets and reported in previous studies.[19]

2. (*The bottom section of columns 2–5 in Tables 4 and 7.*) There is 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 detected for subsidiary buy limit orders. This can be generally interpreted as an evidence of limit order book updating by order anticipators receiving signal that the trading pressure is shifting market price in the opposite direction.

3. (*The mid-section of columns 5–9 in Tables 4 and 7.*) The prolonged periods of low submission rates for subsidiary ask limit orders are more likely to occur following the periods of massive seller-initiated transaction activity. However, only a weak evidence of similar effects is found for subsidiary bid limit orders. An explanation can be based on the common perception of the US dollar in the late 1990s as a currency with stronger fundamentals than the Deutsche Mark, which could be translated into the lower sensitivity of limit order bids to strong "Buy" signals. However, this apparent asymmetry may also be period-specific.

4. (*Tables 5 and 8.*) The model provides empirical support for the observation that order submission and order cancellation rates increase after the periods of high transaction activity. The effect of transaction activity is very persistent at all price levels in the limit order book. Interestingly, the persistence of activity appears to be stronger for the arrival rates of subsidiary limit orders relative to other types of events. This also points to subsidiary limit orders and, more generally, to stop-loss order execution activity (Osler [42]) as a possible transmission mechanism of shocks and the source of memory and fat tails in the foreign exchange returns.

5.4 Summary

The nature of competition among aggressive and non-aggressive traders can be summarized by the following digest of the main empirical implications of the continuous time model. Special attention is drawn to the role of price and depth improvements with and without trade for the price discovery process.

- The continuous time model estimated in this section implies that the buy market order activity is spurred by depth improvements due to higher competition of limit

[19] Lo *et al.* [37] report similar results for limit order data collected from stock markets. Hausman *et al.* [30] reports similar findings for transaction level data.

order traders for time priority on the bid side. On the other hand, the buy market order activity is deterred if depth improvements on the ask side are not accompanied by simultaneous price improvements (i.e., spread reductions).

- The buy market order activity is encouraged by bid quote improvements originating from the competing limit order buyers, and by ask quote improvements originating from the providers of liquidity on the sell side of the market. However, the effect of a bid quote improvement on the buy market order activity is captured almost entirely by the concurrent reduction of the bid-ask spread, and appears to be much smaller in magnitude.

- In the time period covered by our data, price improvements by competitors generally encouraged liquidity provision, while depth improvements on the opposite side made aggressive limit order and market order traders defer their transactions and consider taking different actions.

- Overall, the prevalence of buyer-initiated trades among the recent transactions encouraged the buy market order activity. Similar effects are observed on the opposite side of the market for seller-initiated trades.

6 Principal Components of Activity Indices

The main purpose of principal component analysis (PCA) in the present context is reduction of a large set of competing risk indices to a much smaller set of indices (factors) that still generate most of short-term market dynamics. Besides the data compression, PCA represents an important step toward automatic and efficient generation of short-term forecasts for the market activity. In retrospect, careful inspection of the market activity levels around the times of likely news arrivals reveals a typical market reaction pattern, which starts with a sudden and numerous withdrawals of orders on the "weaker" side of the market as it comes under pressure. The limit order withdrawals are accompanied by larger than usual values of bid-ask spread and strong directional trading volume driven primarily by market orders. This activity creates a considerable imbalance between the number of buy and sell orders in the limit order book, which is gradually restored after the prices adjust to the new level. The price adjustment occurs in a staggering wave-like fashion and is accompanied by higher than usual volatility of the price. No matter whether the news confirmed dealers' expectations about Bundesbank cutting German interest rates, the government crisis in Italy, or their worries about East Asian financial markets, as long as the pattern of traders' reaction to such information has some common characteristics, it may be captured by a small number of principal components that could potentially be much smaller than the number of market participants or the number of initially identified market components.

The version of PCA conducted in this section is based on the analysis of the sample variance-covariance matrix $\widehat{\mathbf{\Sigma}}$ of competing risk indices $y_r = \mathbf{z}'\boldsymbol{\beta}_r$, $r = 1, ..., R$. Matrix $\widehat{\mathbf{\Sigma}}$ is formed by R sample variances of risk indices

$$\widehat{\sigma}_r^2 = \tfrac{1}{N}\sum_{n=1}^{N}(y_{nr} - \overline{y}_r)^2, \quad r = 1, 2, ..., R,$$

as well as $R(R-1)/2$ sample covariances among the pairs of risk indices

$$\widehat{\sigma}_{rr'} = \tfrac{1}{N}\sum_{n=1}^{N}(y_{nr} - \overline{y}_r)(y_{nr'} - \overline{y}_{r'}), \quad r, r' = 1, 2, ..., R, \quad r \neq r',$$

and contains a large amount of information about the contemporaneous variations in the components of covariate vector $\mathbf{Z}$ characterizing the state of the limit order book. The data compression can be achieved since PCA approximates the parametric covariate-dependent competing risk indices $y_r = \mathbf{z}'\boldsymbol{\beta}_r$ by alternative risk indices $\widetilde{y}_r = \mathbf{f}'\boldsymbol{\gamma}_r$ that depend on a small set of Q common factors $\mathbf{f} = (f_1, ..., f_Q)'$. These factors are formed as simple linear combinations of the covariates with the coefficients chosen in such a way that the first factor f_1 explains the largest portion of the sample variance-covariance matrix[20] of the competing risk indices, the second factor f_2 explains the next largest portion, and so on. The factors constructed in such a manner and normalized to have unit variance are called the *principal components*, or *PCA risk factors*. Then the factor loadings of these observable factors (which are, by construction, the linear combinations of covariates) can be obtained by application of standard regression techniques.

6.1 Relative Contributions of PCA Factors

Since, mathematically, the principal components are normalized eigenvectors of the variance-covariance matrix $\widehat{\mathbf{\Sigma}}$, their number is equal to the number of competing risk indices R. As the last few eigenvectors of $\widehat{\mathbf{\Sigma}}$ point to the directions where the risk indices f_1, f_2, ..., f_R jointly exhibit little or no variation, most of the information content of the data is likely to be represented by some smaller number of PCA risk factors $f_1, ..., f_Q$, $Q < R$. In fact,

[20] Since the ultimate goal of analysis in this chapter is construction of short-term forecasts of trading activity, our main concern to the end of this chapter will be approximating the *short-run* variances and covariances of the competing risk indices. Therefore, the analysis throughout the rest of this chapter will be based primarily on the properties of short-run variances and covariances. It remains to be seen whether the similar data compression performed at the lower frequencies leads to useful forecasts.

the situation when $Q \ll R$, and the principal components $f_1, ..., f_Q$ result in a much smaller data set, also cannot be ruled out.

The importance of PCA risk factors $f_1,\ f_2,\ ...,\ f_Q$ across the set of competing risk indices is measured by the eigenvalues $\lambda_1,\ \lambda_2,\ ...,\ \lambda_Q$ of the extracted factors. In particular, the proportion of the total variance described by the first Q principal components is given by the ratio

$$\frac{\sum_{q=1}^{Q} \mathrm{Var}(f_q)}{\sum_{r=1}^{R} \mathrm{Var}(f_r)} = \frac{\mathrm{tr}(\mathrm{Var}(\mathbf{ff}'))}{\mathrm{tr}(\mathbf{\Sigma})} = \frac{\sum_{q=1}^{Q} \lambda_q}{\sum_{r=1}^{R} \lambda_r}.$$

The natural question arises as to how many principal components should be retained to capture systematic variation in the original data set and avoid capturing what is likely to be a random noise. Since there is no single universally accepted statistical approach in the statistical literature to the number of PCA factors to be retained, it would be reasonable to take an eclectic approach applying a spectrum of alternative criteria to this problem.

Figure 10 represents the so-called *scree plot* showing the eigenvalues corresponding to the first 13 principal components against the number of those components. Inspection of the graph suggests that the first five PCA factors capture the major portion of variation in the risk indices, even though the sixth and seventh factors might also be marginally important. Table 13 shows that the first seven PCA factors capture almost 90% of variation in the indices, whereas the first five PCA factors capture almost 86% of the variation. The "subjective" choice of $Q = 5$ appears to be reasonable in the present context.

There is a large number of "objective" decision rules frequently applied to decide on the number of "significant" principal components to be retained. Jolliffe [34], Chapter 6 gives a good survey of formal and informal approaches, while Jackson [33] investigates performance of alternative decision rules applied to some artificial and real data. Even though a large amount of research has been done on the rules for choosing the number of retained components, there is no universally accepted rule that is applied in the literature in all circumstances.

One of the popular statistical decision rules which is strongly favored in Jackson [33] can be derived from the so-called broken stick model (Frontier [18]). According to this model, if the total variance, represented by the sum of the eigenvalues of the variance-covariance matrix of indices, can be divided randomly among Q components, then the distribution of components follows a "broken-stick" distribution, with the expected kth largest eigenvalue calculated as

$$\lambda_k^* = \frac{1}{Q} \sum_{q=k}^{Q} \frac{1}{q},$$

when the number of components Q is large enough. One way of deciding whether the proportion of variance accounted by the kth PCA factor is sufficiently large for this component to be retained is to compare this proportion with λ_k^*. The test based on λ_k^* leads to the conclusion that the first four components should be retained. However, since there are no systematic results on the size and power properties of the "broken stick" rule based on the expected eigenvalues, and in view of frequently cited evidence that the rule retains too few components, we check for the number of components using a bootstrap procedure that recovers the entire distribution of the kth largest eigenvalue under the null of equal eigenvalues for the last $Q-k+1$ PCA factors. This alternative procedure leads to the conclusion that $Q=5$ components should be retained, with the fifth component being only marginally significant at the 95% confidence level.

Yet another, rather subjective procedure for selection of the number of retained components uses the method of log-eigenvalue (LEV) diagrams (Wilks [45], Chapter 9). The method is motivated by the idea that, if the last $R-Q$ principal components pick up random noise, then the magnitudes of their eigenvalues decrease exponentially with the component number. The Q retained PCA factors correspond to the log-eigenvalues deviating from the straight-line portion of the plot on the LEV diagram. Figure 11 shows that the LEV plot deviates from the linear pattern implied by exponential decay of the plot under the null hypothesis of random noise for the number of factors as low as five. Even though it cannot be unambiguously seen on the plot that $Q=5$, it appears to be a reasonable choice again. Therefore, the further analysis will be conducted for five PCA factors. However, to ensure robustness of our results, we repeated the analyses for seven retained PCA factors and obtained similar results.

6.2 Interpretation of PCA Factors

Table 9 gives the representations of covariates in the original Cox proportional hazard regressions for the competing risks in terms of the extracted PCA factors. Tables 10, 11, and 12 report the estimates $\widehat{\boldsymbol{\gamma}}_r$ of factor loadings on the first five PCA factors in the semiparametric Cox proportional hazard model for the competing risk indices $\widetilde{y}_r = \mathbf{f}'\widehat{\boldsymbol{\gamma}}_r$ of sell order arrivals, buy order arrivals, and cancellation activity in the limit order book, respectively. The estimation period is between 6 a.m. and 5 p.m. GMT on the week of October 6-8, 1997. The t-statistics for the estimates of Tables 10, 11, and 12 are shown in parentheses, and statistically significant factor loading estimates coefficients (at the 95% level) are marked by stars. To facilitate interpretation of the extracted PCA factors, we also show the off-diagonal and diagonal terms of the cross-correlogram (Figures 12 and 13, respectively) for the first five PCA factors.

The first principal component, which is obviously nonstationary, dominates the dynamics of risk indices. The first PCA factor captures approximately 44% of the total variation in the covariate indices of the competing risks. Since all types of risk have large positive factor loadings on the first factor, the first component can be readily interpreted as the general level of *limit order book activity.*

The second principal component contributes less than 16% to the total variation of risk indices dynamics. Inspection of its autocorrelogram reveals its stationarity, and its cross-correlogram with the first principal component shows that they interact only marginally at all leads and lags. The factor loadings on the second factor are uniformly positive for log-hazard rates of buyer-initiated events (associated with submission and cancellation of bid limit orders and buy market orders) and almost uniformly negative for the log-hazards of seller-initiated events. Therefore, we can interpret the second factor as the *short-term activity momentum*, which identifies the more active side of the limit order book (buy or sell) without differentiating across the types of activity (whether it is submission or cancellation of limit or market orders). Since the active order-driven market intrinsically represents the dynamic interaction of buyers and sellers, it should be no surprise that buyers' or sellers' actions cannot dominate the market for long periods of time. Indeed, the autocorrelation of this factor becomes indistinguishable from zero for lags as low as 50 time periods (epochs), where each epoch is assumed to be terminated by an observable limit order book event and lasts about one second on average during the hours of liquid trading.

The third principal component capturing 12.4% of variation in the hazard rate dynamics has slowly decaying autocorrelogram and therefore has long memory. The rate of decay of its autocorrelogram on Figure 13, characterizing persistence of the third PCA factor, is close to the rate of decay of the autocorrelogram for the first factor. Similarly to the second principal component, the third PCA factor is uncorrelated to the first factor at all leads and lags but interacts rather non-trivially with the second principal component. Since the factor loadings of the third component are negative for submissions of non-aggressive bids and aggressive seller-initiated orders, and positive for submissions of non-aggressive ask orders and aggressive buyer-initiated orders, we can think of the third factor as the "*buying pressure*" on the market. The cancellation activity pattern that can be uncovered from the factor loadings of cancellation risks generally countervails those for the second PCA factor (short-term activity momentum) and conforms to the intuitive notion of "buying pressure." Aggressive buyers who tend to cancel buy limit orders several ticks below the market bid price, keep limit orders just below the market bid price in the hope of price reversal, and do just the opposite on the ask side of the market. Interestingly, since the cancellation activity patterns at the best market bid and ask quotes are similar to the submission patterns at the more competitive prices (leading to price improvement and narrowing the spread), this also conforms to the activity of aggressive sellers who might be testing the market before resubmitting their orders if their earlier offers have not been hit promptly.

The analysis of cross-correlations between factors 2 and 3 is conducted using the augmented graph of autocorrelogram highlighting the interaction between factors 2 and 3 (Figure 14). The graph shows that the unusually low activity on the ask-side of the limit order book relative to the bid side (small values of factor 2) precede the aggressive buyer pressure (high values of factor 3), leading to moderately higher ask-side activity in the short run that eventually reverts to persistently lower ask-side activity in the long run. This observation reveals the non-trivial interaction between the two factors, which implies a richer story than most theoretical microstructure models can tell.

Since the interpretation of the consecutive principal components becomes increasingly difficult, we make an attempt to interpret only the fourth and fifth components that contribute, respectively, to 9% and 4.7% of variation in the parametric parts of the log-hazard rates of competing risks. The loadings on factor 4 appear symmetric for buyer- and seller-initiated events, and are positive for limit order submissions within the spread that do not cause immediate transactions, as well as for limit order at least five ticks away from the prevailing market bid-ask spread. The loadings on factor 4 are negative for the arrivals of market and limit orders causing transactions, as well as for submissions and cancellations of subsidiary orders in the vicinity of the market bid-ask spread. Therefore, high values of factor 4 can be associated with the tendency to quote more competitive prices within the spread, while high values of factor 4 are associated with the tendency to take limit orders on the opposite side of the book without much bargaining. Therefore, low values of factor 4 can be associated with "choppy" markets when trades tend to occur without much bargaining, which often happens at the high levels of the market spread and might be associated with the relatively high adverse selection component (Harris [26]). We may attach the term "*adverse selection*" to the fourth PCA factor, and stick to the terminology in the future.

Note that the "adverse selection" factor does not appear to interact much with any of the other major principal components except the first one, which provides an illustration of the frequently reported phenomenon that the general level of trading activity is higher when some information is present in the market (i.e., the "adverse selection" component is high). Indeed, active markets lead to higher competition among traders who tend to submit more quotes before making a deal. The inverse causality also appears to be at work. Trading at the relatively large levels of bid-ask spread might be a signal of "choppy market," at least for some, presumably uninformed, traders, causing adjustments of their quotes until the market returns to the "smoother" state.

Finally, we make an attempt to interpret the fifth PCA factor. The loadings on this factor are positive for all aggressive buy limit order arrivals and cancellations, for submissions of buy limit orders well below the bid market price, for submissions of sell limit orders just above the ask market price, and for all cancellations of subsidiary limit orders on the ask side. In all other cases, the loadings on factor 5 are negative and their signs and magnitudes are roughly symmetric to the signs and magnitudes of this factor loadings for similar risks on the opposite side of the book. Even though, in many respects, factor 5 behaves similarly to factor 3, large values of factor 5 imply that the aggressive buyer activity tends to be accompanied by cancellations of subsidiary sell limit orders rather than arrivals of subsidiary sell limit orders, in contrast to factor 3. Also, factor 5 does not appear to be as persistent as factor 3 (Figures 13 and 14), and generally is more sensitive to the changes of depth (Table 9). All these properties might be interpreted as a tentative evidence of directional information on the market that signals a permanent shift of the market price in the near future. For this reason we will call the fifth PCA factor the "*bull market momentum*".

In a small forecasting exercise of the next section, the second, third, and fifth principal components come out as the major determinants of directional activity in the limit order book, whereas the first and fourth principal components affect only the absolute level of activity in the limit order book.

7 Evaluation of Probability Forecasts

In this section, we generate probability forecasts for the next limit order book event based on the formulas shown in Appendix B. Then we apply the method of *reliability diagrams* (Wilks [45], Chapter 7) to measure goodness-of-fit of the five-factor principal component competing risks model (section 6) calibrated to the data collected over the liquid trading period (6 a.m. to 5 p.m. GMT) on October 6–8, 1997, and evaluate the out-of-sample predictive performance of this model in the liquid trading period on October 9–10, 1997. This exercise is especially interesting since the out-of-sample period in our data set is dominated by highly volatile trading following the Bundesbank announcement about an increase in the repo rate around 11:30 a.m. GMT on Thursday, October 9, 1997. The volatility and spread over a large portion of the out-of-sample period remained much higher than they were during the period October 6–8, 1997. The ability of the competing risks model calibrated on historical data to generate credible out-of-sample probability forecasts will be a clear indicator of its promise in improvement of short-term probability forecasts of market events and in their applications to the analysis of alternative scenarios in real time.

The probability estimates (11) and (12) in Appendix B form the backbone of our forecasting results. We set $t_0 = 0$ in the formulas (11) and (12), since these probabilities are evaluated immediately after the limit order book gets updated. The forecasts constructed by these formulas can be made dynamic in the sense that they take into account the passage of time as the duration t_0 of uneventful period is steadily growing from zero to the moment of epoch termination.

7.1 Classification of Forecasted Events

In the present paper, we concentrate on the directional forecasts for future transactions. Direction of future transaction is an important issue for limit order traders and their sponsors who are intrinsically interested in fast execution of their limit orders at favorable prices. The prevalence of buyer- or seller-initiated transactions on the market may be intimately, but non-trivially related to the appreciation or depreciation of exchange rates as it might signal the informational advantage of the counterparties initiating the trades. Investigation of dynamic links between the order flow, appreciation or depreciation of transaction price, and various measures of limit order book liquidity is a dominant topic in the modern empirical microstructure literature and one of the main objectives of the present research. Therefore it would be natural to verify the ability of the model to issue the warnings about unusually high or low probabilities of transactions on the sell and buy side of the limit order book.

First, we evaluate the ability of the PCA factor model with five principal components (section 6), based on the estimates of risk indices $\widetilde{y}_r = \mathbf{f}'\boldsymbol{\gamma}_r$ reported in Tables 10, 11, and 12, to provide a good in-sample fit between the realized and forecast probabilities of buyer- and seller-initiated transactions. As we have some freedom to select the number of event categories, we choose a coarser event classification scheme than the one with $S = 14$ observable risks used in the original model estimated in section 5. Even though we started with $S = 14$ observable risks (marked by single and double stars in Table 1), their number was later reduced by pooling the events of types A1, A2, A3 in the "Sell Trade" category AA and the events of types B1, B2, B3 in the "Buy Trade" category BB, and by collapsing the events of types A4 and A5 into the "Ask Price Improvement" category AP+ and the events of types B4 and B5 in to the "Bid Price Improvement" category BP+. The composition of the event classes A6, B6, AC6, and BC6 corresponding, respectively, to "Ask Depth Improvement", "Bid Depth Improvement", "Ask Touch Cancellation", and "Bid Touch Cancellation" events were unchanged but the categories were renamed as AD+, BD+, A−, and B−, respectively. The modified event classification scheme with $S = 8$ observable categories is shown in Table 14.

Figure 15 gives an example of dynamic evolution of forecast probabilities for buyer- and seller-initiated transactions. The sample period shown on the graph is chosen to be identical to the one used to demonstrate the evolution of bid, ask, and transaction prices on Figure 9. The probability forecasts are based on formulas (9) and (10) from Appendix B and evaluated under the assumption that no events occurred at least for one second after the previous observable limit order book event, which explicitly takes into account the reaction time of the potential forecast user. The first seven minutes of the sample period captured on Figures 9 and 15 was the period of heavy buyer pressure accompanied by a rapid growth of market price by almost ten tick points. This was also the period when our competing risks model produces the probability forecasts which are much larger for buy than for sell transactions. The last portion of the sample covers the period when the price stabilizes just under the new level DEM 1.7575 per US dollar and is supported by a fairly strong "resistance" on the sell side. The sell and buy transaction probability forecasts for this period are approximately equal to each other on average, even though the fraction of epochs terminated by seller-initiated transactions is slightly higher than predicted in this subperiod.

7.2 Reliability Plots for In-Sample and Out-of-Sample Forecasts

The quality of probability forecasts based on formulas (9) and (10) from Appendix B is evaluated using the method of reliability diagrams (Wilks [45], Chapter 7). The graphs on Figure 16 show the reliability plots for the forecasts of buyer- and seller-initiated transactions made under assumption that no events occurred at least for one second after the previous observable limit order book event. The first two graphs show the reliability plots for the probability forecasts of buyer-initiated transactions (graph 1) and seller-initiated transactions (graph 2) matched with the frequencies of buyer- and seller-initiated transactions observed over the estimation period of the first three days (31391 epochs). The graphs in the lower portion of Figure 16 show the reliability plots for the probability forecasts of buyer- and seller-initiated trades (graphs 3 and 4, respectively) when the forecast probabilities are matched with the corresponding transaction frequencies observed in the out-of-sample period covering the last two days (19385 epochs). Both sell and buy event forecasts offer high resolution with a broad range of covered probabilities. The forecasts also have good reliability properties as they do not reveal a strong tendency to deviate systematically from the main diagonal line that corresponds to perfect reliability. Even though the probability forecasts slightly underpredict the probabilities of buy and sell trades when the predicted probabilities are very small, this tendency towards overconfidence (underprediction of unlikely seller- and buyer-initiated transactions) is fairly weak. The overconfidence (underprediction) bias for the rare events is slightly stronger when the forecasts are evaluated out-of-sample, which can be easily detected from the deviations of left tails of reliability plots on the last two graphs from the perfect reliability line. This might be an issue of concern if the forecasts are ultimately used to measure the risk of transactions that might occur in the undesirable direction. However, even with this small caveat that must be taken into consideration by the ultimate users of forecast, the out-of-sample performance of the competing risks model turns out to be surprisingly good.

The high degree of persistence detected on the diagrams for cross-correlograms of PCA factors (Figures 16 and 17) suggests that at least some of the factors are unlikely to change much over a relatively short time period that might be covering more than one epoch. Therefore, it may be interesting to compare the quality of forecasting rules that rely on the principal components derived from the competing risks model and the performance of some benchmark forecasting rules based on the current and lagged directional indicators. In our last exercise, we evaluate the quality of forecasts for the event that the next transaction in the limit order book will be initiated by buyer or seller within the next 30 seconds, or no trade will be recorded in the next 30 seconds since the time of forecast. The "naïve" benchmark used in our comparisons will be based on the trinomial logit regression of buy, sell, or no-trade indicator with the covariates given by the signs of ten most recent transactions (± 1 if transaction was initiated by seller/buyer, and zero if no transaction occurred). The only information used to predict the direction of next transaction in this simple "directional momentum" model is the direction of the last several trades. The performance of this simple forecasting model will be compared with the multinomial logit model with the index of covariates given by the first five PCA factors.[21]

The two diagrams on the left-hand side of Figure 17 show the reliability plots for the forecast of the event that the next limit order book transaction will be buyer-initiated and occur in 30 seconds after the time of forecast. Similarly, the two diagrams on the right-hand side of Figure 17 show the reliability plots for probability forecasts of the event that the next transaction in limit order book will be seller-initiated and occur in 30 seconds since the time of forecast. The reliability plots in the upper portion of Figure 17 are based on the trinomial logit regression with the covariates given by the five competing risks PCA factors. The reliability plots in the bottom portion of Figure 17 are based on the benchmark forecasting model based on the trinomial logit regression with the covariates given by signs of last 10 transactions (± 1 for transactions initiated by seller/buyer). All diagrams are produced from the sample covering the estimation period 6 am to 5 pm GMT on October 6–8, 1997.

Since the plots on the upper two diagrams on Figure 17 match very closely the main diagonal, we can conclude that the PCA factor model fits the data much better than the "naïve" benchmark based on the directional indicators. Moreover, the PCA factor model also has better discrimination properties as the range of probability forecasts based on this model is substantially wider. Apart from the slight downward bias of probability forecasts for small probabilities, the PCA factor model appears to provide a better fit to the empirical data in comparison to the alternative model.

Now we check whether the PCA factor model is capable of giving the warnings about unusually high (or unusually low) probabilities that the next limit order book transaction will be buyer- or seller-initiated. Figure 18 is similar to Figure 17, except that it shows the reliability plots of forecasts matched to the data from the out-of-sample period 6 am to 5 pm GMT on October 9 and October 10, 1997. Again, the PCA factor model delivers the forecasts that are more reliable and have much better resolution properties than the forecasts based on the alternative "momentum-based" model. Figure 18 also indicates a problem that seems to deteriorate with the forecast horizon.[22] In particular, the forecasts based on our model appear to be overconfident in the sense that the model underpredicts low probability events (left tails on the reliability plots are bent upward) and overpredicts

[21] Since the efficient algorithm generating multistep forecasts in the competing risks framework is currently unavailable, and would require, in particular, the dynamic multistep forecasts of the covariate structure or the factor structure, we come up with a shortcut solution that ignores the dynamic properties of covariates but keeps intact the general structure of the model. The development of a truely dynamic forecasting model is left for future investigation.

[22] The reliability plots of one-minute-ahead probability forecasts (not reported here) have similar properties, except for the more substantial biases of these plots for high and low forecast probabilities.

high probability events (right tails on the reliability plots are bent downward), which may also serve as an evidence of rapid reaction of market participants to extremely high or low probability signals. As the quality of information contained in the PCA components quickly deteriorates as the forecast horizon increases, continuous monitoring of the relevant market information and updating the risk indices are crucial conditions for the success in this highly competitive segment of foreign exchange market.

8 Conclusion

The main contributions of this paper can be summarized as follows. First, the unified competing risks framework is proposed for the analysis of high-frequency heterogeneous trading activity in financial markets organized as limit order books. The model has an attractive behavioral interpretation and can be applied to any irregularly spaced observations with heterogeneous attributes which are frequently encountered in empirical finance and economics. The asymptotic theory developed by Andersen *et al.* [3] for counting processes can be adopted with minimal adjustments to conduct the inference for competing risks. The competing risks model of this paper represents a semiparametric alternative to the fully parametric model of limit order trading by Bisière and Kamionka [5]. Our model is more general, since it does not impose unjustified restrictions on the form of the baseline hazard rates, which can lead to potential biases in the estimated covariate effects and incorrect inferences. It also serves as a flexible hazard-based alternative to the multinomial logit model of a limit order market (Ellul *et al.* [13]).

The model is applied to analyze the timing and interaction between quotes and trades in the Reuters D2000-2 electronic brokerage system. Despite its rapid development in the 1990s, this segment of the foreign exchange market has been rarely studied before. The major stylized facts about the dynamics of 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 applied to a variety of high- and low-frequency financial data. One natural application of the competing risks technique involves the empirical analysis of thinly or irregularly traded financial instruments, corporate bonds, and emerging market securities. In combination with conventional asset 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.

Appendix

A. Cox Proportional Hazard Model for Competing Risks

Consider N independent random vectors of latent durations $(T_{n1}, ..., T_{nR})$, $n = 1, ..., N$, and the associated hazard functions $h_1(t|\mathbf{z}), ..., h_R(t|\mathbf{z})$. The random variables $T_{n1}, ..., T_{nR}$ are assumed to be conditionally independent given the current covariates $\mathbf{z}_n$. Consider N realizations of multivariate single-jump counting processes

$$\mathbf{N}_n(t) = (N_{n1}(t), ..., N_{nR}(t)), \quad n = 1, ..., N,$$

with

$$N_{nr}(t) = \mathbf{1}\{T_{nr} = \min_{r'} T_{nr'} \text{ and } T_{nr} \leq t\}.$$

Every individual counting process N_{nr} satisfies the multiplicative intensity model

$$\lambda_{nr}(t|\mathbf{z}) = Y_{nr} h_r(t|\mathbf{z}), \quad r = 1, ..., R;\ n = 1, ..., N,$$

where Y_{nr} is an observable indicator that contains information whether or not the market in period n is at risk of experiencing an event of type r.

The pdf of duration T conditional on the next event being of type r and the current covariate vector (partially determined by the previous event as explained above) being equal to $\mathbf{z}$, is

$$\begin{aligned} p_r(t|\mathbf{z}) &= h_r(t|\mathbf{z}) S(t-|\mathbf{z}) = h_r(t|\mathbf{z}) \prod_{r'=1}^{R} S_{r'}(t-|\mathbf{z}) \\ &= h_r(t|\mathbf{z}) \cdot \exp\left[-\int_0^t \sum_{r'=1}^{R} h_{r'}(u|\mathbf{z}) du \right]. \end{aligned} \tag{1}$$

On the other hand, by the definition of survivor function,

$$S(t-|\mathbf{z}) = \frac{f(t|\mathbf{z})}{h(t|\mathbf{z})},$$

where $h(t|\mathbf{z})$ and $p(t|\mathbf{z})$ are the conditional hazard function and pdf of duration t, whatever is the type of event associated with it. Therefore formula (1) can be rewritten as follows

$$p_r(t|\mathbf{z}) = \frac{h_r(t|\mathbf{z})}{h(t|\mathbf{z})} p(t|\mathbf{z}) = \pi_r(t|\mathbf{z}) p(t|\mathbf{z}),$$

where $\pi_r(t|\mathbf{z})$ is the probability that an event of type r occurs exactly t units of time since the last event was observed, conditional on covariates $\mathbf{z}$ and given the information that some event happens at time t at all. Thus, the probability density function of duration between the previous observable event in transient state $\mathbf{z}$ and the next event (not necessarily observable) is

$$p(t|\mathbf{z}) = \sum_{r=1}^{R} h_r(t|\mathbf{z}) \exp\left[-\int_0^t \sum_{r'=1}^{R} h_{r'}(u|\mathbf{z})du\right],$$

and the associated hazard rate is

$$h(t|\mathbf{z}) = \sum_{r=1}^{R} h_{r'}(t|\mathbf{z}).$$

The hazard rates of these risks are modelled using the Cox proportional hazard (CPH) specification

$$h_r(t|\mathbf{z}) = h_{0r}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_r), \quad r = 1, ..., R, \tag{2}$$

where $h_{0r}(t) = h_r(t|\mathbf{0})$ is the baseline hazard function, t is the time at risk, and $\mathbf{z}$ is the covariate vector. Covariates in proportional hazard models always act multiplicatively on the hazard rate of the specified type of event. The model (2) can be rewritten in the integrated form as follows

$$H_r(t|\mathbf{z}) = \int_0^t h_r(u|\mathbf{z})du = \int_0^t h_{0r}(u)du\exp(\mathbf{z}'\boldsymbol{\beta}_r) = H_{0r}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_r),$$

where $H_r(t|\mathbf{z})$ is the cumulative hazard function of risk r. The third equivalent representation of the model (2) is based on the expression for the survival function

$$S_r(t|\mathbf{z}) = \exp(-H_r(t|\mathbf{z})) = \exp(-H_{0r}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_r)) = (S_{0r}(t))^{\exp(\mathbf{z}'\boldsymbol{\beta}_r)},$$

where $S_{0r}(t) = S_r(t|\mathbf{0}) = \exp(-H_{0r}(t))$ is the baseline survivor function of risk r.To validate the Singh–Maddala parametric form of duration dependence for the baseline hazard functions (Bisière and Kamionka [5]), this dependence can be estimated nonparametrically. The notional hazard rate of type r as a function of time since inception (i.e., the time measured since the last publicly observed event) is assumed to be of the CPH form

$$h_r(t|\mathbf{z};\boldsymbol{\theta}_r) = h_{0r}(t)\exp(\mathbf{z}'\boldsymbol{\beta}_r)$$

with the unknown parameter $\boldsymbol{\theta}_r = (\boldsymbol{\beta}_r', h_{0r}(\cdot))'$, and the dependence of notional hazards on time being determined by functions $h_{0r}(\cdot)$ which are permitted to vary arbitrarily over the R types of risk.

Assuming there are N distinct arrival and cancellation events during the trading day, the likelihood function is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \mathcal{L}(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_R) = \prod_{n=1}^{N} p_{r_n}(t_n | \mathbf{z}_n; \boldsymbol{\theta}) \\
&= \prod_{n=1}^{N} h_{r_n}(t_n | \mathbf{z}_n; \boldsymbol{\theta}_r) S(t_n - | \mathbf{z}_n; \boldsymbol{\theta}) \\
&= \prod_{r=1}^{R} \prod_{n=1}^{N} h_r(t_n | \mathbf{z}_n; \boldsymbol{\theta}_r)^{\delta_{nr}} S_r(t_n | \mathbf{z}_n; \boldsymbol{\theta}_r) \\
&= \prod_{r=1}^{R} \mathcal{L}_r(\boldsymbol{\theta}_r),
\end{aligned}$$

where

$$\mathcal{L}_r(\boldsymbol{\theta}_r) = \prod_{n=1}^{N} h_r(t_n | \mathbf{z}_n; \boldsymbol{\theta}_r)^{\delta_{nr}} S_r(t_n | \mathbf{z}_n; \boldsymbol{\theta}_r),$$

t_n is the duration measured from the last observable event s_n associated with the present covariate vector $\mathbf{z}_n = (\mathbf{x}_n', \mathbf{d}_n')'$, and

$$\delta_{nr} = \begin{cases} 1 & \text{if the } n\text{th event is of type } r, \\ 0 & \text{otherwise.} \end{cases}$$

The maximum likelihood estimates of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_R)$ are obtained by independent maximization of functions $\mathcal{L}_r(\boldsymbol{\theta}_r)$ with respect to $\boldsymbol{\theta}_r$.

Let $t_{(1)r} < ... < t_{(N_r)r}$ denote the N_r distinct ordered durations of type r ($r = 1, ..., R$), and let $\mathbf{z}_{(j)r}$ characterize the covariates for the observed duration $t_{(j)r}$. The partial likelihood function

$$\mathcal{L}_{\text{part}}(\boldsymbol{\beta}) = \mathcal{L}_{\text{part}}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_R) = \prod_{r=1}^{R} \prod_{j=1}^{N_r} \left(\frac{\exp(\mathbf{z}_{(j)r}' \boldsymbol{\beta}_r)}{\sum_{n \in \mathcal{R}(t_{(j)r})} \exp(\mathbf{z}_n' \boldsymbol{\beta}_r)} \right)$$

depends on a finite dimensional parameter $\boldsymbol{\beta}$, and the risk set $\mathcal{R}(t_{(j)r})$ is defined as the set of observed durations that are equal to or larger than $t_{(j)r}$. Insertion of maximum partial likelihood estimators $\widehat{\boldsymbol{\beta}}_1, ..., \widehat{\boldsymbol{\beta}}_R$ into the expressions for hazard functions yields the Breslow estimators of cumulative hazard functions

$$\widehat{H}_{0r}(u, \widehat{\boldsymbol{\beta}}_r) = \sum_{j:\ t_{(j)r} \leq t} \left[\sum_{n \in \mathcal{R}(t_{(j)r})} \exp(\mathbf{z}_n' \widehat{\boldsymbol{\beta}}_r) \right]^{-m_{jr}}, \tag{3}$$

where m_{jr} is the number of tied durations of type r at $t_{(j)r}$. The estimates of survivor functions $S_r(t|\mathbf{z};\widehat{\boldsymbol{\beta}}_r)$ are obtained exactly as shown above for the case of a single risk.

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

$$P_r(0,t;\mathbf{Z}_0) = \int_0^t P_0(0,u;\mathbf{Z}_0)\exp(\mathbf{Z}_0'\boldsymbol{\beta}_r)dH_{0r}(u), \quad r = 1,...,R, \tag{4}$$

and can be estimated by

$$\widehat{P}_r(0,t;\mathbf{Z}_0) = \int_0^t \widehat{P}_0(0,u;\mathbf{Z}_0)\exp(\mathbf{Z}_0'\widehat{\boldsymbol{\beta}}_r)d\widehat{H}_{0r}(u,\widehat{\boldsymbol{\beta}}_r), \tag{5}$$

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

$$\begin{aligned} P_0(0,t;\mathbf{Z}_0) &= \prod_{u\in[0;t)}\left(1-\sum_{r=1}^{R}\exp(\mathbf{Z}_0'\boldsymbol{\beta}_r)dH_{0r}(u)\right) \\ &= \exp\left[-\sum_{r=1}^{R}\exp(\mathbf{Z}_0'\boldsymbol{\beta}_r)H_{0r}(t)\right], \end{aligned}$$

and its estimator

$$\widehat{P}_0(0,t;\mathbf{Z}_0) = \exp\left[-\sum_{r=1}^{R}\exp(\mathbf{Z}_0'\widehat{\boldsymbol{\beta}}_r)\widehat{H}_{0r}(t,\widehat{\boldsymbol{\beta}}_r)\right].$$

B. Forecasting the Probability of Next Event

Our probability forecasts will be based on the expressions (4) and (5) with the version of kernel estimator of the baseline hazard function (Appendix A). We also maintain the regularity assumption

$$\lim_{\tau\to+\infty}\int_0^\tau h_{0r}(u)du = \infty, \quad \text{for some } r = 1,2,...,R, \tag{6}$$

which guarantees that the probability of no event in interval $[0;\tau]$ converges to zero as the period without arrivals of new observations expands. Then t_0 seconds after the arrival of a new observable event the updated risk index $Y_{nr} = \mathbf{z}'_n \boldsymbol{\beta}_r$ for the risk of type r feeds into the expression of the hazard rate to yield

$$H_r(t; y_{nr}) = H_{0r}(t) \exp(y_{nr})$$

and into expression (4) for the incidence rate to obtain the conditional probability that event of type r will be next to occur less than t seconds after the previous observable event provided that the observation period has started t_0 seconds after the previous observable event[23]

$$P_r(t_0, t; y_{n\cdot}) = \int_{t_0}^{t} P_0(t_0, u; y_{n\cdot}) \exp(y_{nr}) dH_{0r}(u), \quad r = 1, ..., R, \tag{7}$$

where

$$\begin{aligned} P_0(t_0, t; y_{n\cdot}) &= \prod_{u \in [t_0; t)} \left(1 - \sum_{r=1}^{R} \exp(y_{nr}) dH_{0r}(u) \right) \\ &= \exp\left[-\sum_{r=1}^{R} (H_{0r}(t) - H_{0r}(t_0)) \exp(y_{nr}) \right], \end{aligned}$$

is the conditional probability that no event occurs over the period $[t_0; t)$ after the previous observable event given that no event has occurred over the period $[0; t_0)$ following the previous observable event. The conditional probabilities (7) are estimated by the incidence rates (4), using the formula

$$\widehat{P}_r(t_0, t; \widehat{y}_{n\cdot}) = \int_{t_0}^{t} \widehat{P}_0(t_0, u; \widehat{y}_{n\cdot}) \exp(\widehat{y}_{nr}) d\widehat{H}_{0r}(u, \widehat{y}_{nr}), \quad r = 1, ..., R, \tag{8}$$

where

$$\widehat{P}_0(t_0, t; \widehat{y}_{n\cdot}) = \exp\left[-\sum_{r=1}^{R} (\widehat{H}_{0r}(t, \widehat{y}_{nr}) - \widehat{H}_{0r}(t_0, \widehat{y}_{nr})) \exp(\widehat{y}_{nr}) \right],$$

[23] A more parsimonious alternative characterization of the risk indices could be obtained in terms of the PCA factors $\mathbf{u}_n = (u_1, u_2, ..., u_Q)'$ (section 6), since

$$Y_{nr} = \mathbf{u}'_n \boldsymbol{\gamma}_r + \widetilde{v}_{nr} = \widetilde{Y}_{nr} + \widetilde{v}_{nr}, \quad r = 1, ..., R.$$

In this case the set of baseline cumulative hazard functions $H_{0r}(u)$ need to be recalculated to incorporate the portion $\widetilde{v}_{nr}$ of the covariate index $Y_{nr} = \widetilde{Y}_{nr} + \widetilde{v}_{nr}$ assigned to the noise component. The conditional probability formulas look similar, except that the indices $\widetilde{Y}_{nr}$ and baseline hazard functions $\widetilde{H}_{0r}(u)$ are captured by the more parsimonious factor structure.

$\widehat{H}_{0r}(\cdot, \widehat{y}_{nr})$ can be, for example, the Breslow estimator (3), and $\widehat{y}_{n\cdot} = (\widehat{y}_{nr})_{r=1}^{R}$ is the estimated vector of risk indices. In particular, if only the subset of observable types of events $r = 1, ..., S$ $(S < R)$ and the epochs terminated less than τ seconds $(0 < \tau \leq \infty)$ after the last observable event are considered in the analysis, then the probability evaluated $t_0 < \tau$ seconds after the last event that the given epoch will be terminated with an event of an observable type $r = 1, ..., S$ is given by the ratio

$$\pi_r(t_0, \tau; y_{n\cdot}) = \frac{P_r(t_0, \tau; y_{n\cdot})}{\sum_{s=1}^{S} P_s(t_0, \tau; y_{n\cdot})}. \tag{9}$$

Finally, after substitution of the expression for the incidence rates $P_s(t_0, u; y_{n\cdot})$ into the integral (7) and some manipulations, we obtain

$$\begin{aligned} P_r(t_0, \tau; y_{n\cdot}) &= \int_{t_0}^{\tau} \exp\left[y_{nr} - \sum_{s=1}^{S} \exp(y_{ns}) H_{0s}(u) \right] dH_{0r}(u) \\ &= \int_{t_0}^{\tau} \frac{\exp(y_{nr}) h_{0r}(u)}{\sum_{s=1}^{S} \exp(y_{ns}) h_{0s}(u)} \exp(-H(u, y_{n\cdot})) dH(u, y_{n\cdot}) \qquad (10) \\ &= \int_{H(t_0, y_{n\cdot})}^{H(\tau, y_{n\cdot})} \frac{\exp(y_{nr}) \widetilde{h}_{0r}(\theta)}{\sum_{s=1}^{S} \exp(y_{ns}) \widetilde{h}_{0s}(\theta)} \exp(-\theta) d\theta \\ &= \int_{\exp(-H(\tau, y_{n\cdot}))}^{\exp(-H(t_0, y_{n\cdot}))} \frac{\exp(y_{nr}) \widetilde{h}_{0r}(-\log(v))}{\sum_{s=1}^{S} \exp(y_{ns}) \widetilde{h}_{0s}(-\log(v))} dv, \end{aligned}$$

where

$$H(u, y_{n\cdot}) = \sum_{s=1}^{S} \exp(y_{ns}) H_{0s}(u)$$

is the cumulative hazard function for the termination risk (triggered by an arrival of any observable event $s = 1, ..., S$), and $\widetilde{h}_{0r}(\cdot) = h_{0r}(H^{-1}(\cdot, y_{n\cdot}))$ is the baseline hazard rate of risk r for all $r = 1, ..., S$, expressed in the units of "intrinsic time" $H(\cdot, y_{n\cdot})$. Provided that the last integral in (10) is well approximated by the logistic functional form

$$\frac{A_r(t_0, \tau) \exp(y_{nr})}{\sum_{s=1}^{S} A_s(t_0, \tau) \exp(y_{ns})},$$

then the odds ratio (9) takes the form of conventional logistic function (adjusted for the passage of time) and forecasts can be based on the conventional multinomial logit estimates. On the other hand, if the main interest is to obtain the forecasts of instantaneous relative risks exactly t_0 seconds after the last observable event, one should directly substitute the covariate and baseline hazard estimates into the ratio

$$\frac{\widehat{h}_{0r}(t_0)\exp(\widehat{y}_{nr})}{\sum_{s=1}^{S}\widehat{h}_{0s}(t_0)\exp(\widehat{y}_{ns})}.$$

More generally, the probability forecasts can be based on the sample analogues

$$\widehat{P}_r(t_0,\tau;\widehat{y}_{n\cdot}) = \int_{t_0}^{\tau}\widehat{h}_{0r}(u)\exp(\widehat{y}_{nr}-\widehat{H}(u,\widehat{y}_{n\cdot}))du, \quad r=1,...,S, \tag{11}$$

$$\widehat{\pi}_r(t_0,\tau;\widehat{y}_{n\cdot}) = \frac{\widehat{P}_r(t_0,\tau;\widehat{y}_{n\cdot})}{\sum_{s=1}^{S}\widehat{P}_s(t_0,\tau;\widehat{y}_{n\cdot})} \tag{12}$$

for the expression (10) and formula (9).

References

[1] Admati, A. and P. Pfleiderer, "A Theory of Intraday Patterns: Volume and Price Variability." *Review of Financial Studies*, **1** (1988), 3–40.

[2] Admati, A. and P. Pfleiderer, "Divide and Conquer: A Theory of Intraday and Day-of-the-Week Mean Effects." *Review of Financial Studies*, **2** (1989), 189–224.

[3] Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Kieding, *Statistical Models Based on Counting Processes.* Springer-Verlag, New York (1993).

[4] Biais, B., P. Hillion, and C. Spatt, "An Empirical Analysis of the Limit-Order Book and the Order Flow in the Paris Bourse." *Journal of Finance*, **50** (1995), 1655–89.

[5] Bisière, C. and T. Kamionka, "Timing of Orders, Order Aggressiveness and the Order Book in the Paris Bourse." *Working Paper*, University of Toulouse (1999).

[6] Brown, P., N. Thomson, and D. Walsh, "Characteristics of the Order Flow Through an Electronic Open Limit Order Book." *Journal of International Financial Markets, Institutions and Money*, **9** (1999), 335–357.

[7] Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, *The Econometrics of Financial Markets.* Princeton University Press, Princeton, NJ (1997).

[8] Chakravarty, S. and C. Holden, "An Integrated Model of Market and Limit Orders." *Journal of Financial Intermediation*, **4** (1995), 213–241.

[9] Dacorogna, M.M., R. Gençay, U.A. Müller, R.B. Olsen, and O.V. Pictet, *An Introduction to High-Frequency Finance*. Academic Press, San Diego, CA (2001).

[10] Daníelsson, J. and R. Payne, "Measuring and Explaining Liquidity on an Electronic Limit Order Book: Evidence from Reuters D2000-2." *Working Paper*, Financial Markets Group, London School of Economics (2001).

[11] Diamond, D.W. and R.E. Verrecchia, "Information Aggregation in a Noisy Rational Expectation Model." *Journal of Financial Economics*, **16** (1987), 277–311.

[12] Easley, D. and M. O'Hara, "Time and the Process of Security Price Adjustment." *Journal of Finance*, **47** (1992), 577–606.

[13] Ellul, A., C. Holden, P. Jain, and R. Jennings, "Determinants of Order Choice on the New York Stock Exchange." *Working Paper*, Kelley School of Business, Indiana University (2003).

[14] Engle, R.F. and J.R. Russell, "Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the Autoregressive Conditional Duration Model." *Journal of Empirical Finance*, **4** (1997), 187–212.

[15] Evans, M.D.D. and R.K. Lyons, "Order Flow and Exchange Rate Dynamics." *NBER Working Paper* #7317, National Bureau of Economic Research, Cambridge, MA (1999).

[16] Foucault, T., "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market." *Journal of Financial Markets*, **2** (1999), 99–134.

[17] Foucault, T., O. Kadan, and E. Kandel, "Limit Order Book as a Market for Liquidity." *Working Paper*.

[18] Frontier, S., "Étude de la decroissance des valeurs propres dans une analyze en composantes principales: comparison avec le modèle de baton brisé." *Journal of Experimental Marine Biology and Ecology*, **25** (1976), 67–75.

[19] Geman, H. and T. Ané, "Stochastic Subordination." *Risk*, **9** (1996), 145–149.

[20] Glosten, L., "Is the Electronic Order Book Inevitable?" *Journal of Finance*, **49** (1994), 1127–61.

[21] Goodhart, C.A.E. and M. O'Hara, "High Frequency Data in Financial Markets: Issues and Applications." *Journal of Empirical Finance*, **4** (1997), 73–114.

[22] Goodhart, C., T. Ito, and R. Payne, "One Day in June 1993: A Study of the Working of the Reuters 2000-2 Electronic Foreign Exchange Trading System." In Frankel, J., G. Galli, and A. Giovannini (Eds.), *The Microstructure of Foreign Exchange Markets*. The University of Chicago Press (1996), 107–179.

[23] Griffiths, M.D., B.F. Smith, D.A.S. Turnbull, and R.W. White, "The Costs and Determinants of Order Aggressiveness." *Journal of Financial Economics*, **56** (2000), 65–88.

[24] Hamao, Y. and J. Hasbrouck, "Securities Trading in the Absence of Dealers: Trades and Quotes on the Tokyo Stock Exchange." *Review of Financial Studies*, **8** (1995), 849–878.

[25] Handa, P. and R.A. Schwartz, "Limit Order Trading." *Journal of Finance*, **51** (1996), 1835–61.

[26] Harris, L., *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, UK (2003).

[27] Harris, L. and J. Hasbrouck, "Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy." *Journal of Financial and Quantitative Analysis*, **31** (1996), 213–231.

[28] Hasbrouck, J., "Stalking the 'Efficient Price' in Market Microstructure Specifications: An Overview." *Journal of Financial Markets*, **5** (2002), 329–339.

[29] Hasbrouck, J. and G. Saar, "Limit Orders and Volatility in a Hybrid Market: The Island ECN." *Working Paper*, New York University.

[30] Hausman, J.A., A.W. Lo, and A.C. MacKinlay, "An Ordered Probit Analysis of Transaction Stock Prices." *Journal of Financial Economics*, **31** (1992), 319–379.

[31] Hollifield, B., R.A. Miller, and P. Sandås, "An Empirical Analysis of Limit Order Markets." *Working Paper*, Carnegie Mellon University, Pittsburgh, PA (1999).

[32] Huang, R.D. and H.R. Stoll, "The Components of the Bid-Ask Spread: A General Approach." *Review of Financial Studies*, **10** (1997), 995–1034.

[33] Jackson, D.A., "Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches." *Ecology*, **74** (1993), 2201–14.

[34] Jolliffe, I.T., *Principal Component Analysis* (2nd Ed.) Springer-Verlag, New York (2002).

[35] Jones, C., G. Kaul, and M. Lipson, "Transactions, Volume and Volatility." *Review of Financial Studies*, **7** (1994), 631–651.

[36] Lin, D. Y. and L. J. Wei, "The Robust Inference for the Cox Proportional Hazards Model." *Journal of American Statistical Association*, **84** (1989), 1074–78.

[37] Lo, A.W., A.C. MacKinlay, and J. Zhang, "Econometric Models of Limit-Order Executions." *Journal of Financial Economics*, **65** (2002), 31–71.

[38] Lyons, R., *The Microstructure Approach to Exchange Rates*. MIT Press, Cambridge, MA (2001).

[39] Lyons, R.K., "New Perspective on FX Markets: Order-Flow Analysis." *International Finance*, **4** (2001), 303–320.

[40] Marsh, T.A. and K. Rock, "The Transaction Process and Rational Stock Price Dynamics." *Working Paper*, University of California at Berkeley (1986).

[41] O'Hara, M., *Market Microstructure Theory*. Blackwell Publishers, Cambridge, U.K. (1995).

[42] Osler, C.L., "Currency Orders and Exchange Rate Dynamics: Explaining the Success of Technical Analysis." *Working Paper*, Federal Reserve Bank of New York (2001).

[43] Osler, C.L., "Stop-Loss Orders and Price Cascades in Currency Markets." *Staff Report* #150, Federal Reserve Bank of New York (2002).

[44] Parlour, C.A., "Price Dynamics in Limit Order Markets." *Review of Financial Studies*, **11** (1998), 789–816.

[45] Wilks, D.S., *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, CA (1995).

Tables and Graphs

Table 1: Classification of arrival and cancellation events on the sell side of limit order book

Risk (r)	Limit order price P^*	Price change	Quantity change
A1**	$P^* < P_{\text{bid}}$	$P_{\text{bid}} \searrow P_{\text{ask}} \searrow$	$Q_{\text{bid}} \downarrow\uparrow Q_{\text{ask}} \downarrow\uparrow$
A2**	Market sell order	$P_{\text{bid}} \searrow P_{\text{ask}}$ same	$Q_{\text{bid}} \downarrow$ if $\Delta P_{\text{bid}} = 0$
A3**	$P^* = P_{\text{bid}}$	$P_{\text{bid}} \searrow P_{\text{ask}} \searrow$	$Q_{\text{bid}} \downarrow\uparrow Q_{\text{ask}} \downarrow\uparrow$
A4*	$P^* - P_{\text{ask}} < -1$	$P_{\text{ask}} \downarrow$	$Q_{\text{ask}} \rightarrow Q^*$(lim.order size)
A5*	$P^* - P_{\text{ask}} = 1$	$P_{\text{ask}} \downarrow$	$Q_{\text{ask}} \rightarrow Q^*$(lim.order size)
A6*	$P^* - P_{\text{ask}} = 0$	No price effect	$Q_{\text{ask}} \uparrow$
A7	$P^* - P_{\text{ask}} = 1$	No price effect	No quantity effect
A8, A9, A10, and A11 defined similarly			
A12	$5 < P^* - P_{\text{ask}} \leq 10$	No price effect	No quantity effect
A13	$10 < P^* - P_{\text{ask}} \leq 20$	No price effect	No quantity effect
A14	$P^* - P_{\text{ask}} > 20$	No price effect	No quantity effect
AC6*	$P^* - P_{\text{ask}} = 0$	$P_{\text{ask}} \uparrow$	$Q_{\text{ask}} \downarrow$ if $\Delta P_{\text{ask}} = 0$
AC7	$P^* - P_{\text{ask}} = 1$	No price effect	No quantity effect
AC8, AC9, AC10, and AC11 defined similarly			
AC12	$5 < P^* - P_{\text{ask}} \leq 10$	No price effect	No quantity effect
AC13	$10 < P^* - P_{\text{ask}} \leq 20$	No price effect	No quantity effect
AC14	$P^* - P_{\text{ask}} > 20$	No price effect	No quantity effect

Table 2: Covariates characterizing the state of limit order book and recent trading history

Covariate	Description
Slippage	Current midquote minus last transaction price
Lt.Return	Last minus 2nd-to-last transaction price
$\text{Spread}_{>0}$	Best ask minus best bid quote, or zero, whatever is larger
$\text{Spread}^2_{>0}$	(Best ask minus best bid quote)^2, or zero, whatever is larger
ΔP_{ask}	Change of ask quote between last and 2nd-to-last events
$\Delta P_{\text{ask},-1}$	Change of ask quote between 2nd- and 3rd-to-last events
ΔP_{bid}	Change of bid quote between last and 2nd-to-last events
$\Delta P_{\text{bid},-1}$	Change of bid quote between 2nd- and 3rd-to-last events
$\log(Q_{\text{ask}})$	Log depth at best ask quote, or log \$10M, whatever is less
Q^+_{ask}	1 if log depth at ask equals \$10M, zero otherwise
$\Delta\log(Q_{\text{ask}})$	Change of $\log(Q_{\text{ask}})$ if $\Delta P_{\text{ask}}=0$, zero otherwise
$\Delta\log(Q_{\text{ask},-1})$	Change of $\log(Q_{\text{ask},-1})$ if $\Delta P_{\text{ask}}=0\&\Delta P_{\text{ask},-1}=0$, zero otherwise
$\log(Q_{\text{bid}})$	Log depth at best bid quote, or log \$10M, whatever is less
Q^+_{bid}	1 if log depth at bid equals \$10M, zero otherwise
$\Delta\log(Q_{\text{bid}})$	Change of $\log(Q_{\text{bid}})$ if $\Delta P_{\text{bid}}=0$, zero otherwise
$\Delta\log(Q_{\text{bid},-1})$	Change of $\log(Q_{\text{bid},-1})$ if $\Delta P_{\text{bid}}=0\&\Delta P_{\text{bid},-1}=0$, zero otherwise
Side	1 if last trade seller-initiated, -1 if buyer-initiated
Side_{-1}	1 if 2nd-to-last trade seller-initiated, -1 if buyer-initiated
$F_{\text{0-5}''}$	Signed number of trades 0 to 5 sec. prior to last event
$F_{\text{5-10}''}$, $F_{\text{10-15}''}$, $F_{\text{15-30}''}$, $F_{\text{30-60}''}$, $F_{\text{1-2}'}$, $F_{\text{2-5}'}$, $F_{\text{5-15}'}$ defined similarly	
$T_{\text{0-5}''}$	Number of trades 0 to 5 sec. prior to last event
$T_{\text{5-10}''}$, $T_{\text{10-15}''}$, $T_{\text{15-30}''}$, $T_{\text{30-60}''}$, $T_{\text{1-2}'}$, $T_{\text{2-5}'}$, $T_{\text{5-15}'}$ defined similarly	

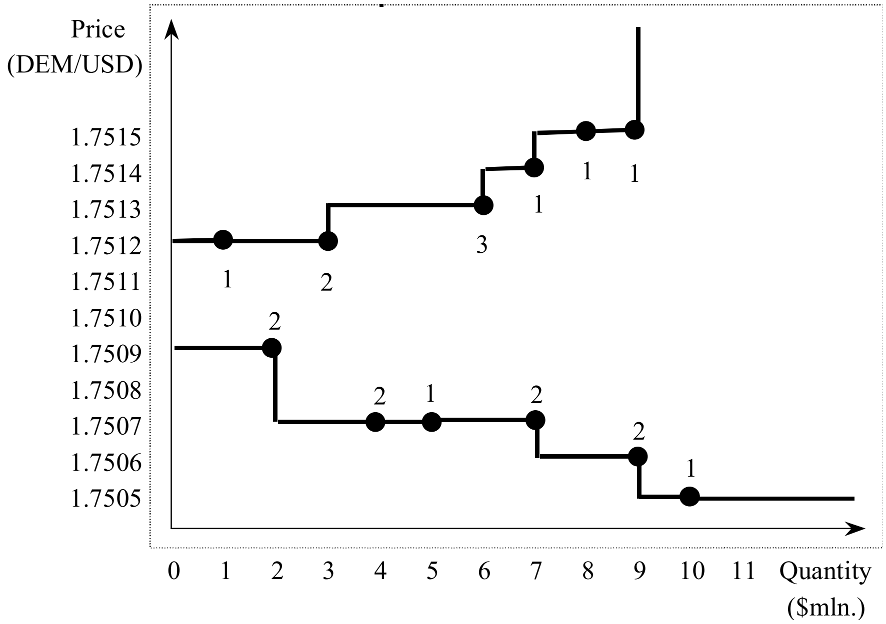

Figure 1:

Example of supply and demand curves from Reuters D2000-2 dealing system

Figure 1 displays the state of the Reuters D2000-2 electronic limit order book at a particular moment in time. The ersatz supply and demand curves on 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 limit 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.

Table 3: Estimated price and quantity coefficients for competing risks of seller-initiated events

Risk type (r) (1)	Slippage (2)	Spread (3)	ΔP_{ask} (4)	ΔP_{bid} (5)	Q_{ask} (6)	ΔQ_{ask} (7)	Q_{bid} (8)	ΔQ_{bid} (9)
Sell $< P_{\text{bid}}$	0.103	−0.754	0.437	0.172	−0.006	−0.479	−0.436	−0.381
Market sell	0.067	−0.525	0.135	−0.069	0.174	0.361	0.068	−1.161
Sell $= P_{\text{bid}}$	0.081	−0.670	0.222	0.091	0.140	0.041	0.005	−1.146
Sell $< P_{\text{ask}}-1$	0.062	0.451	0.089	−0.059	0.120	−0.013	0.053	−0.255
Sell at $P_{\text{ask}}-1$	0.017	0.029	0.048	−0.101	0.302	−0.235	0.005	−0.253
Sell at P_{ask}	−0.105	0.017	−0.051	−0.124	0.122	0.798	0.017	−0.100
Sell at $P_{\text{ask}}+1$	−0.176	−0.187	−0.130	−0.159	−0.101	0.060	0.058	0.025
Sell at $P_{\text{ask}}+2$	−0.126	−0.162	−0.142	−0.136	−0.080	0.081	0.125	0.003
Sell at $P_{\text{ask}}+3$	−0.138	−0.037	−0.120	−0.113	−0.026	0.037	0.080	−0.005
Sell at $P_{\text{ask}}+4$	−0.139	−0.153	−0.115	−0.012	−0.043	0.067	0.042	−0.009
Sell at $P_{\text{ask}}+5$	−0.208	−0.091	0.007	0.042	0.103	0.004	−0.045	0.051
Sell $\leq P_{\text{ask}}+10$	−0.135	−0.101	−0.080	−0.048	−0.081	0.013	0.025	−0.000
Sell $\leq P_{\text{ask}}+20$	−0.027	0.011	−0.098	−0.043	0.042	−0.028	0.009	−0.003
Cancel at P_{ask}	−0.149	−0.013	−0.014	−0.052	0.600	0.205	0.017	−0.228
Canc.at $P_{\text{ask}}+1$	−0.257	−0.128	0.039	−0.096	−0.046	−0.040	0.014	0.101
Canc.at $P_{\text{ask}}+2$	−0.299	−0.112	−0.102	0.050	0.029	−0.072	0.020	0.092
Canc.at $P_{\text{ask}}+3$	−0.158	−0.116	−0.078	−0.152	0.007	−0.058	−0.030	0.152
Canc.at $P_{\text{ask}}+4$	−0.212	−0.152	−0.153	0.021	−0.067	−0.015	0.015	0.127
Canc.at $P_{\text{ask}}+5$	−0.167	−0.159	−0.037	0.013	0.191	−0.170	0.029	0.145
Canc.$\leq P_{\text{ask}}+10$	−0.153	−0.016	−0.009	0.005	0.009	−0.047	0.048	0.091
Canc.$\leq P_{\text{ask}}+20$	−0.036	−0.088	−0.037	−0.070	−0.050	−0.007	−0.028	0.121

Table 4: Estimated lagged signed order flow coefficients for competing risks of seller-initiated events

Risk type (r) (1)	Side (2)	Side-1 (3)	$F_{0\text{-}5''}$ (4)	$F_{5\text{-}10''}$ (5)	$F_{10\text{-}15''}$ (6)	$F_{15\text{-}30''}$ (7)	$F_{30\text{-}60''}$ (8)	$F_{1\text{-}2'}$ (9)
Sell $< P_{\text{bid}}$	0.311	0.288	−0.017	0.018	−0.028	−0.003	0.008	0.002
Market sell	0.319	0.170	0.044	0.013	−0.003	−0.004	0.002	−0.001
Sell $= P_{\text{bid}}$	0.238	0.104	0.029	−0.001	−0.024	−0.006	0.003	0.000
Sell $< P_{\text{ask}}-1$	0.216	0.075	0.018	−0.008	0.004	0.005	0.004	−0.000
Sell at $P_{\text{ask}}-1$	0.105	0.047	0.028	−0.019	−0.014	−0.009	0.000	−0.001
Sell at P_{ask}	0.022	−0.017	0.034	−0.013	−0.017	−0.008	−0.005	−0.004
Sell at $P_{\text{ask}}+1$	−0.021	−0.018	0.021	−0.022	−0.014	−0.010	−0.005	−0.006
Sell at $P_{\text{ask}}+2$	0.024	−0.017	−0.001	−0.022	−0.023	−0.008	−0.008	−0.002
Sell at $P_{\text{ask}}+3$	−0.057	−0.027	0.011	−0.034	−0.024	−0.009	−0.007	−0.005
Sell at $P_{\text{ask}}+4$	−0.085	0.035	0.001	−0.028	−0.034	−0.017	−0.008	−0.005
Sell at $P_{\text{ask}}+5$	0.011	−0.008	−0.001	−0.007	−0.030	−0.016	−0.006	−0.008
Sell $\leq P_{\text{ask}}+10$	−0.006	0.010	0.007	−0.025	−0.026	−0.013	−0.008	−0.005
Sell $\leq P_{\text{ask}}+20$	−0.066	−0.049	0.001	−0.033	−0.025	−0.021	−0.009	−0.010
Cancel at P_{ask}	0.042	0.085	0.014	0.022	0.007	−0.007	−0.002	−0.003
Canc.at $P_{\text{ask}}+1$	0.112	0.083	0.030	0.032	−0.009	−0.012	−0.003	−0.003
Canc.at $P_{\text{ask}}+2$	0.095	0.167	0.025	0.040	−0.005	−0.007	−0.004	−0.004
Canc.at $P_{\text{ask}}+3$	0.054	0.175	0.025	0.043	0.015	−0.002	−0.002	−0.007
Canc.at $P_{\text{ask}}+4$	0.138	0.097	0.039	0.042	0.015	−0.003	−0.005	−0.008
Canc.at $P_{\text{ask}}+5$	0.138	0.144	0.026	0.018	0.028	−0.008	0.001	−0.003
Canc.$\leq P_{\text{ask}}+10$	0.150	0.159	0.023	0.036	0.020	0.010	0.005	−0.001
Canc.$\leq P_{\text{ask}}+20$	0.076	0.041	0.020	0.005	0.028	0.013	0.013	0.007

Table 5: Estimated lagged trading activity coefficients for competing risks of seller-initiated events

Risk type (r) (1)	$T_{0\text{-}5''}$ (2)	$T_{5\text{-}10''}$ (3)	$T_{10\text{-}15''}$ (4)	$T_{15\text{-}30''}$ (5)	$T_{30\text{-}60''}$ (6)	$T_{1\text{-}2'}$ (7)	$T_{2\text{-}5'}$ (8)	$T_{5\text{-}15'}$ (9)
Sell $< P_{\text{bid}}$	0.069	0.009	0.037	0.004	0.004	0.006	−0.001	0.001
Market sell	0.019	0.008	0.018	0.006	0.004	0.003	0.001	0.001
Sell $= P_{\text{bid}}$	0.043	0.025	0.014	0.007	0.005	0.002	0.001	0.001
Sell $< P_{\text{ask}}-1$	0.037	0.011	0.008	−0.005	0.002	0.003	0.002	0.001
Sell at $P_{\text{ask}}-1$	0.014	0.022	0.008	0.001	0.007	0.003	0.002	0.001
Sell at P_{ask}	0.019	0.027	0.022	0.009	0.008	0.003	0.002	0.001
Sell at $P_{\text{ask}}+1$	0.020	0.034	0.016	0.011	0.007	0.003	0.002	0.001
Sell at $P_{\text{ask}}+2$	0.041	0.028	0.029	0.017	0.009	0.000	0.002	0.001
Sell at $P_{\text{ask}}+3$	0.024	0.034	0.024	0.022	0.012	0.005	0.001	0.001
Sell at $P_{\text{ask}}+4$	0.021	0.041	0.011	0.014	0.010	0.004	0.002	0.002
Sell at $P_{\text{ask}}+5$	0.035	0.033	0.033	0.033	0.006	0.005	0.001	0.001
Sell $\leq P_{\text{ask}}+10$	−0.002	0.039	0.032	0.026	0.013	0.007	0.001	0.001
Sell $\leq P_{\text{ask}}+20$	0.004	0.017	0.028	0.015	0.013	0.005	0.004	0.000
Cancel at P_{ask}	0.044	0.008	0.002	0.001	0.002	−0.000	0.000	0.001
Canc.at $P_{\text{ask}}+1$	0.027	0.003	0.005	0.004	0.005	0.005	0.002	0.001
Canc.at $P_{\text{ask}}+2$	0.036	0.011	0.018	0.008	0.006	0.004	0.002	0.001
Canc.at $P_{\text{ask}}+3$	0.035	0.010	0.015	0.015	0.005	−0.001	0.003	0.001
Canc.at $P_{\text{ask}}+4$	0.019	0.014	0.009	0.016	0.002	0.002	0.003	0.001
Canc.at $P_{\text{ask}}+5$	0.041	0.042	0.020	0.012	0.008	−0.001	0.003	0.000
Canc.$\leq P_{\text{ask}}+10$	0.031	0.018	0.005	0.008	0.004	0.003	0.001	0.002
Canc.$\leq P_{\text{ask}}+20$	0.029	0.021	0.016	0.012	0.002	−0.004	0.002	0.002

Table 6: Estimated price and quantity coefficients for competing risks of buyer-initiated events

Risk type (r) (1)	Slippage (2)	Spread (3)	ΔP_{bid} (4)	ΔP_{ask} (5)	Q_{bid} (6)	ΔQ_{bid} (7)	Q_{ask} (8)	ΔQ_{ask} (9)
Buy $> P_{\text{ask}}$	−0.269	−0.687	−0.224	−0.212	0.107	−0.518	−0.543	−0.503
Market buy	−0.132	−0.568	−0.133	0.087	0.194	0.552	0.053	−1.102
Buy $= P_{\text{ask}}$	−0.221	−0.760	−0.168	−0.050	0.199	−0.132	−0.032	−1.136
Buy $> P_{\text{bid}}+1$	−0.110	0.468	−0.102	0.037	0.175	−0.359	−0.002	−0.328
Buy at $P_{\text{bid}}+1$	−0.023	0.071	−0.097	0.110	0.234	−0.265	−0.003	−0.090
Buy at P_{bid}	0.047	0.061	0.066	0.080	0.113	0.709	0.030	−0.179
Buy at $P_{\text{bid}}-1$	0.066	−0.177	0.178	0.168	−0.092	0.062	0.069	0.017
Buy at $P_{\text{bid}}-2$	0.053	−0.161	0.211	0.135	−0.040	0.052	0.005	0.045
Buy at $P_{\text{bid}}-3$	0.053	−0.126	0.179	0.035	−0.054	0.071	−0.012	0.045
Buy at $P_{\text{bid}}-4$	0.058	−0.119	0.119	0.116	−0.094	0.068	0.085	0.013
Buy at $P_{\text{bid}}-5$	0.020	0.079	0.158	0.002	0.096	−0.043	0.002	0.028
Buy $\geq P_{\text{bid}}-10$	0.055	0.002	0.079	0.144	0.042	−0.009	−0.036	0.038
Buy $\geq P_{\text{bid}}-20$	0.043	−0.002	0.018	0.055	0.014	−0.036	0.021	0.008
Cancel at P_{bid}	0.077	−0.019	−0.023	0.035	0.600	0.111	0.003	−0.199
Canc.at $P_{\text{bid}}-1$	0.179	−0.106	−0.018	0.059	−0.043	−0.026	0.009	0.092
Canc.at $P_{\text{bid}}-2$	0.286	−0.134	0.066	0.029	−0.031	−0.072	−0.011	0.128
Canc.at $P_{\text{bid}}-3$	0.156	−0.088	0.118	0.047	−0.054	−0.085	−0.002	0.143
Canc.at $P_{\text{bid}}-4$	0.175	−0.057	0.140	0.014	0.042	−0.098	−0.131	0.182
Canc.at $P_{\text{bid}}-5$	0.110	0.219	0.139	−0.039	0.097	−0.132	−0.100	0.182
Canc.$\geq P_{\text{bid}}-10$	0.045	−0.033	0.072	0.036	0.048	−0.108	0.013	0.143
Canc.$\geq P_{\text{bid}}-20$	0.022	0.021	0.017	0.079	−0.060	−0.023	0.030	0.067

Table 7: Estimated lagged signed order flow coefficients for competing risks of buyer-initiated events

Risk type (r) (1)	Side (2)	$\text{Side}_{\text{-1}}$ (3)	$F_{\text{0-5''}}$ (4)	$F_{\text{5-10''}}$ (5)	$F_{\text{10-15''}}$ (6)	$F_{\text{15-30''}}$ (7)	$F_{\text{30-60''}}$ (8)	$F_{\text{1-2'}}$ (9)
Buy $> P_{\text{ask}}$	−0.042	−0.167	−0.041	−0.016	0.000	−0.010	−0.005	−0.003
Market buy	−0.246	−0.125	−0.053	−0.020	0.001	0.001	−0.003	−0.001
Buy $= P_{\text{ask}}$	−0.158	−0.103	−0.031	−0.002	0.009	0.004	−0.004	−0.003
Buy $> P_{\text{bid}}+1$	−0.157	0.040	−0.018	−0.001	0.011	−0.000	−0.007	−0.007
Buy at $P_{\text{bid}}+1$	−0.138	0.027	−0.028	0.010	0.021	0.007	−0.002	−0.001
Buy at P_{bid}	0.013	0.012	−0.025	0.012	0.021	0.009	−0.000	0.001
Buy at $P_{\text{bid}}-1$	0.056	0.030	−0.008	0.012	0.028	0.012	0.002	−0.001
Buy at $P_{\text{bid}}-2$	0.075	0.054	−0.004	0.018	0.020	0.003	0.003	0.000
Buy at $P_{\text{bid}}-3$	0.070	0.044	−0.007	0.017	0.024	0.003	0.002	0.001
Buy at $P_{\text{bid}}-4$	0.090	0.034	−0.006	0.015	0.019	0.007	0.004	−0.001
Buy at $P_{\text{bid}}-5$	0.041	0.096	0.002	0.026	0.009	0.001	0.006	−0.000
Buy $\geq P_{\text{bid}}-10$	0.086	0.060	0.001	0.015	0.020	0.005	0.009	0.004
Buy $\geq P_{\text{bid}}-20$	0.054	0.039	−0.020	0.011	0.017	0.013	0.003	0.006
Cancel at P_{bid}	−0.070	−0.069	−0.015	−0.025	−0.002	0.004	−0.002	0.002
Canc.at $P_{\text{bid}}-1$	−0.119	−0.023	−0.031	−0.029	−0.005	0.017	0.002	−0.001
Canc.at $P_{\text{bid}}-2$	−0.045	−0.132	−0.028	−0.024	0.005	0.008	0.000	−0.001
Canc.at $P_{\text{bid}}-3$	−0.102	−0.090	−0.049	−0.016	−0.017	0.007	0.005	0.001
Canc.at $P_{\text{bid}}-4$	−0.164	−0.120	−0.013	−0.026	−0.014	−0.000	0.000	0.002
Canc.at $P_{\text{bid}}-5$	−0.027	−0.144	−0.046	−0.051	−0.003	−0.001	0.001	−0.001
Canc.$\geq P_{\text{bid}}-10$	−0.098	−0.080	−0.046	−0.037	−0.032	−0.009	0.001	0.002
Canc.$\geq P_{\text{bid}}-20$	−0.073	−0.105	−0.011	−0.030	−0.035	−0.012	−0.012	−0.004

Table 8: Estimated lagged trading activity coefficients for competing risks of buyer-initiated events

Risk type (r) (1)	$T_{0\text{-}5''}$ (2)	$T_{5\text{-}10''}$ (3)	$T_{10\text{-}15''}$ (4)	$T_{15\text{-}30''}$ (5)	$T_{30\text{-}60''}$ (6)	$T_{1\text{-}2'}$ (7)	$T_{2\text{-}5'}$ (8)	$T_{5\text{-}15'}$ (9)
Buy $> P_{\text{ask}}$	0.075	0.022	0.036	0.015	0.006	0.001	0.003	0.000
Market buy	0.030	0.008	0.018	0.012	0.007	0.002	0.001	0.001
Buy $= P_{\text{ask}}$	0.050	0.017	0.017	0.009	0.010	0.001	0.001	0.000
Buy $> P_{\text{bid}}+1$	0.051	0.017	0.005	0.009	0.003	0.004	0.000	0.001
Buy at $P_{\text{bid}}+1$	0.017	0.014	−0.002	0.002	0.006	0.004	0.001	0.001
Buy at P_{bid}	0.025	0.017	0.017	0.007	0.006	0.002	0.001	0.001
Buy at $P_{\text{bid}}-1$	0.012	0.008	0.014	0.009	0.006	0.004	0.001	0.001
Buy at $P_{\text{bid}}-2$	0.026	0.023	0.030	0.015	0.006	−0.000	0.001	0.001
Buy at $P_{\text{bid}}-3$	0.031	0.014	0.012	0.019	0.008	−0.001	0.001	0.001
Buy at $P_{\text{bid}}-4$	0.012	0.008	0.020	0.018	0.009	0.004	0.002	0.001
Buy at $P_{\text{bid}}-5$	0.028	0.013	0.037	0.024	0.011	−0.004	0.002	0.001
Buy $\geq P_{\text{bid}}-10$	0.006	0.018	0.032	0.018	0.007	0.002	0.001	0.001
Buy $\geq P_{\text{bid}}-20$	0.014	−0.002	0.018	0.010	0.011	0.004	0.001	0.001
Cancel at P_{bid}	0.049	0.003	0.008	0.001	0.004	−0.001	0.000	0.000
Canc.at $P_{\text{bid}}-1$	0.031	−0.006	0.007	0.002	0.004	0.004	0.001	0.001
Canc.at $P_{\text{bid}}-2$	0.051	0.011	0.003	0.007	0.002	0.005	0.000	0.001
Canc.at $P_{\text{bid}}-3$	0.037	0.011	0.014	0.007	0.005	0.002	0.001	0.001
Canc.at $P_{\text{bid}}-4$	0.038	0.027	−0.006	0.010	0.010	0.002	0.002	0.001
Canc.at $P_{\text{bid}}-5$	0.056	0.030	0.014	0.011	0.013	0.008	−0.001	0.001
Canc.$\geq P_{\text{bid}}-10$	0.031	0.021	0.025	0.006	0.011	0.005	−0.000	0.001
Canc.$\geq P_{\text{bid}}-20$	0.027	0.029	0.017	0.016	0.007	0.008	0.001	0.001

Table 9: Representation of PCA factor indices in terms of observable characteristics of the limit order book and recent trading history

Covariate	Fact.1	Fact.2	Fact.3	Fact.4	Fact.5	Fact.6	Fact.7
Constant	-1.186	-0.287	-0.363	-2.496	0.227	-0.639	0.477
Slippage	-0.068	0.330	-0.230	-0.014	-0.280	0.042	-0.166
$\text{Spread}_{>0}$	-0.185	0.095	0.080	0.779	-0.058	-0.169	-0.076
ΔP_{ask}	0.028	0.129	-0.229	0.036	-0.103	0.150	-0.130
ΔP_{bid}	0.016	0.144	-0.221	-0.055	-0.152	-0.367	-0.101
$\log(Q_{\text{ask}})$	0.033	-0.131	-0.103	0.165	-0.367	0.712	0.124
$\Delta\log(Q_{\text{ask}})$	-0.004	-0.026	-0.051	0.193	-0.765	-0.380	0.459
$\log(Q_{\text{bid}})$	0.078	0.089	0.160	0.191	0.386	0.637	-0.318
$\Delta\log(Q_{\text{bid}})$	-0.058	0.042	0.012	-0.020	0.684	-0.753	-0.725
Side	0.034	-0.241	-0.223	0.025	-0.184	-0.086	-0.091
Side_{-1}	0.037	-0.211	-0.140	0.072	0.013	-0.015	-0.118
$F_{0-5''}$	-0.002	-0.077	-0.001	0.011	-0.038	-0.002	0.000
$F_{5-10''}$	-0.003	-0.048	-0.059	0.009	0.061	-0.013	0.002
$F_{10-15''}$	-0.001	-0.007	-0.059	0.019	0.066	0.000	-0.041
$F_{15-30''}$	-0.004	0.012	-0.031	0.003	0.027	0.003	-0.010
$F_{30-60''}$	-0.002	0.002	-0.022	0.003	0.011	-0.001	0.016
$F_{1-2'}$	-0.003	0.004	-0.013	-0.000	0.010	0.001	0.013
$T_{0-5''}$	0.059	0.016	-0.003	0.024	0.020	0.073	0.017
$T_{5-10''}$	0.038	-0.002	0.027	0.012	-0.022	-0.018	-0.011
$T_{10-15''}$	0.034	0.002	-0.002	0.006	0.002	-0.028	0.034
$T_{15-30''}$	0.027	0.002	0.004	0.010	0.004	-0.039	0.021
$T_{30-60''}$	0.016	0.004	0.004	0.008	-0.004	-0.011	0.002
$T_{1-2'}$	0.005	0.003	0.004	0.002	-0.007	0.002	0.000
$T_{2-5'}$	0.003	-0.001	0.001	0.002	-0.001	-0.003	-0.006
$T_{5-15'}$	0.002	0.000	-0.001	0.001	0.000	-0.000	-0.000

Table 10: Cox five-factor regressions for competing risks of sell order arrivals

Risk type (r)	Fact.1 (activity)	Fact.2 (imbalance)	Fact.3 (b.pressure)	Fact.4 (adv.selection)	Fact.5 (b.momentum)
A1: Sell below P_{bid}	0.988* (23.21)	−0.115* (−3.47)	−0.589* (−16.37)	−0.861* (−11.74)	−0.279* (−8.53)
A2: Market Sell	0.565* (40.04)	−0.463* (−32.42)	−0.423* (−33.19)	−0.275* (−16.55)	−0.325* (−22.63)
A3: Sell at P_{bid}	0.717* (43.34)	−0.245* (−14.23)	−0.391* (−24.78)	−0.495* (−21.15)	−0.329* (−18.62)
A4: Sell at $\leq P_{\text{ask}} - 1$	0.053* (2.45)	−0.078* (−3.00)	−0.114* (−4.76)	0.652* (26.61)	−0.140* (−5.25)
A5: Sell at $P_{\text{ask}} - 1$	0.340* (22.45)	−0.169* (−10.57)	−0.016 (−1.07)	0.168* (9.61)	−0.243* (−14.89)
A6: Sell at P_{ask}	0.412* (37.44)	−0.199* (−16.66)	−0.181* (17.22)	0.110* (9.38)	−0.132* (−10.30)
A7: Sell at $P_{\text{ask}} + 1$	0.537* (36.32)	−0.199* (−12.04)	0.329* (23.15)	−0.154* (−9.12)	0.091* (5.38)
A8: Sell at $P_{\text{ask}} + 2$	0.587* (29.20)	−0.117* (−4.89)	0.313* (15.66)	−0.108* (−4.56)	0.084* (3.71)
A9: Sell at $P_{\text{ask}} + 3$	0.621* (24.09)	−0.066 (−2.40)	0.362* (14.82)	−0.010 (−0.35)	−0.000 (−0.02)
A10: Sell at $P_{\text{ask}} + 4$	0.609* (16.56)	−0.063 (−1.66)	0.385* (11.24)	−0.092 (−2.24)	0.003 (0.07)
A11: Sell at $P_{\text{ask}} + 5$	0.624* (16.69)	−0.082 (−1.64)	0.334* (8.79)	0.015 (0.31)	−0.073 (−1.42)
A12: Sell at $\leq P_{\text{ask}} + 10$	0.599* (21.36)	−0.061 (−2.20)	0.336* (12.17)	0.019 (0.62)	−0.012 (−0.39)
A13: Sell at $\leq P_{\text{ask}} + 20$	0.431* (10.53)	−0.004 (−0.09)	0.394* (9.43)	0.096 (2.23)	−0.214* (−4.27)
A14: Sell above $P_{\text{ask}} + 20$	0.355* (8.17)	−0.038 (−0.89)	0.271* (6.82)	0.060 (1.30)	−0.160* (−3.69)

Table 11: Cox five-factor regressions for competing risks of buy order arrivals

Risk type (r)	Fact.1 (activity)	Fact.2 (imbalance)	Fact.3 (b.pressure)	Fact.4 (adv.selection)	Fact.5 (b.momentum)
B1: Buy above P_{ask}	1.013^* (22.83)	-0.004 (-0.08)	0.495^* (9.99)	-0.797^* (-9.55)	0.585^* (10.87)
B2: Market Buy	0.554^* (39.36)	0.366^* (29.45)	0.300^* (21.42)	-0.486^* (-30.06)	0.387^* (26.26)
B3: Buy at P_{ask}	0.767^* (46.34)	0.096^* (5.73)	0.249^* (14.64)	-0.650^* (-27.28)	0.463^* (24.83)
B4: Buy at $\geq P_{\text{bid}}+1$	0.123^* (5.59)	0.167^* (6.44)	0.286^* (12.52)	0.586^* (22.19)	0.156^* (5.16)
B5: Buy at $P_{\text{bid}}+1$	0.297^* (19.72)	0.176^* (11.47)	0.053^* (3.62)	0.159^* (9.31)	0.184^* (10.40)
B6: Buy at P_{bid}	0.364^* (33.95)	0.177^* (17.61)	-0.149^* (-14.29)	0.112^* (10.05)	0.107^* (9.65)
B7: Buy at $P_{\text{bid}}-1$	0.484^* (33.31)	0.094^* (7.58)	-0.316^* (-20.39)	-0.129^* (-7.64)	-0.048^* (-3.24)
B8: Buy at $P_{\text{bid}}-2$	0.536^* (27.67)	0.067^* (3.90)	-0.331^* (-17.11)	-0.089^* (-3.77)	-0.039 (-2.19)
B9: Buy at $P_{\text{bid}}-3$	0.493^* (20.07)	0.057 (2.35)	-0.292^* (-11.86)	-0.082^* (-2.72)	-0.004 (-0.18)
B10: Buy at $P_{\text{bid}}-4$	0.527^* (16.84)	0.074 (2.33)	-0.279^* (-9.15)	-0.035 (-0.88)	-0.077^* (-2.64)
B11: Buy at $P_{\text{bid}}-5$	0.454^* (13.99)	0.043 (1.27)	-0.232^* (-6.74)	0.200^* (5.34)	0.070 (1.92)
B12: Buy at $\geq P_{\text{bid}}-10$	0.471^* (20.73)	0.080^* (3.55)	-0.323^* (-13.46)	0.120^* (4.57)	0.023 (1.10)
B13: Buy at $\geq P_{\text{bid}}-20$	0.425^* (12.82)	0.143^* (4.54)	-0.242^* (-7.27)	0.169^* (4.78)	0.085^* (2.76)
B14: Buy below $P_{\text{bid}}-20$	0.334^* (9.45)	0.121^* (3.49)	-0.209^* (-6.16)	0.210^* (6.05)	0.133^* (3.75)

Table 12: Cox five-factor regressions for competing risks of limit order cancellations

Risk type (r)	Fact.1 (activity)	Fact.2 (imbalance)	Fact.3 (b.pressure)	Fact.4 (adv.selection)	Fact.5 (b.momentum)
AC6: Canc.at P_{ask}	0.292^* (25.56)	-0.351^* (-24.44)	-0.057^* (-4.91)	0.151^* (12.55)	-0.171^* (-11.40)
AC7: Canc.at $P_{\text{ask}}+1$	0.446^* (22.85)	-0.427^* (-20.52)	0.096^* (5.30)	-0.016 (-0.73)	0.080^* (4.01)
AC8: Canc.at $P_{\text{ask}}+2$	0.513^* (20.98)	-0.511^* (-17.50)	0.069^* (2.88)	-0.002 (-0.06)	0.095^* (3.48)
AC9: Canc.at $P_{\text{ask}}+3$	0.553^* (18.30)	-0.518^* (-13.57)	-0.008 (-0.24)	0.017 (0.45)	0.093 (2.32)
AC10: Canc.at $P_{\text{ask}}+4$	0.589^* (14.29)	-0.549^* (-11.10)	-0.004 (-0.10)	-0.090 (-1.76)	0.108 (2.20)
AC11: Canc.at $P_{\text{ask}}+5$	0.581^* (12.42)	-0.484^* (-8.77)	-0.103 (-2.27)	0.028 (0.50)	-0.008 (-0.16)
AC12: Canc.at $\leq P_{\text{ask}}+10$	0.468^* (17.53)	-0.387^* (-12.35)	-0.162^* (-5.81)	0.176^* (6.22)	0.118^* (4.21)
AC13: Canc.at $\leq P_{\text{ask}}+20$	0.444^* (12.68)	-0.223^* (-5.36)	-0.229^* (-5.97)	0.074 (1.72)	0.182^* (4.70)
AC14: Canc.above $P_{\text{ask}}+20$	0.397^* (9.88)	0.039 (0.95)	-0.179^* (-4.21)	0.249^* (6.02)	0.182^* (4.33)
BC6: Canc.at P_{bid}	0.214^* (18.67)	0.324^* (29.02)	0.105^* (8.96)	0.079^* (6.87)	0.188^* (14.53)
BC7: Canc.at $P_{\text{bid}}-1$	0.365^* (18.63)	0.259^* (16.80)	-0.062^* (-3.20)	-0.045 (-2.08)	0.025 (1.40)
BC8: Canc.at $P_{\text{bid}}-2$	0.418^* (18.18)	0.496^* (20.75)	-0.123^* (-4.88)	-0.155^* (-5.68)	-0.105^* (-4.26)
BC9: Canc.at $P_{\text{bid}}-3$	0.384^* (12.21)	0.536^* (16.92)	-0.055 (-1.57)	-0.103^* (-2.83)	-0.086^* (-2.61)
BC10: Canc.at $P_{\text{bid}}-4$	0.463^* (10.85)	0.317^* (10.99)	0.047 (1.13)	-0.043 (-0.89)	0.036 (1.03)
BC11: Canc.at $P_{\text{bid}}-5$	0.506^* (12.43)	0.559^* (13.65)	0.138^* (3.56)	0.124^* (2.84)	-0.000 (-0.01)
BC12: Canc.at $\geq P_{\text{bid}}-10$	0.439^* (16.60)	0.360^* (18.27)	0.188^* (7.24)	0.036 (1.27)	0.011 (0.52)
BC13: Canc.at $\geq P_{\text{bid}}-20$	0.426^* (11.50)	0.233^* (9.09)	0.252^* (8.03)	0.131^* (3.69)	-0.158^* (-5.37)
BC14: Canc.below $P_{\text{bid}}-20$	0.372^* (9.32)	0.096^* (2.56)	0.291^* (7.88)	0.172^* (4.30)	-0.101 (-2.44)

Table 13: Eigenvalues and cumulative contribution of principal components to the competing risk indices

	1	2	3	4	5	6	7	8	9	10
Eigenvalue	20.16	7.26	5.70	4.12	2.15	1.01	0.88	0.66	0.61	0.56
Cumulative	0.438	0.596	0.720	0.810	0.857	0.879	0.898	0.912	0.925	0.937

Table 14: Modified classification of observable events in the forecasting model

Risk category (s)	Description of event	Risk types (r) included
AA	Seller-initiated transaction	Risks A1–A3 in Table **??**
AP+	Ask price improvement	Risks A4–A5 in Table **??**
AD+	Ask depth improvement	Risk A6 in Table **??**
A−	Ask touch cancellation	Risk AC6 in Table **??**
BB	Buyer-initiated transaction	Risks B1–B3 in Table **??**
BP+	Bid price improvement	Risks B4–B5 in Table **??**
BD+	Bid depth improvement	Risk B6 in Table **??**
B−	Bid touch cancellation	Risks BC6 in Table **??**

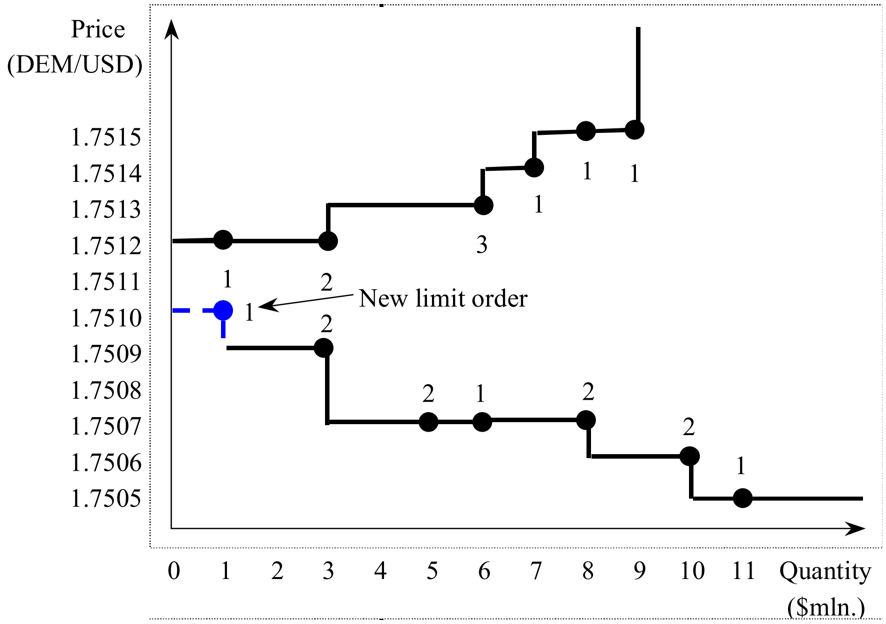

Figure 2:
Supply and demand curves after the price improvement occurs on demand side

Figure 2 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a new limit order to purchase one million dollars at the bid price DEM 1.7510 per dollar. The arrival of new limit order leads to reduction of the market bid-ask spread, and shifts the prior ersatz demand curve to the right. The improved market liquidity associated with such an event validates the term "price improvement". However, the quantity (depth) available at the improved bid price DEM 1.7510 per dollar is smaller than the bid depth at the previous bid quote DEM 1.7509 per dollar on Figure 1.

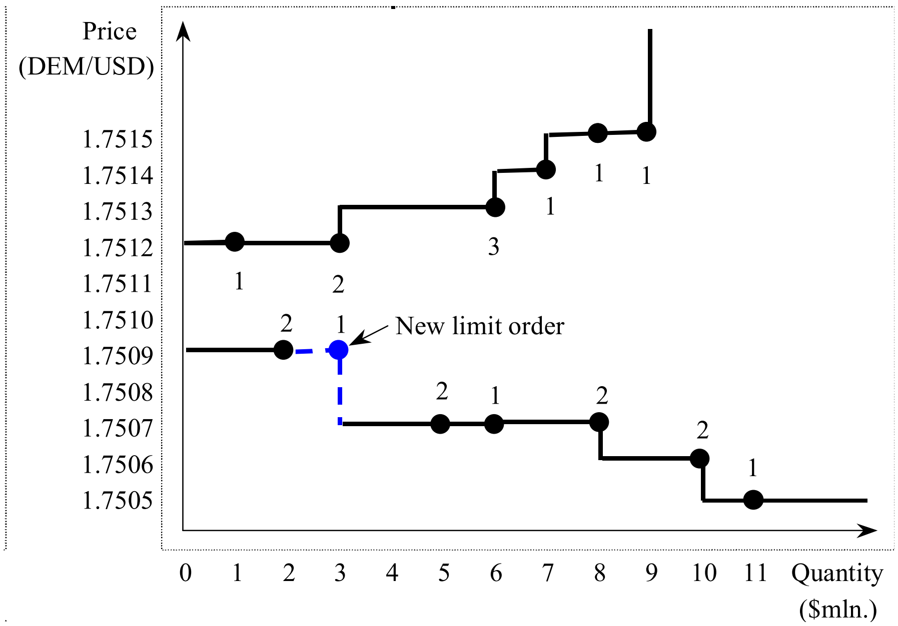

Figure 3:
Supply and demand curves after the depth improvement occurs on demand side

Figure 3 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a new limit order to purchase one million dollars at the bid price DEM 1.7509 per dollar. The arrival of new limit order does not change the market bid-ask spread, but it shifts the portion of prior ersatz demand curve below the best bid price DEM 1.7509 per dollar to the right. The improved market depth on the bid side associated with such an event validates the term "depth improvement".

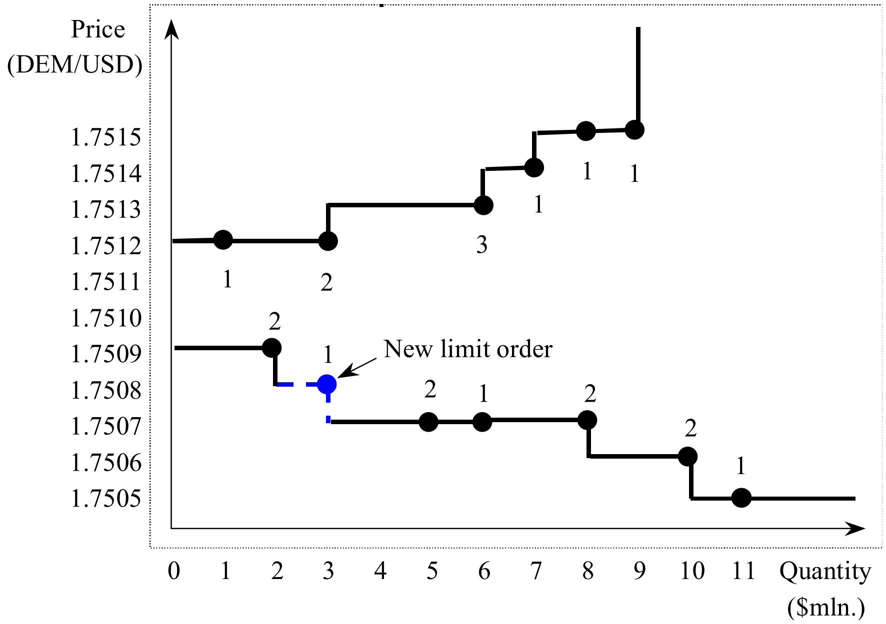

Figure 4:
Supply and demand curves after arrival of subsidiary bid one tick below the touch

Figure 4 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a new limit order to purchase one million dollars at the bid price DEM 1.7508 per dollar, which is lower than the best bid price DEM 1.7509 available on the market. The arrival of new limit order does not change the public information on the Reuters D2000-2 trading screens, in particular, it does not affect the size of the bid-ask spread and the market depth at the touch (best bid quote). However, the market liquidity improves in a broader sense as the market depth one tick below the best bid quote increases.

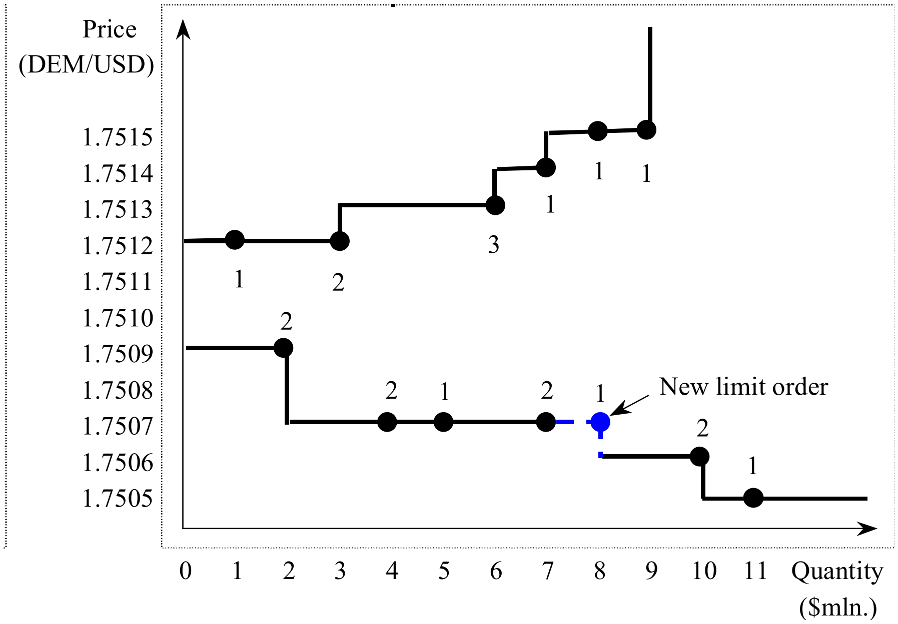

Figure 5:
Supply and demand curves after arrival of subsidiary bid two ticks below the touch

Figure 5 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a new limit order to purchase one million dollars at the bid price DEM 1.7507 per dollar, which is two ticks below the best bid price DEM 1.7509 available on the market. The arrival of new limit order does not change the public information on the Reuters D2000-2 trading screens, since the market event is associated with improvement of market liquidity deep on the bid side of the limit order book.

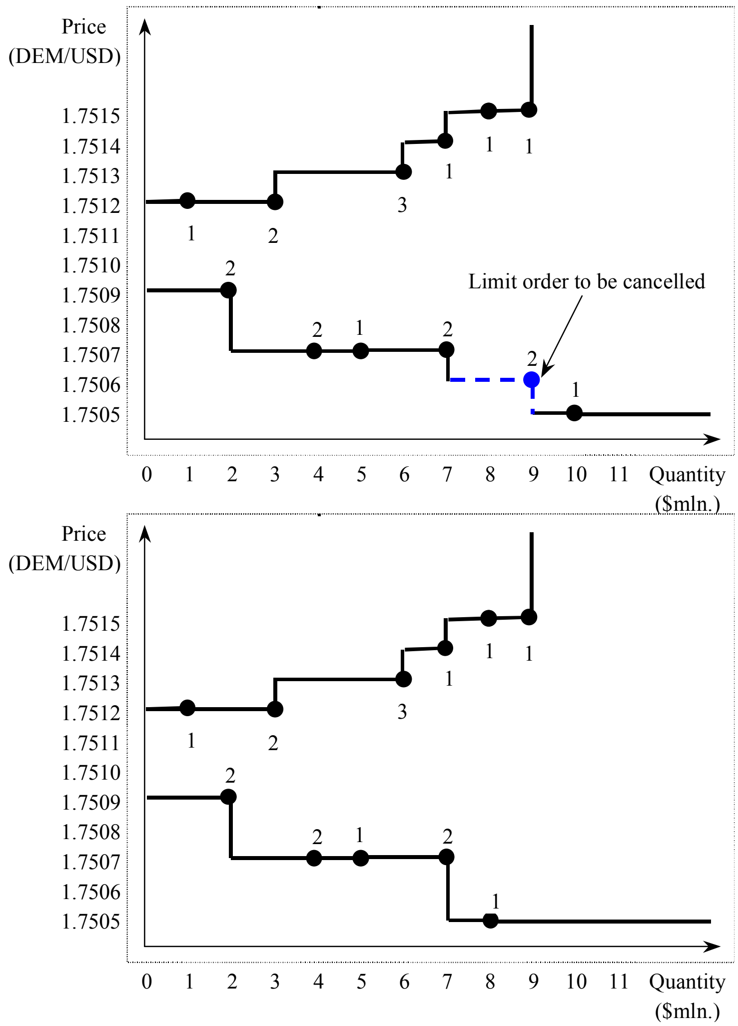

Figure 6:
Supply and demand curves after subsidiary bid cancellation

Figure 6 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after cancellation of the subsidiary limit order to purchase two million dollars at the bid price DEM 1.7506 per dollar. The cancellation event does not affect public information on the Reuters D2000-2 trading screens, as it is associated with the deterioration of market liquidity deep inside the bid side of limit order book.

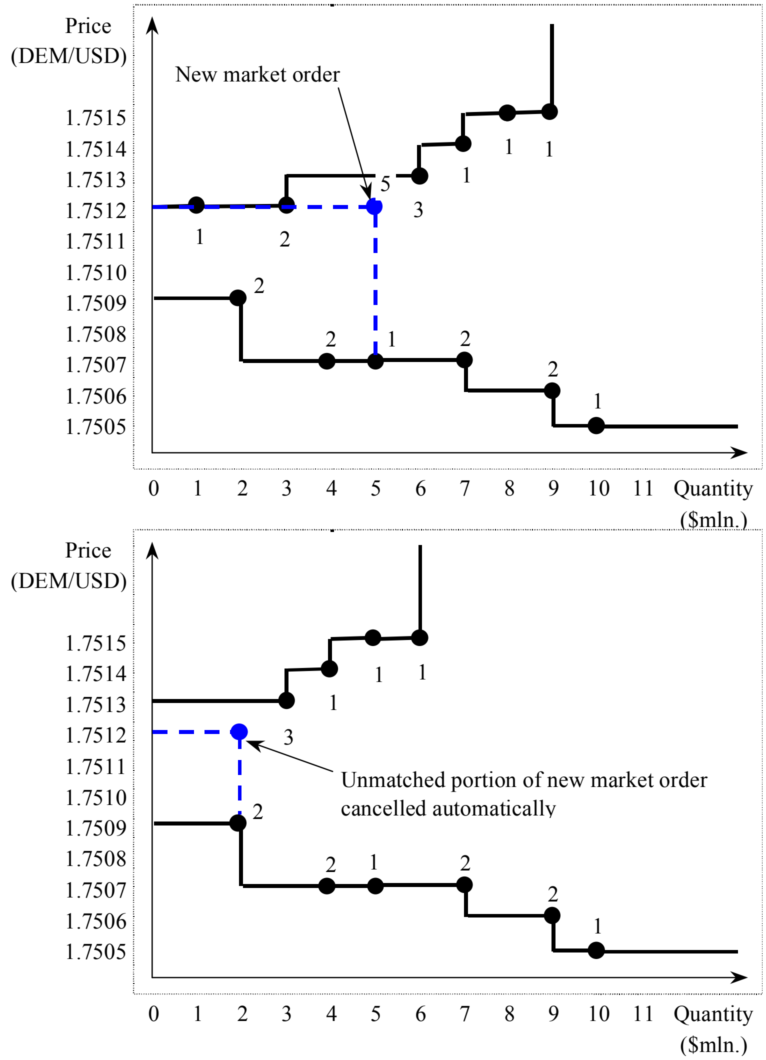

Figure 7:
Supply and demand curves after arrival of large market buy order

Figure 7 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a market order to purchase five million dollars, which was submitted at the best ask market price of DEM 1.7512 per dollar. Since the quantity available at this price is only three million dollars, part of demand for liquidity created by the new market order arrival is not satisfied. The unmatched portion of the market order gets cancelled immediately, while the best ask price goes up one tick to DEM 1.7513 per dollar.

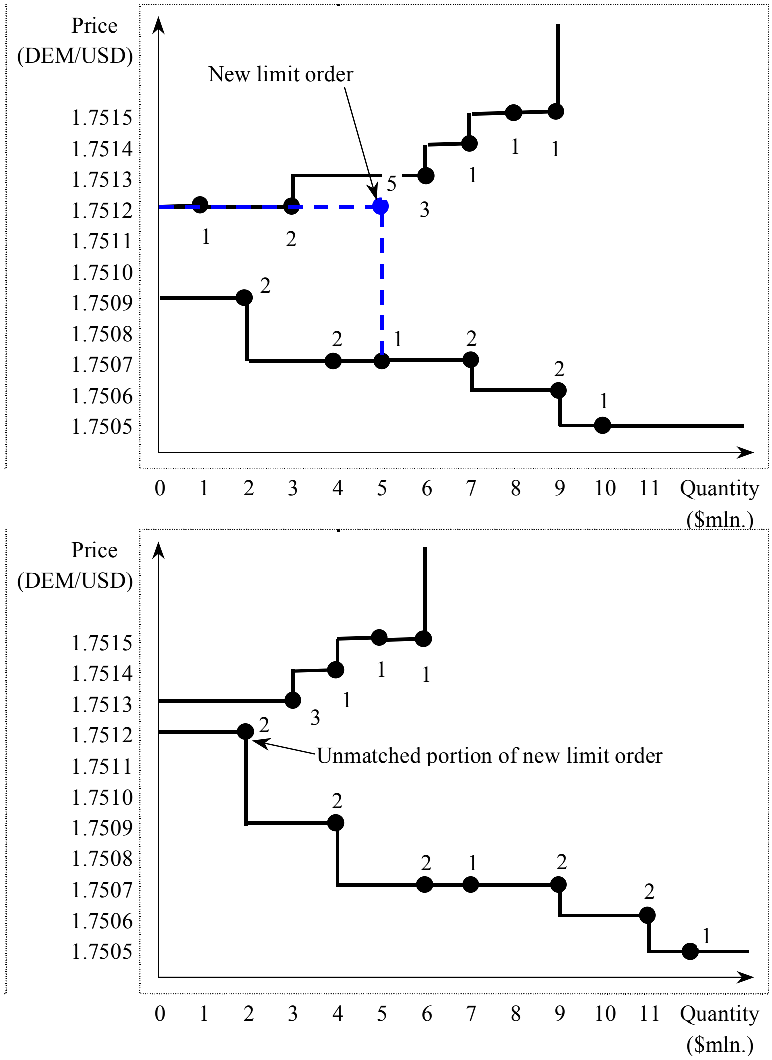

Figure 8:
Supply and demand curves after arrival of large aggressive limit bid

Figure 8 displays the change in the state of the Reuters D2000-2 electronic limit order book shown on Figure 1 after arrival of a limit order to purchase five million dollars at the limit order price which coincides with the best ask market price of DEM 1.7512 per dollar previously available on the market. The situation is analogous to the submission of buy market order (Figure 7), except that the unmatched portion of the arriving aggressive limit order remains on the limit order book, leading to the improvement of the best bid price by three ticks to DEM 1.7512 per dollar. The best bid price moves one tick up to DEM 1.7513 per dollar.

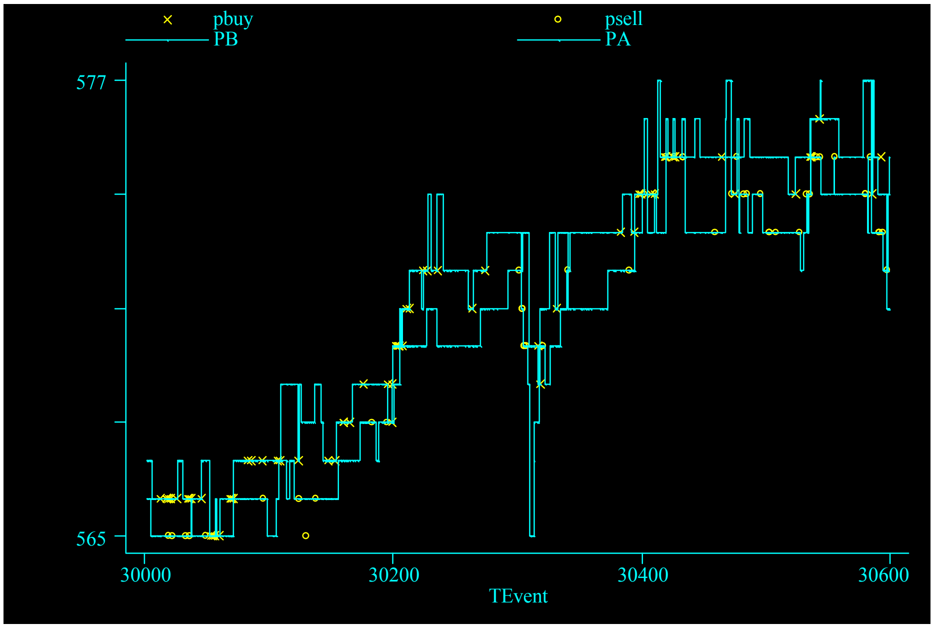

Figure 9:
A representative ten-minute sample of best market bid and ask quotes and transactions in Reuters D2000-2 trading system

Figure 9 shows a small subsample of continuously sampled best market bid and ask quotes, as well as the times and prices of buyer- and seller-initiated transactions (marked by small white crosses and knots, respectively). All prices were obtained by matching limit and market orders from the original Reuters D2000-2 data set. The sampled time period covers the trading hours 8:20 to 8:30 a.m. GMT on Monday, October 6, 1997.

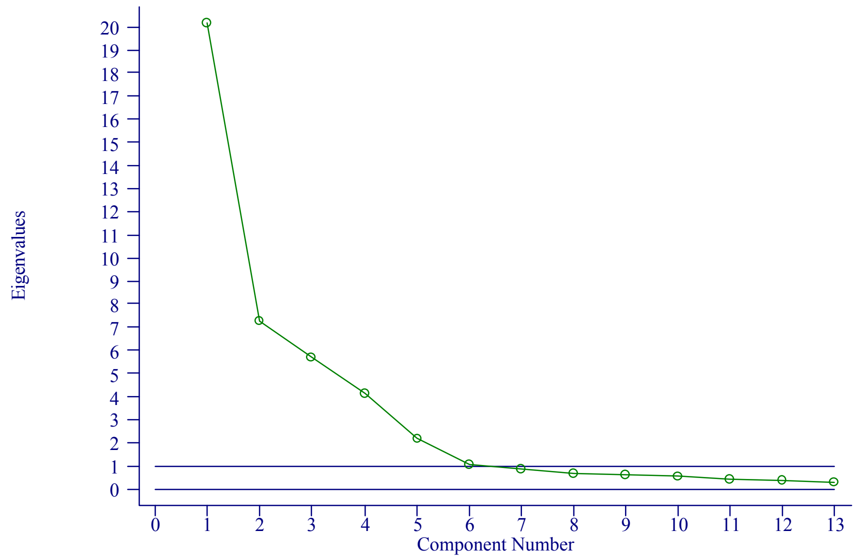

Figure 10:
Eigenvalues of principal components for the covariate indices of competing risks

Figure 10 displays the plot of eigenvalues corresponding to the first $Q = 13$ PCA factors of the competing risk indices $r = 1, ..., R$. Since only five of these eigenvalues are unambiguously above the horizontal line $\lambda = 1$, which suggests PCA factors 1 through 5 can be treated as independent pervasive components driving the market dynamics. Even though factors 6 and 7 are only marginally significant, they are retained to prevent wrongful exclusion of additional, marginally significant factors.

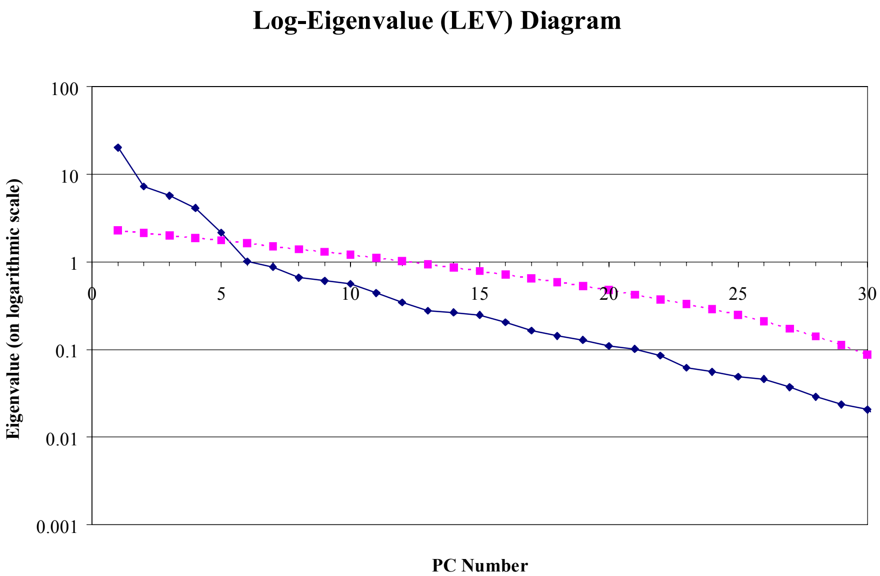

Figure 11:
Log-eigenvalue (LEV) diagram for principal components of the competing risk indices

Figure 11 displays the LEV diagram, which is obtained by replotting Figure **??** on the logarithmic scale. The darker line plots the eigenvalues corresponding to the PCA factors of the competing risk indices $r = 1, ..., R$. Deviations of plot from the linear pattern on the left for low eigenvalues suggests that retention of $Q = 5$ PCA factors for further analysis would be appropriate. The choice of $Q = 5$ is confirmed by the comparison of the actual eigenvalues with the 95% confidence bounds shown on the graph in pink color that are obtained by a bootstrap procedure under the null hypothesis that all eigenvalues corresponding to the Qth and higher-order PCA factors are equal to each other.

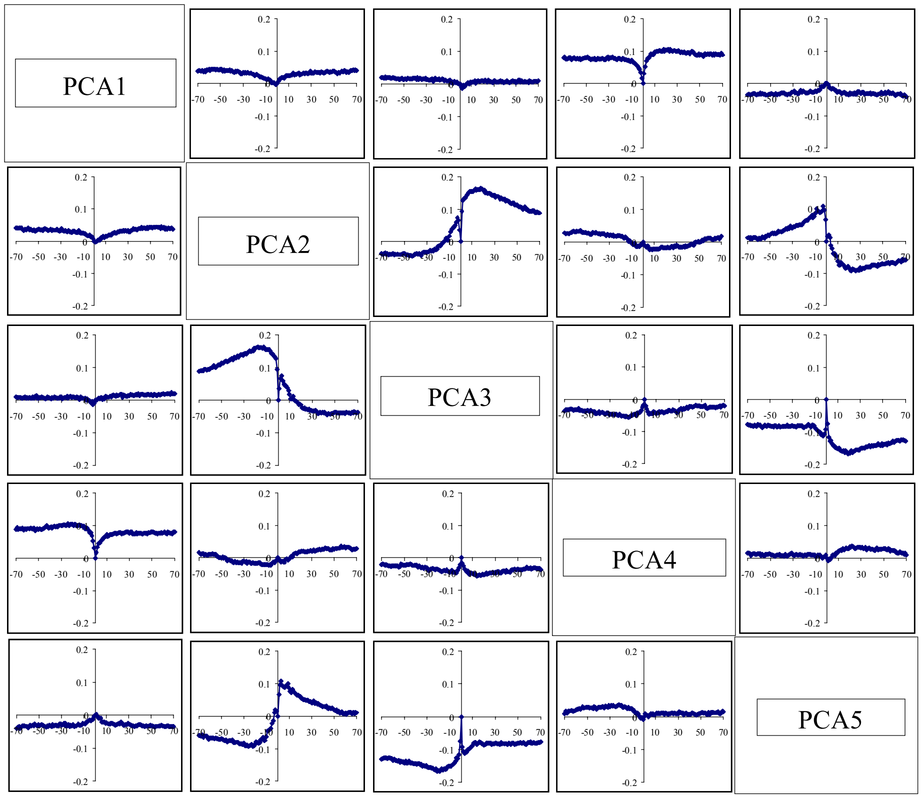

Figure 12:
Cross-correlograms of the first five PCA factors (principal components)

Figure 12 displays the estimated off-diagonal terms of cross-correlograms of the first five PCA factors associated with the aggregate risks of limit order book events. Cell (i, j) of the chart contains the estimated cross-correlogram of factors i and j defined by the formula

$$\widehat{\rho}_{ij}(h) = \widehat{\text{Corr}}(f_{i,n},\ f_{j,n+h}) = \tfrac{1}{N-h} \sum_{n=1}^{N-h} f_{i,n} f_{j,n+h},$$

where h is the lead (forward shift) of factor j relative to factor i measured by the number of epochs. All calculations are performed for the values of h between -70 and 70 and based on the subsample covering the liquid trading hours (6 a.m. to 5 p.m. GMT) on October 6–8, 1997.

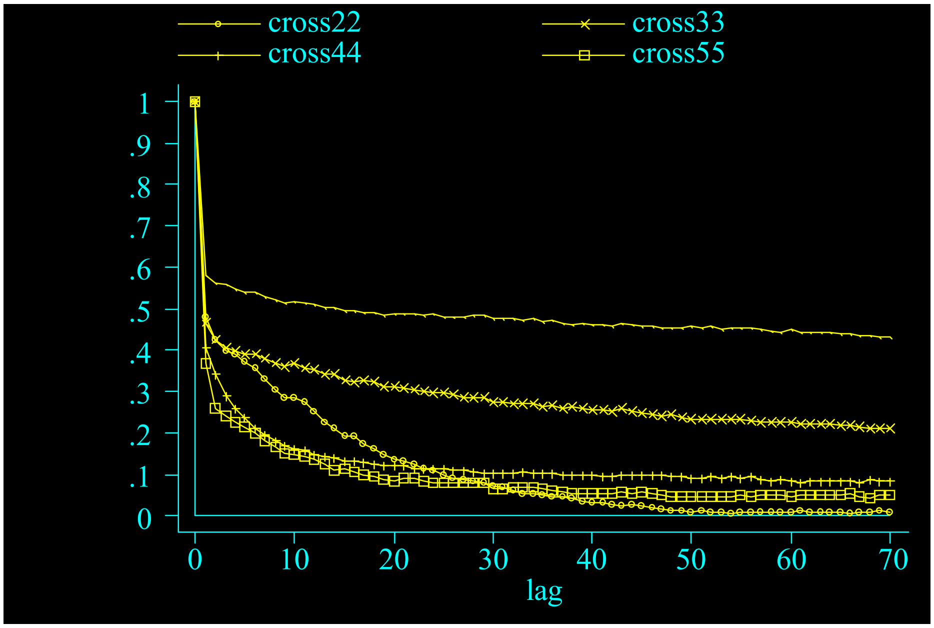

Figure 13:
Autocorrelograms of the first five PCA factors (principal components)

Figure 13 displays the estimated autocorrelation functions (autocorrelograms) for the first five PCA factors associated constructed from the risk indices of the aggregated limit order book activity. Autocorrelograms of factor 1 (shown by the unmarked solid line) and factor 3 (highlighted by "x" symbols) demonstrate slow rate of decline with the lag order, which is a clear evidence of long memory and potential nonstationarity. Autocorrelograms of factors 4 and 5 (highlighted by pluses and squares, respectively) also decline relatively slowly which serves as an evidence of long memory. The autocorrelogram of factor 2 (highlighted by circles) rapidly declines to zero with the lag order and becomes indistinguishable from zero at lag 50. All calculations are performed using the subsample covering only the liquid trading hours 6 am to 5 pm GMT on October 6–8, 1997.

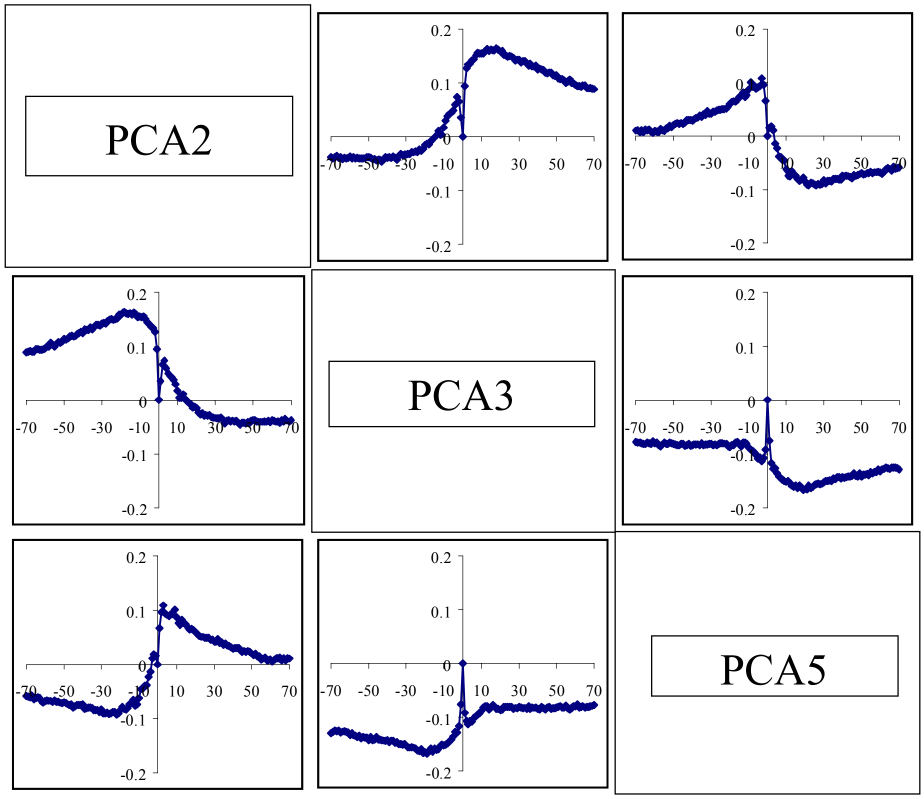

Figure 14:
Cross-correlograms of activity imbalance, buyer pressure, and bull market momentum (PCA factors 2, 3 and 5)

Figure 14 highlights the portion of the estimated cross-correlogram 12 that displays the interaction of the second, third, and fifth PCA factors. These factors are identified in the text as the limit order book activity imbalance (factor 2), the buyer pressure (factor 3), and the bull market momentum (factor 5). Cell (i, j) of the chart contains the estimated cross-correlogram

$$\widehat{\rho}_{ij}(h) = \widehat{\mathrm{Corr}}(f_{i,n},\ f_{j,n+h}) = \frac{1}{N-h}\sum_{n=1}^{N-h} f_{i,n}f_{j,n+h},$$

of PCA factors i and j, where h is the lead (forward shift) of factor j relative to factor i measured as the number of consecutive ticks. All calculations are performed for the values of h between -70 and 70 and based on the subsample covering the liquid trading hours (6 a.m. to 5 p.m. GMT) on October 6–8, 1997.

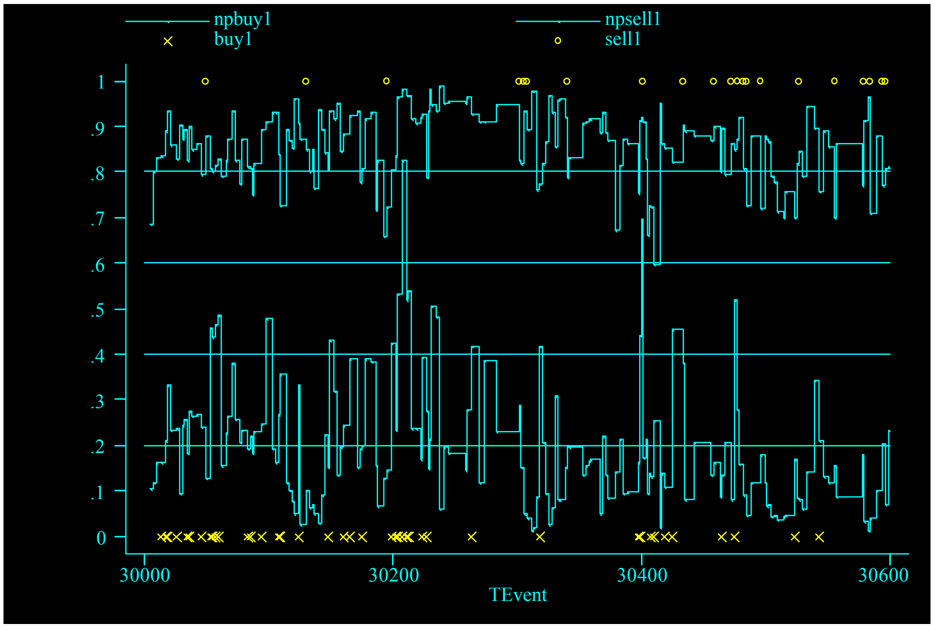

Figure 15:
A representative ten-minute sample for one-step-ahead forecast probabilities of buy and sell transactions

Figure 15 shows a small subsample of one-step forecast probabilities of buyer- and seller-initiated transactions made after at least one second elapsed since the previous event, as well as the times of actual buyer-initiated transactions (shown by crosses) and seller-initiated transactions (shown by knots). The forecasts are based on the version of PCA factor competing risks model with $S = 8$ types of observable risks and the covariate structure comprised by $Q = 5$ PCA factors as described in section 6. The sample period covers the episode 8:20 to 8:30 a.m. GMT on Monday, October 6, 1997, which is identical to the period used to produce the graph of the best market bid and ask quotes and transactions (Figure 9). The forecast probabilities of buyer-initiated trades are shown on the plot as the distance of lower solid line from the horizontal zero-probability line. The forecast probabilities of seller-initiated trades are shown as the distance of the upper solid line from the 100% probability horizontal line.

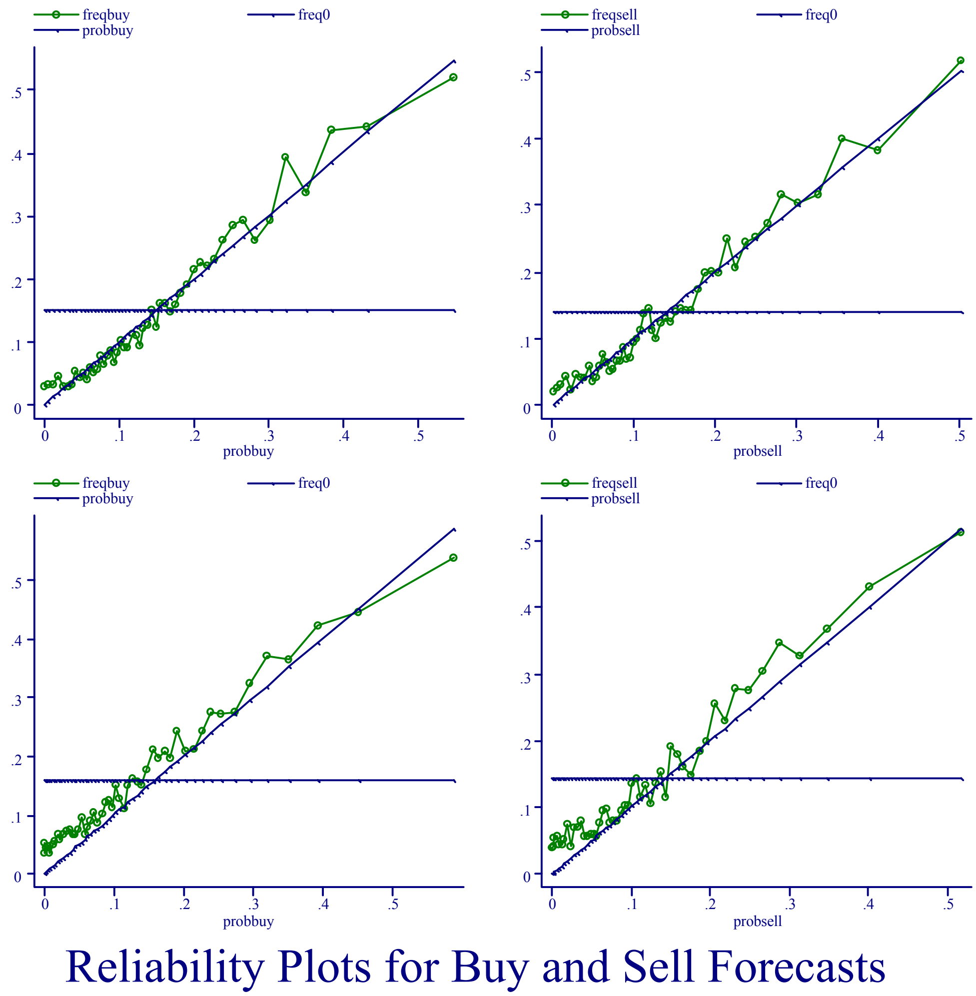

Figure 16:
Reliability plots for one-step-ahead forecast probabilities of buy and sell transactions

Figure 16 displays the reliability plots for one-step-ahead forecast probabilities of buyer- and seller-initiated transactions made after at least one second elapsed since the previous event. The forecasts are based on the version of PCA factor competing risks model with $S = 8$ types of observable risks and the covariate structure comprised by $Q = 5$ PCA factors as described in section 6. Two diagrams on the top are based on the liquid trading hours of the first three trading days (the model estimation period). The left and right diagrams on the top plot, respectively, the fractions of epochs terminated with buy and sell transactions against the forecast probabilities of such events. Two diagrams on the bottom are based on the liquid trading hours of the last two days (the out-of-sample period). The left and right diagrams on the bottom plot, respectively, the fractions of epochs terminated with buy and sell transactions against the forecast probabilities of such events.

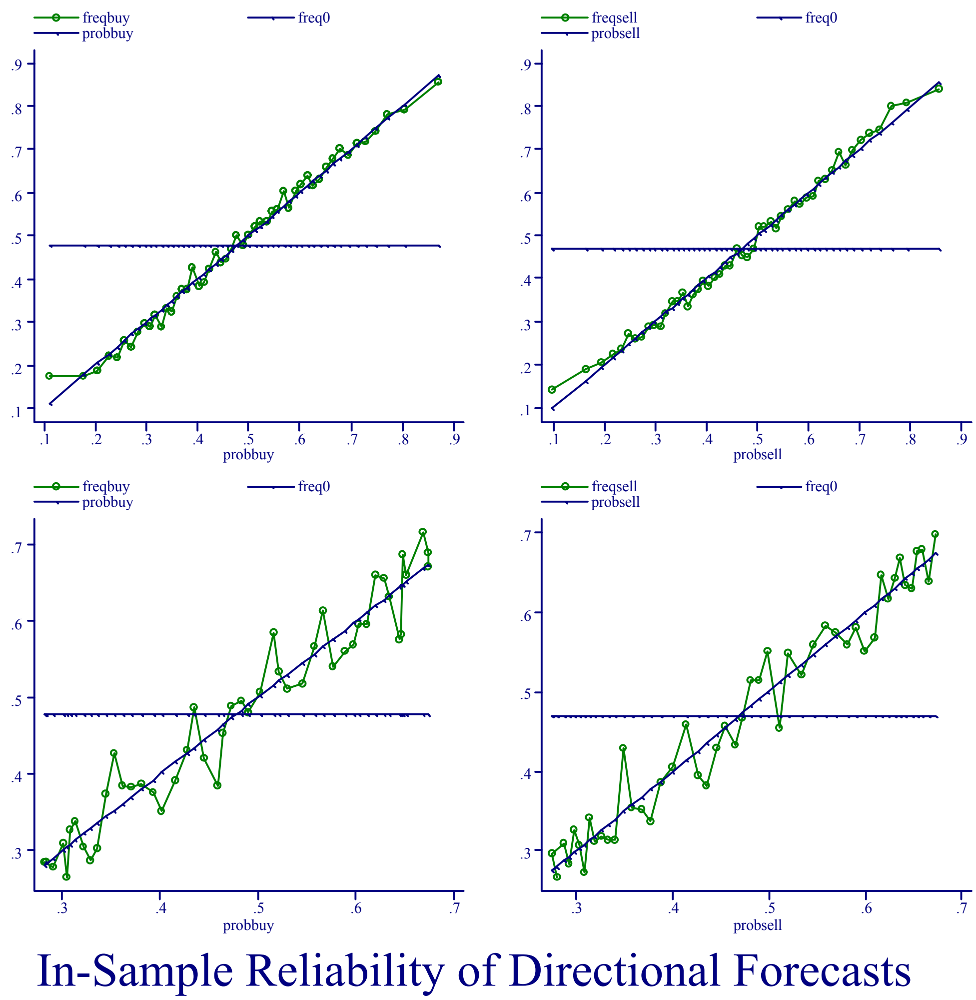

Figure 17:
In-sample reliability plots for probability forecasts of next trade in 30 seconds

Two diagrams on the left-hand side of Figure 17 show reliability plots for probability forecasts of the event that the next transaction in the limit order book will be initiated by a buyer and will occur in 30 seconds since the time of forecast. Two diagrams on the right-hand side of Figure 17 show reliability plots for probability forecasts of the event that the next transaction in the limit order book will be initiated by a seller and will occur in 30 seconds since the time of forecast. The forecasts evaluated on the top section of the graph are produced by the trinomial logit regression with the covariates given by the five competing risks PCA factors. The forecasts evaluated on the lower section of the graph are produced by the benchmark forecasting model based on the trinomial logit regression with the covariates given by the signs of last 10 transactions (± 1 if the transaction was initiated by seller/buyer). All diagrams are based on the data from the estimation period 6 a.m. to 5 p.m. GMT on October 6–8, 1997.

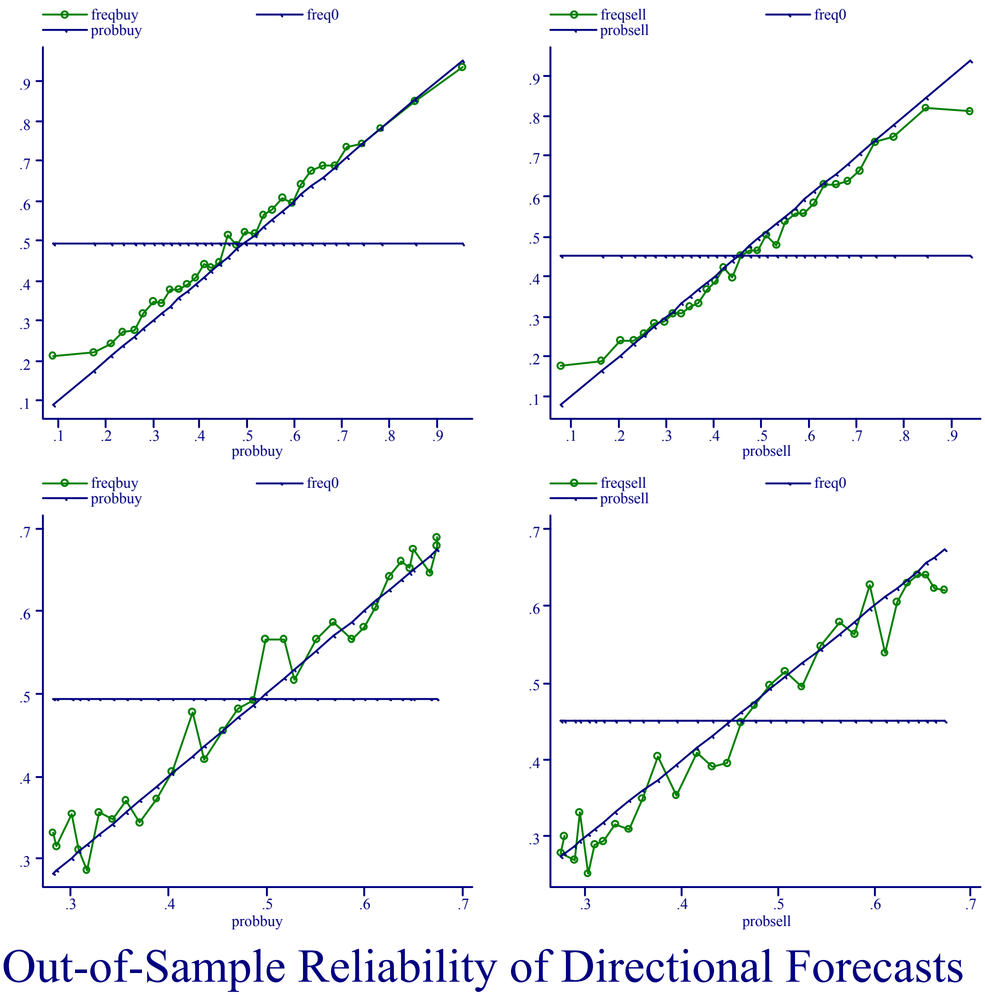

Figure 18:
Out-of-sample reliability plots for probability forecasts of next trade in 30 seconds

Two diagrams on the left-hand side of Figure 18 show reliability plots for probability forecasts of the event that the next transaction in the limit order book will be initiated by a buyer and will occur in 30 seconds since the time of forecast. Two diagrams on the right-hand side of Figure 18 show reliability plots for probability forecasts of the event that the next transaction in the limit order book will be initiated by a seller and will occur in 30 seconds since the time of forecast. The forecasts evaluated on the top section of the graph are produced by the trinomial logit regression with the covariates given by the five competing risks PCA factors. The forecasts evaluated in the lower section of the graph are produced by the benchmark forecasting model based on the trinomial logit regression with the covariates given by the signs of last 10 transactions (± 1 if the transaction was initiated by seller/buyer). All diagrams are based on the data from the out-of-sample period covering the trading hours 6 a.m. to 5 p.m. GMT on October 9–10, 1997.