CAEPR INDIANA UNIVERSITY
CENTER FOR APPLIED ECONOMICS AND POLICY RESEARCH

CAEPR Working Paper
#2006-015

A Competing Risk Analysis of Executions and Cancellations in a Limit Order Market

Bidisha Chakrabarty
Saint Louis University

Zhaohui Han
Financial Engineering Group, ITG Inc.

Konstantin Tyurin
Indiana University Bloomington

Xiaoyong Zheng
North Carolina State University

4 October 2006

A Competing Risk Analysis of Executions and Cancellations in a Limit Order Market*

Bidisha Chakrabarty† Zhaohui Han‡ Konstantin Tyurin§
Xiaoyong Zheng¶

October 4, 2006

Abstract

The competing risks technique is applied to the analysis of times to execution and cancellation of limit orders submitted on an electronic trading platform. Time-to-execution is found to be more sensitive to the limit price variation than time-to-cancellation, even though it is less sensitive to the limit order size. More importantly, investors who aim to reduce the expected time-to-execution for their limit orders *without inducing any significant increase in the risk of subsequent cancellation* should submit their orders when the market depth is smaller on the side of their orders or when the market depth is greater on the opposite side of their orders. We also provide a new diagnostic plots method for evaluating the goodness-of-fit of different competing risks models.

JEL classification: G14, G23

Keywords: Market microstructure, limit order, competing risks, hazard rate, frailty

---

*We thank Craig Holden, Patrik Sandas, and participants of the 2005 FMA annual meeting and the Financial Engineering Group weekly seminar at ITG, Inc. for their valuable comments. All remaining errors are our own. No senior authorship is assigned.

†Department of Finance, Saint Louis University. E-Mail: chakrab@slu.edu.

‡Financial Engineering Group, ITG Inc. E-Mail: dhan@itginc.com.

§Department of Economics, Indiana University. E-Mail: ktyurin@indiana.edu.

¶Department of Agricultural and Resource Economics, North Carolina State University. E-Mail: xzheng@ncsu.edu.

I. Introduction

The role of limit orders in the process of price discovery in financial markets has been extensively studied. Limit orders, defined as price-contingent instructions to buy or sell a financial security at the specified (limit) price, represent the major part of liquidity on organized exchanges. Most electronic equity markets around the globe are organized as pure limit order books; even on hybrid markets such as the New York Stock Exchange (NYSE), limit orders account for more than half of all trading activities.

Much of the related literature discusses the following aspects of limit order markets: (1) transaction costs, (2) spread decomposition, (3) order submission strategies, and (4) price formation.[1] Parlour (1998) presents a model of limit order book using a stochastic dynamic game and characterizes the optimal order choice between submitting a limit order and a market order. However, in her model limit order cancellations are not allowed. Hollifield *et al.* (2004) build a structural model of a pure limit order market which captures the trade-off among the order price, the order execution probability, and the winner's curse risk associated with different feasible order choices. Battalio *et al.* (2002) gauge execution quality across markets by comparing the limit order fill rates and times to execution on primary and regional exchanges. Harris and Panchapagesan (2005) examine whether the limit order book has a signaling effect on the future price change and whether the NYSE specialists take advantage of this information. Overall, the focus has been on limit order executions. Little attention has been paid to limit order cancellations.[2] In this paper we explicitly model cancellation of limit orders (jointly with execution) and show that limit order cancellations contain information that should also be incorporated into decisions regarding order submission strategies.

Limit order traders face two types of risk: (i) execution uncertainty, since trade at the limit price is not guaranteed; and (ii) adverse selection (also known as a "pick-off" risk), realized when better informed market participants take advantage of slow (or less informed) limit order traders. Limit order traders need analytical tools for efficient management of these risks. In particular, models that predict the possible outcomes (execution or cancellation) of their orders given these two types of risk should be of benefit to such traders. In this paper we will develop three general classes of models, separately for buy and sell orders, and study their properties.

Duration of limit order (defined as the time interval between the limit order arrival and its termination) plays an important role in the determination of transactions cost and

[1] See Henker and Martens (2003), Peterson and Sirri (2002), Biais *et al.* (1995, 1999).

[2] Notable exceptions are Lo *et al* (2002) and Hollifield et al (forthcoming).

opportunity cost of limit orders. Although a limit order trader does not face the price risk associated with a market order, at the time of limit order submission it is unknown how long it will take to fill the order, and whether the limit order will be executed at all. Expected limit order execution and cancellation times also affect market liquidity and are likely to be major factors driving the dynamics of the bid-ask spread. Lo *et al.* (2002) estimate an econometric model for the conditional distribution of limit order execution times as a function of economic variables such as the limit price, order size, and current market conditions. They find that limit order execution times are very sensitive to the limit price, but are not sensitive to limit order size. In a related strand of literature, Bisière and Kamionka (2000) and Tyurin (2003) use the competing risk methodology popular in survival analysis to model the hazard rates of order arrival, execution, and cancellation on limit order markets. Their models can be viewed as price formation models, which capture the interaction between quotes, transactions, and cancellation events in financial markets. In a study that is most closely related to this one, Hollifield et al (forthcoming) consider both executions and cancellations of limit orders on the Vancouver Stock Exchange to estimate gains from trade in a limit order trading platform.

In spite of the relative lack of focus on limit order cancellations, it is interesting to note that cancellation events represent the most common cause of limit order termination in modern electronic limit order markets. For example, according to Hasbrouck and Saar (2004) more than 80% of all limit orders submitted on Island ECN do not receive execution and are ultimately cancelled. Yeo (2004) reports that approximately 40% of all limit orders submitted on NYSE in 2001 were cancelled. Ellul *et. al.* (2003) and Lo and Sapp (2005) examine the determinants of order choice strategies of investors, including the option to cancel previously placed limit orders, while Yeo (2004) studies investors' strategies after cancellation of limit orders. All three papers are implemented in a discrete time framework and use discrete choice models to study traders' decisions.

In this paper we further investigate executions and cancellations of limit orders, extending the previous studies in several directions. We analyze recent data from the INET trading platform – a completely automated limit order book. We use the ITCH® data feed, which contains information on the entry, processing, and execution of all orders submitted to INET. The rich and detailed data in the ITCH® files allow us to investigate the relationship between the duration times and alternative outcomes of limit orders, and study the dependence of duration times on order characteristics and market conditions.

We find that of the various models we empirically test, the Weibull proportional hazard model with independent competing risks best describes the times to limit order executions and cancellations. Our results agree with previous findings and intuition. Time-to-execution is found to be more sensitive to the limit price variation than time-to-cancellation, even though it is less sensitive to the limit order size. Therefore, a safe way to reduce the time-to-execution for both buy and sell orders with a minimal increase in non-execution risk is to increase the price aggressiveness of the order. More importantly, investors who aim to reduce the expected time-to-execution for their limit orders *without inducing any significant increase in the risk of subsequent cancellation* should submit their orders when the market depth is smaller on the side of their orders or when the market depth is greater on the opposite side of their orders.

Our paper contributes to the existing literature in several ways. First, in a continuous time framework we model jointly the limit order cancellation and execution events. From an econometric perspective, the process of limit order execution is best approached by the use of competing risk models that are popular in the survival analysis literature. In the competing risks framework, several risk factors compete against each other for the termination event. After a limit order is placed, the limit order can be executed, cancelled/modified, or expires if expiration time is provided. Therefore, the execution and cancellation outcomes can be treated as competing risks, and the expiration outcome can be treated as a censoring time for those competing risks. Our models allow us to capture the information in cancellations, in addition to executions of limit orders.

Second, we treat limit order cancellation as a random event rather than an investor's choice as in Ellul *et al.* (2003) or as a censoring event as in Lo *et al.* (2002). Although traders are free to choose whether and when to cancel the limit orders they submit, they might have little if any knowledge at the time of order submission as to whether their orders would be ultimately cancelled or executed. Therefore, it is reasonable to consider cancellation as an additional risk that could cause the termination of a limit order.

Third, the competing risk model allows us to obtain different effects of the explanatory variables on different risks. This is not only of academic interest but also has useful practical implications. For example, if an investor increases her limit price for a buy limit order, then the probability of the order being executed is increased. But at the same time, she conveys information to the market by increasing the limit buy price, indicating that her private valuation of the asset is high. In response, other investors might place more aggressive limit orders or marketable limit orders, and the bid-ask spread changes accordingly. This causes changes in the market conditions, which might induce the investor who placed the limit order at the beginning to cancel her order, submit a new one, or stay away from the market altogether. Thus, an increase in the limit order price might also increase the probability of later cancellation. Our methodology recognizes this interdependence, and provides a way for investors to develop optimal limit order submission strategies. Modeling this interdependence is a contribution of our methodology.

Fourth, we test several competing models that can be used to analyze limit order executions and cancellations jointly. We consider three general classes of models: (a) Weibull proportional hazard model with independent competing risks, (b) the Accelerated Failure Time (AFT) model, and (c) Gamma Frailty model. Weibull is the most popular model specification in duration analysis, AFT is used to benchmark against the findings of Lo *et. al* (2002), and finally we consider the case where the two competing risks, execution and cancellation, depend on each other through an unobserved latent variable (frailty). Frailty allows us to control for the unobserved heterogeneity of limit orders and market conditions, and is an important econometric issue. Unobserved heterogeneity refers to the factors that are observed by investors placing the orders but unobserved (or uncaptured) by the econometrician. In general, failure to control unobserved heterogeneity in empirical research could severely bias the estimates for the parameters of interest and yield misleading conclusions and strategy recommendations (Lancaster, 1990). We introduce diagnostic plots as a useful method of verifying the goodness-of-fit for all three models.

Last but not least, our empirical data covers a very recent period in which we have Regulation NMS, decimal pricing, and even subpenny pricing for some stocks. Chung *et al.* (2004) and Chakravarty *et al.* (2004) studied the impact of decimalization on transaction costs, market quality and liquidity, and found that quoted depth as well as the quoted and effective bid-ask spreads declined significantly following decimalization.

The rest of the paper is organized as follows. Section II presents the econometric model. We introduce the data, explanatory variables, and provide their summary statistics in Section III. Section IV discusses empirical results and their implications. Section V describes goodness-of-fit diagnostics of the alternative models. A brief summary and an outline for future research conclude this paper in Section VI. The technical derivations are relegated to the Appendix.

II. The Econometric Model

To develop an econometric model for the times to limit order execution and cancellation, it is important not only to distinguish between these two causes of limit order termination, but also incorporate all observed characteristics of the limit order and capture the influence of market conditions at the time of the limit order submission. We accomplish this through the application of a well-known statistical technique – competing risks analysis.

Competing risks analysis is typically applied in situations where we have multiple failure types. A prototypical competing risks situation was first considered in the eighteenth century when small pox vaccination was discovered and popularized.[3] For many years, competing risks models have been a popular tool for analyses of failure time data in various disciplines, such as biostatistics, medicine, and engineering. Recently the competing risks model found numerous applications in actuarial science, criminal justice, economics, finance, and many other areas. Following the seminal papers of Heckman and Honoré (1989), Han and Hausman (1990), Sueyoshi (1992), and McCall (1996), competing risks models became popular in the economic analysis of time to an economic event, where the duration intervals can have multiple causes of termination. For example, an unemployment spell may end due to transition out of unemployment into either a new job or recall. Similarly, a mortgage loan contract can terminate because of the homeowner's default or prepayment of the loan.[4]

In the context of duration analysis of limit orders, we intend to capture the effects of market-wide conditions and order-specific characteristics, such as the level of market activity, stock price volatility, limit order price, limit order size, bid and ask quotes, market depth, and other covariates, on the survival probability of limit orders. For different risks of failure, these variables may have different effects, in magnitude and direction, on the risk-specific limit order duration. There are several empirical papers on limit orders that extensively use a single risk analysis, but little work has been done in a multiple risks framework.

The idea behind our econometric model can be best explained using the latent failure time approach. Often one wishes to analyze the failure time data where one of the several mutually exclusive causes (or types) of failure (i.e. competing risks) is assigned as the reason for the failure.[5] Since in most real life situations one cannot observe each of the latent causes of failure separately (e.g., the different possible causes of death; types of recidivism among released prisoners; possible reasons for ending a spell of unemployment), each individual survival time can be interpreted as if it represents a component of an unobserved random vector of survival times associated with different causes of failure (risks). The failure can happen because of any of these competing risks. The system is said to have failed because of the cause (or type) that happens to *realize first*. The lack of observability of the underlying paths of each component of the random vector of failure times and the posterior knowledge of the cause that actually triggers the failure justify the term "latent" duration. So, one can only observe the shortest time to failure and the cause (or type) of that failure.

[3] Daniel Bernoulli (1766) posed the question: "How much would the mortality be reduced or expected life be increased if the risk of death due to small pox is totally eliminated, the other risks persisting as before?"

[4] Ciochetti *et al.* (2003) applies the proportional hazard model with competing risks to the analysis of time to termination of commercial mortgage contacts.

[5] In some studies more complicated failure patterns may emerge. In the current framework, such patterns may be handled by defining additional failure types. For example, in our application to the limit order market, a limit order might be partially executed, with the remainder of the order staying on the book and being cancelled eventually.

Formally, let $T_1, T_2, \ldots, T_m$ denote the latent failure times of an individual subjected to m competing risks. What is actually observed is the time to failure $T = \min(T_1, T_2, \ldots, T_m)$ and the cause of failure $J = \arg\min_{j=1,2,\ldots,m} T_j$.[6] Denote by δ the right-censoring indicator ($\delta = 0$ if censored, $\delta = 1$ if uncensored) and denote by T_c the right-censoring time.[7] Often it is also useful to assume that the sample includes only the orders that have survived by time t_0 since the beginning of the duration episode, which implies that the observed times to failure are larger than t_0. In summary, observation for the ith individual included in the sample is either in the form $(t_i, j_i, \delta_i = 1)$ or $(t_i, \delta_i = 0)$, where $t_i \in (t_0; t_{ci}]$ is the realized time to failure in the uncensored case, and $t_i = t_{ci}$ is the censoring time if observation i is censored.

In our context, for each uncensored observation of a limit order, its termination can be triggered by two mutually exclusive competing risks: execution and cancellation.[8] We observe the duration to one and only one of those two causes, whichever occurs first. Hence, the observed survival time is $T = \min(T_1, T_2)$, where (T_1, T_2) is the random vector of the two underlying failure times of the limit order. In addition, we observe the cause of the termination J, which can be execution ($J = 1$) or cancellation ($J = 2$). As a result, the random vector (T, J) gives us the observable part of our data. Let $T > t_0$ and assume that the random vector (T_1, T_2) has a well defined absolutely continuous distribution function. Then the joint survival function

$$S(t_1, t_2 | t_0) = \Pr\{T_1 > t_1, T_2 > t_2 | T_1 > t_0, T_2 > t_0\} \tag{1}$$

is defined as the joint probability that the limit order surviving by time t_0 will not be terminated due to cause 1 (not executed) until time $t_1 > t_0$ and will not be terminated due to cause 2 (not cancelled) until time $t_2 > t_0$.

The sample censoring is said to exist when certain subsets of the population cannot be sampled, but the econometrician either knows or can consistently estimate the probability of not sampling this subset of the population. Here, censored observations may correspond to

[6] Random variables $T_1, T_2, \ldots, T_m$ are assumed to have continuous distribution, so that the cause of failure J is uniquely defined.

[7] The realizations of random variable T_c can vary across limit orders and generally depend on the market conditions. However, for day limit orders considered in this paper it is useful to assume that the right-censoring occurs at 4 p.m. EST or 10 minutes after the limit order submission, whichever comes first. Since the time of such right-censoring is known in advance, this censoring scheme can be considered deterministic. Note that although INET operates after the major exchanges close, we do not consider the extended hours in our analysis since market quality in the after hours is different from the regular trading hours (Barclay and Hendershott, 2003).

[8] In this paper we focus on the bivariate case, and assume that there are only two distinct causes of limit order termination. Generalization to the multivariate case is conceptually straightforward. Implementation of the multivariate competing risks model, possibly including revised and resubmitted orders, is left for future analysis.

the limit orders staying on the book until the end of the analyzed time interval or duration episode without being executed or cancelled. Even though neither execution nor cancellation events can be observed for such limit orders within the duration episode, partial information is still available. Specifically, it is known that limit order i survived by the censoring time t_{ci} corresponding to the duration between the time of limit order submission and the end of the analyzed duration episode. The probability of right-censoring for limit order i can then be expressed in terms of its joint survival function as follows

$$\Pr\{\delta_i = 0 | T_{1i} > t_0, T_{2i} > t_0\} = S\left(t_{ci}, t_{ci} | t_0\right). \tag{2}$$

Assuming in the above setup that the censoring rule is defined exogenously, the log-likelihood function takes the form[9]

$$\ln L = \prod_{i=1}^{n} \left[(1 - \delta_i) \ln S\left(t_{ci}, t_{ci} | t_0\right) + \delta_i \ln f\left(t_i, j_i | t_0\right)\right], \tag{3}$$

where $\delta_i \in \{0, 1\}$, $t_i \in (t_0, t_{ci}]$, $j_i \in \{1, 2\}$ are the sample observations indexed by $i = 1, 2, \ldots, n$, and the risk-specific density function $f\left(\cdot, \cdot | t_0\right)$ is given by

$$f\left(t, j | t_0\right) = -\left.\frac{\partial S\left(t_1, t_2 | t_0\right)}{\partial t_j}\right|_{t_1 = t_2 = t}, \quad j = 1, 2. \tag{4}$$

It can be clearly seen from the expression (3) that to carry out the likelihood-based estimation one only needs to specify the bivariate survival function $S\left(t_1, t_2 | t_0\right)$. Equivalently, one can specify $F\left(t_1, t_2 | t_0\right)$, the joint distribution function of (T_1, T_2) for limit orders surviving at least t_0 seconds since the onset of the duration episode, or the joint density function $f\left(t_1, t_2 | t_0\right)$. Indeed, the values of $S\left(t_1, t_2 | t_0\right)$ and $F\left(t_1, t_2 | t_0\right)$ are linked by the following relationship[10]

$$S\left(t_1, t_2 | t_0\right) = F(t_c, t_c | t_0) - F\left(t_1, t_c | t_0\right) - F(t_c, t_2 | t_0) + F\left(t_1, t_2 | t_0\right), \quad (t_1, t_2) \in (t_0, t_{ci}] \times (t_0, t_{ci}]. \tag{5}$$

In the analysis of failure time data, researchers often prefer to specify the cumulative hazard rate $H(t_1, t_2 | t_0) = -\ln(S(t_1, t_2 | t_0))$, providing yet another way to characterize the underlying distribution of the bivariate failure times. Often there are compelling reasons to model the cumulative hazard rate directly, as it may be more convenient to think of the limit order as a survivor subjected at any moment of its lifetime to instantaneous risks, characterized by partial derivatives of the joint cumulative hazard function. Formally, the

[9] See the Appendix for derivation of formula (3).
[10] See the Appendix for derivation of formula (5).

cause-specific hazard functions are defined for $t > t_0$ as follows

$$h_j(t|t_0) = \left.\frac{\partial H(t_1,t_2|t_0)}{\partial t_j}\right|_{t_1=t_2=t} = \lim_{\triangle t\to 0}\frac{P(t \le T_j < t+\triangle t \,|\, T_j \ge t > t_0)}{\triangle t}, \quad j=1,2. \tag{6}$$

In other words, the cause-specific hazard rates $h_j(t|t_0)$ $(j = 1, 2)$ can be viewed as the instantaneous rates of type j failure at time t, $t \in (t_0; t_{ci})$.

In the special case where random variables T_1 and T_2 are independent, we have

$$H(t_1,t_2|t_0) = \int_{t_0}^{t_1} h_1(\tau|t_0)\, d\tau + \int_{t_0}^{t_2} h_2(\tau|t_0)\, d\tau, \tag{7}$$

and the following relationship between $S(t_1, t_2|t_0)$ and $h_1(t|t_0)$ and $h_2(t|t_0)$ holds true[11]

$$S(t_1,t_2|t_0) = \exp\left[-\left(\int_{t_0}^{t_1} h_1(\tau|t_0)\, d\tau + \int_{t_0}^{t_2} h_2(\tau|t_0)\, d\tau\right)\right]. \tag{8}$$

Therefore, in the case of independent competing risks, it would be sufficient to specify the two cause-specific hazard functions $h_j(t)$, $j = 1, 2$.[12]

Finally, it must be emphasized that the above discussion can be reformulated in terms of the conditional joint distribution function $F(t_1, t_2|t_0, \mathbf{z})$, conditional joint survival functions $S(t_1, t_2|t_0, \mathbf{z})$, conditional joint hazard rate $H(t_1, t_2|t_0, \mathbf{z}) = -\ln(S(t_1, t_2|t_0, \mathbf{z}))$, as well as conditional risk-specific densities $f(t, j|t_0, \mathbf{z})$ or risk-specific hazard rates $h_j(t|t_0, \mathbf{z})$, $j = 1, 2$, given the vector of exogenous covariates $\mathbf{Z} = \mathbf{z}$. Components of the covariate vector $\mathbf{z}$ may include attributes of the limit order such as its size and transparency, as well as the variables characterizing the market conditions at the time of limit order submission.

III. Data

We use a sample of four randomly selected stocks – American Capital Strategies Ltd. (ACAS), Associated BanCorp (ASBC), Imclone Systems Inc. (IMCL), and Career Education Corp (CECO), – and the stock of Intel Corp (INTC), which is one of the most liquid equities traded on US markets. The sample period is July–December, 2005. We estimate and test our competing risk models using order book data for these five stocks from INET. INET is an automated limit order platform for trading equities. Thus, unlike the NYSE

[11] See the Appendix for derivation of formula (8).

[12] Later in this paper we will relax the conditional independence assumption for the competing risks by specifying the unobserved factor V affecting the joint hazard rate multiplicatively. It will be shown that the survival function obtained by integrating V out of the formula will imply dependence between the latent durations of the competing risks.

or NASDAQ, trading on INET is completely order driven. INET is open when the U.S. equity markets are open, and generally accepts orders between 7 a.m. and 8 p.m. EST. Only broker-dealers can submit orders to INET, and the only type of order allowed is limit order, which can be either open for display or hidden. A trader who wants an immediate INET execution can place a marketable limit order that meets or crosses the best price on the opposite side of the limit order book. A trader who is prepared to wait and deal with uncertainty of INET execution can place a non-marketable limit order that does not hit the best price instantaneously available on the opposite side of the limit order book.

Upon receiving an order and performing a series of checks to establish its validity, the INET trading system scans its limit order book to determine if a matching order is present in the system. If a matching order is found, the incoming order is executed immediately. If a matching order for the newly arrived display order does not exist, the display order is placed on the limit order book and remains visible until a matching order is received or until the display order originally submitted to the system is cancelled. If the newly arrived order was entered as a non-display (hidden) order, and a matching order is not available immediately, the non-display order is also placed on the electronic limit order book but remains hidden from view by other traders. All unmatched orders are automatically cleared from the book at the end of the trading day.

The INET electronic trading platform matches incoming orders with existing orders in the book based on the following priorities:

(a) Price: the limit order price specified at the time of order arrival,

(b) Display: non-display (hidden) orders have lower execution priority than display orders with the same limit order price, and

(c) Time: the exact time of limit order arrival (in milliseconds).

The information from the ITCH® database that is relevant for our empirical work is as follows: the limit order reference number, the limit price, size, time stamps (in milliseconds) at order entry, execution information (partial or complete, as can be inferred by comparing the order reference numbers), cancellation (partial or complete, as can be inferred by comparing the order reference numbers). We follow Odders-White (2000) in inferring the buy-sell direction of orders by assuming that the initiator of a transaction is the investor who places his or her order last, chronologically. To construct the National Best Bid and Offer (NBBO) quotes, we use quote data from the NYSE Trade and Quote (TAQ) database.

A. Data Cleaning

To reduce the influence of outliers and data errors, we apply the following filters to the raw ITCH® data:

(1) We confine our analysis to buy and sell limit orders submitted between 10 a.m. to 4 p.m. EST. The orders that are not executed or cancelled before 4 p.m. EST are treated as right-censored observations. The beginning time of 10 a.m. EST is chosen because some of the covariates that we are going to use are based on the previous hour's limit order book activity. Since the dynamics of limit order executions and cancellations are believed to be driven by a different set of factors for orders submitted after 4 p.m. EST (the market closing time for major US-based exchanges), such limit orders are also not covered by the analysis in the present paper.[13]

(2) For each submitted limit order, we classify its termination event as execution or cancellation and compute its failure time. We treat all partially executed limit orders as executed orders, no matter whether they are subsequently fully executed or cancelled after one or several partial executions. Therefore, the failure time of a limit order is defined as the time to its first fill (partial or complete) or, if no executions are reported, to its cancellation. The limit orders that are not executed and not cancelled within 10 minutes after their submission are treated as right-censored observations.

(3) We delete all orders with limit price more than \$0.25 away from the bid-ask mid-quote at the time of submission,[14] since we believe that such distant limit orders have a different set of factors driving their execution and cancellation dynamics. The median distance between the mid-quote and limit order price for our sample is \$0.025, while the 99% percentile of this variable is \$0.21. Therefore, limit orders with the distance between mid-quote and the limit price larger than \$0.25 are quite rare and can be treated as outliers that should be approached with a separate model.

(4) All orders with durations less than or equal to two seconds are excluded. It should be emphasized that over 90% of limit orders in our data set are terminated with cancellations and most of these cancellations occur within two seconds or less since the time of limit order submission. We exclude such *fleeting* limit orders[15] from our sample and set $t_0 = 2$ seconds in all formulas of Section II used in the subsequent analysis. Executions and cancellations that occur shortly after submission of a limit order require a separate model since their dynamics are likely to be driven by factors different from those for the limit orders surviving at least two seconds.

[13] See Barclay and Hendershott (2003) for a discussion of after-hours trading.

[14] The distance between the limit order price and the bid-ask mid-quote is captured by the covariate $MQLP$ defined below. The bid-ask mid-quote is calculated from the NBBO quotes.

[15] Hasbrouck and Saar (2004) provide more information on fleeting limit orders.

B. Covariates

The covariates (exogenous or pre-determined explanatory variables) are chosen to capture the effect of limit order characteristics and market conditions prevailing at the time of order arrival on the time-to-execution and time-to-cancellation of the limit order. Therefore, the dependence between execution and cancellation risks is partially captured by the covariates included in the model. The residual dependence between the two risks is due to the unobserved factors, which can be viewed as the covariates left out of the model.

Define P_l as the limit order price, P_b and P_o as the NBBO quotes, $P_q \equiv \frac{1}{2}(P_b + P_o)$ as the NBBO mid-quote, S_b and S_o as the number of round lots (100 shares) available at the NBBO quotes, S_l as the limit order size measured in round lots, and $BSID$ as the buy/sell indicator of the previous transaction. Similar to Lo *et al.* (2002), in the buy limit order models we use the following covariates that are all measured at the time of limit order submission[16]

$$
\begin{aligned}
MQLP &= 100(P_q - P_l), \\
BSID &= \begin{cases} 1 & \text{if last trade was a sell trade,} \\ -1 & \text{if last trade was a buy trade,} \end{cases} \\
MKD1 &= \begin{cases} \ln(S_b)\,(1 + 100(P_b - P_l)) & \text{if } P_l \leq P_b, \\ 0 & \text{if } P_l > P_b, \end{cases} \\
MKD2 &= \begin{cases} \ln(S_o)\,/(1 + 100(P_o - P_l)) & \text{if } P_o \geq P_l, \\ \ln(S_o) & \text{if } P_o < P_l, \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
SZSD &= \begin{cases} \ln(S_l)\,(1 + 100(P_o - P_l)) & \text{if } P_o > P_l, \\ \ln(S_l - S_o) & \text{if } P_o = P_l \text{ and } S_l > S_o, \\ 0 & \text{otherwise,} \end{cases} \\
STKV &= \frac{\#\text{ of trades in the last half-hour}}{\#\text{ of trades in the last hour}}, \\
TURN &= \ln(\#\text{ of trades in the last hour}), \qquad (9)
\end{aligned}
$$

Three of the above covariates are redefined for the sell limit-order models in order to retain uniformity in the underlying interpretation of these variables. The redefined covariates are

[16] As discussed in Section VI, this paper applies only to the case of exogenous covariates that are assumed to be fixed or predetermined at the time of limit order submission. Extending the analysis to the case of time-varying and endogenous covariates is left for future research.

listed below:

$$
\begin{aligned}
MKD1 &= \begin{cases} \ln(S_o)\,(1+100(P_l-P_o)) & \text{if } P_l \geq P_o, \\ 0 & \text{if } P_l < P_o, \end{cases} \\
MKD2 &= \begin{cases} \ln(S_b)\,/(1+100(P_l-P_b)) & P_b \leq P_l, \\ \ln(S_b) & P_b > P_l, \end{cases} \\
SZSD &= \begin{cases} \ln(S_l)\,(1+100(P_l-P_b)) & P_l > P_b, \\ \ln(S_l-S_b) & P_l = P_b \text{ and } S_l > S_b, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{10}
$$

The covariates defined above attempt to capture the current state of the market and accommodate the dynamic nature of the marketplace. The variable $MQLP$ measures the distance between the limit order price and the NBBO mid-quote at the time of limit order submission. $BSID$ is an indicator of whether the prior transaction was buyer-initiated or seller-initiated, determined using the Odders-White (2000) chronology test. $BSID$ equals 1 (-1) if the last trade in TAQ database was a seller (buyer)-initiated trade. $MKD1$ is a proxy for the minimum number of shares that have a higher priority of execution. $MKD2$ measures liquidity on the opposite side of the market. $SZSD$ is a measure of liquidity demanded by the marketable limit order scaled by the distance between the limit order price and the best quote on the opposite side of the market. $STKV$ is a variable that attempts to capture the lower-frequency shifts in the absolute level of trading activity; it indirectly approximates the fluctuations of market volatility at the intraday frequencies comparable to common time horizons of limit order traders. $TURN$ is an absolute measure of past trading activity. As previously mentioned, some of these variables are created from the TAQ data set. In the process of merging ITCH® data and TAQ data, we assume that the time stamps in the two data sets are consistent with each other.[17]

C. Descriptive Statistics

Table 1 reports some descriptive statistics for the five stocks analyzed in this paper. It shows the average daily number of outcomes for buy and sell limit orders and some characteristics of daily trading activity in each stock. Note that the reported statistics are for the orders that survive at least two seconds since their submission. Each of the five stocks have well over half million observations (buy and sell limit orders) for the sample period (July–December 2005).[18] More than 90% of all orders get cancelled, and most execution and cancellation

[17] See Chakrabarty *et al.* (2005) for details on mapping ITCH® to TAQ times.

[18] Our sample contains 588,000 observations for ACAS, 685,000 observations for ASBC stock, almost 979,000 observations for IMCL stock, and more than 1.16 mln. observations for CECO stock. The INTL

events occur within the first several seconds of order arrival. The price averages for our sample securities range between \$25 and \$40 per share. More than half of all executions for the four moderately liquid stocks chosen for this study have small size (100 shares), even though the market depth available at the best bid and ask quotes is typically between 400 and 800 shares. For the most liquid stock, INTC, the median transaction size is 300 shares, and the median market depth available at the best bid and ask quotes is around $26,000$ shares. The average bid-ask spread for the four moderately liquid stocks is slightly higher than two ticks, which means that the competing limit order traders can typically offer a one tick price improvement without making their limit orders marketable. Two of the stocks (ACAS and ASBC) are low-volatility stocks, whereas two other stocks (IMCL and CECO) are high-volatility stocks. As expected, the realized volatility of the four chosen stocks tends to increase as the sampling frequency goes up from monthly to hourly, suggesting the presence of substantial microstructure noise. On the other hand, the realized volatility of the INTC stock is practically invariant to the sampling frequency, suggesting very little microstructure noise even at the hourly frequency.

Table 2 presents some descriptive statistics for durations and covariates used in the estimation of alternative competing risks models for the ACAS stock. Note that Table 2 and the subsequent Tables 3–6 report the estimation results for the single stock – ACAS – using the subsamples for two months only (July and December of 2005). We summarize the results for all sample stocks and time intervals in Table 7. The properties of durations for buy and sell limit orders appear to be very similar. For example, the median duration time for ACAS buy limit orders (13.66 seconds) is almost identical to the one for sell limit orders (13.64 seconds). As expected, the difference between the limit order price and the market mid-quote has different signs for buy and sell orders due to the definition of the covariate $MQLP$. The average distance between the limit order price and the mid-quote of the bid-ask spread is close to three ticks (\$0.03). The properties of other covariates are also very similar for buy and sell limit orders. Notice that most covariates are slightly (and some significantly) leptokurtic and skewed.

IV. Estimation Results and Interpretation of Parameters

A. Model Specification

As explained in Section II, it is sufficient for estimation of the competing risks model in the range of durations $t > t_0$ to specify its joint conditional survival function $S(t_1, t_2 | t_0, \mathbf{z})$ used in the expression of the log-likelihood function (3). Equivalently, one can start with the

stock stands out with more than 10.64 mln. observations.

specification of the joint conditional distribution function $F(t_1, t_2|t_0, \mathbf{z})$ or joint conditional cumulative hazard rate $H(t_1, t_2|t_0, \mathbf{z})$, and then derive the function $S(t_1, t_2|t_0, \mathbf{z})$.[19] In this section, we offer three alternative specifications of the conditional survival function that are subsequently used in our empirical analysis. First we discuss two specifications where the underlying risk-specific durations T_1 and T_2 are assumed to be conditionally independent, and then turn to our last specification where the risk-specific duration T_1 and T_2 depend on a common unobserved risk factor.

A.1. Generalized Gamma AFT Model with Independent Risks

We adopt the accelerated failure time (AFT) specification of Lo *et al.* (2002)

$$T_j = T_{0j} \exp(\mathbf{z}'\boldsymbol{\beta}_j), \quad j = 1, 2, \tag{11}$$

where T_{0j} is the risk-specific baseline failure time, $\mathbf{z}$ is a vector of covariates that capture market-wide conditions and limit order characteristics, $\boldsymbol{\beta}_j$ is a risk-specific parameter vector. The distribution of T_{0j} is called the baseline distribution. The risk-specific durations T_j are modeled as scaled transformations of the baseline failure times T_{0j}, where the covariates and the parameter values determine the degree of scaling. If we assume that the baseline failure times T_{0j} have generalized gamma distribution with shape parameters κ_j and p_j, then the risk-specific conditional survival functions for the risk-specific durations $T_j > t_0$ $(j = 1.2)$ are given by

$$S_j(t|t_0, \mathbf{z}; \boldsymbol{\omega}_j) = \frac{\mathbf{1}\{p_j > 0\} \cdot \Gamma(\kappa_j) - \text{sign}(p_j) \cdot \Gamma(\kappa_j, (t \exp(-\mathbf{z}'\boldsymbol{\beta}_j))^{p_j}\kappa_j)}{\mathbf{1}\{p_j > 0\} \cdot \Gamma(\kappa_j) - \text{sign}(p_j) \cdot \Gamma(\kappa_j, (t_0 \exp(-\mathbf{z}'\boldsymbol{\beta}_j))^{p_j}\kappa_j)}, \tag{12}$$

where $\boldsymbol{\omega}_j=(\kappa_j, p_j, \boldsymbol{\beta}_j')'$ is the parameter vector, $\mathbf{1}\{\cdot\}$ is the indicator function which takes the value of 1 if the statement in parentheses is valid and 0 otherwise, $\text{sign}(\cdot)$ is the sign function which takes the value of 1 if its argument is positive and -1 otherwise,

$$\Gamma(a, x) = \int_0^x y^{a-1} e^{-y} dy \tag{13}$$

is the incomplete gamma function, and $\Gamma(a) \equiv \Gamma(a, \infty)$ is the complete gamma function. The corresponding density functions for the risk-specific conditional durations T_j are positive

[19] In fact, it is sufficient to specify any of these functions in a small neighborhood of the main diagonal $t_1 = t_2$ on the latent duration space.

on the support $t > t_0$ and have the following form

$$f_j(t|t_0,\mathbf{z};\boldsymbol{\omega}_j) = \frac{\exp(-\mathbf{z}'\boldsymbol{\beta}_j)\,|p_j|\,\kappa_j^{\kappa_j}(t\exp(-\mathbf{z}'\boldsymbol{\beta}_j))^{p_j\kappa_j-1}\exp(-(t\exp(-\mathbf{z}'\boldsymbol{\beta}_j))^{p_j}\kappa_j)}{\mathbf{1}\{p_j>0\}\cdot\Gamma(\kappa_j)-\mathrm{sign}(p_j)\cdot\Gamma(\kappa_j,(t_0\exp(-\mathbf{z}'\boldsymbol{\beta}_j))^{p_j}\kappa_j)}. \quad (14)$$

The generalized gamma distribution is chosen since it nests a number of other popular distributions (gamma, Weibull, log-normal) as special cases. Since T_1 and T_2 are assumed to be independent, given the observed covariates $\mathbf{z}$ and starting time t_0, the joint conditional survival function $S(t_1,t_2|t_0,\mathbf{z})$ is obtained as

$$S(t_1,t_2|t_0,\mathbf{z};\boldsymbol{\omega}_1,\boldsymbol{\omega}_2) = S_1(t|t_0,\mathbf{z};\boldsymbol{\omega}_1)\,S_2(t|t_0,\mathbf{z};\boldsymbol{\omega}_2)\,, \quad t_1 > t_0,\ t_2 > t_0, \quad (15)$$

where $S_j(t|t_0,\mathbf{z};\boldsymbol{\omega}_j)$, $j = 1,2$, are the risk-specific conditional survival functions defined by (12).

A.2. Weibull and Cox Proportional Hazard Models with Independent Risks

Alternatively, if we pursue the hazard function approach, a popular fully parametric specification for the risk-specific conditional hazard rates is provided by the Weibull proportional hazard model. In this specification, it is assumed that the risk-specific hazard rates $h_j(t|t_0,\mathbf{z})$ $(j = 1,2)$ are conditionally independent of each other at all times $t > t_0$, given the observed vector of covariates $\mathbf{z}$, and have the form

$$h_j(t|t_0,\mathbf{z};\boldsymbol{\theta}_j) = \gamma_j t^{\gamma_j-1}\exp(\mathbf{z}'\boldsymbol{\beta}_j), \quad t > t_0, \quad (16)$$

where $\boldsymbol{\theta}_j = (\gamma_j,\boldsymbol{\beta}_j')'$ are the model parameters $(j = 1,2)$.

In this setup, the joint survival function for risk-specific durations T_1 and T_2 conditional on the observed covariates $\mathbf{z}$ and starting time t_0 can be obtained as follows[20]

$$S(t_1,t_2|t_0,\mathbf{z};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \exp\left[-\left((t_1^{\gamma_1}-t_0^{\gamma_1})\exp(\mathbf{z}'\boldsymbol{\beta}_1)+(t_2^{\gamma_2}-t_0^{\gamma_2})\exp(\mathbf{z}'\boldsymbol{\beta}_2)\right)\right]. \quad (17)$$

For estimation, the expression (17) is plugged into the expression (3) of the log-likelihood function.

The Weibull proportional hazard specification (16) and (17) is a fully parametric model. The parameter estimates based on this specification are efficient if the model is correct but would be biased if the model is misspecified. If the assumption that T_1 and T_2 are conditionally independent given the observed covariates $\mathbf{z}$ and starting time t_0 is maintained, one

[20] Derivation of formula (17) is given in the Appendix.

can use the Cox proportional hazard model, which is robust to the Weibull baseline hazard misspecification. The Cox proportional hazard model, due to Cox (1972), has a long history of use in medical statistics and biostatistics, and it also gained popularity in economics and finance. Similar to the Weibull proportional hazard model, the Cox proportional hazard model is formulated for the risk-specific conditional hazard rates. It assumes that the risk-specific conditional hazard rates at time $t > t_0$ are given by

$$h_j(t|t_0, \mathbf{z}, \boldsymbol{\beta}_j) = h_{0j}(t|t_0) \exp(\mathbf{z}' \boldsymbol{\beta}_j), \quad t > t_0, \ j = 1, 2, \tag{18}$$

where $\mathbf{z}$ is the vector of observed covariates and $h_{0j}(t|t_0)$ are the covariate-free risk-specific baseline hazard rates. The Cox proportional hazard model is semiparametric since no assumptions (other than mild regularity conditions) are made about the shape of the baseline hazard rates $h_{0j}(t|t_0)$.

Cox (1972) shows that inference on the covariate effects $\boldsymbol{\beta}_j$ $(j = 1, 2)$ in the Cox proportional hazard model can be based on the partial likelihood function instead of the full likelihood derived in Section II. Moreover, since the Cox proportional hazard competing risks model includes, as a special case, the proportional hazard model with Weibull independent risks, we can compare the estimates of covariate coefficients coming from the two models. If the coefficient vectors $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ coming from the two models are close to each other, the Weibull proportional hazard assumption is justified (at least indirectly) and can be accepted as a good working hypothesis.[21] To this end, we also estimate the Cox proportional hazard model with conditionally independent competing risks and report our estimation results in Table 5.

A.3. Weibull Proportional Hazard Model with Gamma Frailty

In the last two subsections, we maintained the assumption that T_1 and T_2 are conditionally independent given the observed covariates $\mathbf{z}$ and the starting time t_0. However, allowing T_1 and T_2 to be dependent given the observed covariates often may be a more realistic assumption. For example, if there are some economic factors that could plausibly affect the times to limit order execution and cancellation but cannot be incorporated in the model as components of the covariate vector $\mathbf{z}$, then the risk-specific durations T_1 and T_2 would be dependent random variables, even after their dependence on the observed covariates is controlled. Our last model specification allows for such a possibility.

In this specification, we assume that the risk-specific hazard rates at time t are independent of each other, provided that not only the covariate vector $\mathbf{z}$ but also the value of an

[21] A formal specification test similar to the standard Hausman test (Hausman, 1978) can also be applied.

additional latent factor (frailty) V are observed. We assume the following specification for the risk-specific hazard rates

$$h_j(t|\mathbf{z},v;\boldsymbol{\theta}_j) = \gamma_j t^{\gamma_j - 1} v \exp(\mathbf{z}'\boldsymbol{\beta}_j), \quad j = 1, 2, \tag{19}$$

where $\mathbf{z}$ is the vector of observed covariates, v is the value of latent factor V realized at the beginning of the episode (at time $t = 0$) but unobserved by the econometrician, and $\boldsymbol{\theta}_j = (\gamma_j, \boldsymbol{\beta}_j')'$ are the model parameters as in the Weibull proportional hazard model (16). As a consequence of dependence on the common factor V, the risk-specific hazard rates $h_j(t|\mathbf{z};\boldsymbol{\theta}_j)$, $t > t_0$, $j = 1, 2$, will be dependent given the observed covariates $\mathbf{z}$ only, and hence the risk-specific durations T_1 and T_2 will also be dependent conditional on the observed covariates and the starting time t_0.

If we assume that random variable V has gamma distribution with mean $\tau\alpha$ and variance $\tau^2\alpha$, characterized by the density

$$f(v|\boldsymbol{\varphi}) = \frac{1}{\tau\Gamma(\alpha)} \left(\frac{v}{\tau}\right)^{\alpha - 1} \exp\left(-\frac{v}{\tau}\right), \tag{20}$$

where $\boldsymbol{\varphi} = (\tau, \alpha)'$ is the vector of its scale and shape parameters, then the joint survival function for T_1 and T_2 conditional only on the observed covariate vector $\mathbf{z}$ (given that the limit order survived until time t_0 since its submission) can be obtained as follows[22]

$$S(t_1, t_2|t_0, \mathbf{z}; \boldsymbol{\Theta}) = \left(\frac{1 + \tau(t_0^{\gamma_1} \exp(\mathbf{z}'\boldsymbol{\beta}_1) + t_0^{\gamma_2} \exp(\mathbf{z}'\boldsymbol{\beta}_2))}{1 + \tau(t_1^{\gamma_1} \exp(\mathbf{z}'\boldsymbol{\beta}_1) + t_2^{\gamma_2} \exp(\mathbf{z}'\boldsymbol{\beta}_2))}\right)^{\alpha}, \quad t > t_0, \tag{21}$$

where $\boldsymbol{\Theta} = \left(\alpha, \gamma_1, \gamma_2, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$ is the vector of model parameters.[23] As usual, the expression (21) can be plugged into the formula (3) to obtain the maximum likelihood estimates of the model parameters.

B. Estimation Results

Each of the models discussed in the previous section is estimated twice – once for buy orders and once for sell orders. In our interpretation of the estimation results, we distinguish between two modeling schemes. In the accelerated failure time (AFT) framework, where we model the risk-specific conditional durations directly, positive estimated values of covariate

[22] Derivation of formula (21) can be found in the Appendix.

[23] It can be inferred from the expression of log-likelihood function (3) based on formulas (19) and (21) that parameters τ and α cannot be identified simultaneously. Imposing a restriction on parameter τ (for instance, $\tau = 1$) makes the competing risks model based on the gamma frailty just-identified and leads to a well-defined maximum of the log-likelihood function.

effects imply that larger values of those covariates are associated with longer risk-specific durations. In the second approach, where we model the risk-specific conditional hazard rates, the interpretation of covariate effects is opposite. Positive estimated values of covariate effects imply that an increase in those covariates tends to increase the risk-specific conditional hazard rate and decrease the expected time to realization of the risk under consideration.

Estimation results based on July 2005 and December 2005 subsamples of limit orders for alternative competing risks model specifications are reported in Tables 3–6 for the ticker ACAS; summary results for the entire sample are provided in Table 7. The following discussion is based on the estimates for ACAS limit order data in July 2005. As emphasized in the discussion of Section III, only the orders submitted no further than \$0.25 away from the current NBBO mid-quote, between 10 a.m. and 4 p.m. EST are considered. In addition, orders with less than two-second durations are eliminated and all limit orders with durations larger than 10 minutes are assumed to be right-censored at 600 seconds. Similarly, any limit order that survives by 4 p.m. EST is assumed to be right-censored at the duration equal to the time interval between 4 p.m. and the time of this limit order submission. Additionally, as a robustness check, we report all estimation results based on the ACAS limit order data in December 2005. At the end of this section we briefly discuss robustness of our estimates across stocks and time periods.[24]

B.1. Generalized Gamma AFT Model with Independent Risks

Table 3 reports parameter estimates of the generalized gamma accelerated failure time model for buy and sell orders, along with their t-statistics. Note that the effect of covariate $MQLP$ associated for the execution risk of buy (sell) limit orders equals $+0.99$ (-1.14) and is highly significant, whereas the same covariate effect for the cancellation risk of buy (sell) limit orders is only $+0.05$ ($+0.01$) with a much smaller t-ratio. This demonstrates that the further away the limit order price is relative to the midquote level, the longer it takes to execute the order, and the sooner that order would be cancelled.

The positive sign ($+0.12$) on the covariate effect of $BSID$ for the buy limit order time to execution suggests that if the prior trade occurred below the current midquote ($BSID = -1$), then the smaller time to execution of that buy limit order is expected. The covariate effect of $BSID$ for the buy limit order cancellation risk is quantitatively small, which suggests that the high-frequency mean reversal effect, which is apparently quite substantial for the execution risk, is negligible for the cancellation risk of buy limit orders.

The positive coefficients ($+0.32$ and $+0.04$) on the covariate effects of $MKD1$ indicate that the time-to-execution and the time-to-cancellation of a buy limit order are both in-

[24] Detailed estimates for all five stocks are available upon request.

creasing with the amount of liquidity provided on the same side of the market with a higher price and time priority. The coefficient is much larger for execution risk, which leads to the tentative conclusion that the deeper market on the buy side would be associated with a higher percentage of cancellations (relative to executions) of the buy orders submitted at the inferior limit order prices.

On the other hand, the negative coefficient (-0.21) (the positive coefficient ($+0.01$)) on the covariate effect of $MKD2$ indicate that the time-to-execution (time-to-cancellation) of a buy limit order is decreasing (increasing) with the amount of liquidity provided on the opposite side of the market and with the distance between the buy limit order price and the best offer quote available on the opposite side. To put it differently, the increasing competition among the sellers implies shorter time-to-execution (longer time-to-cancellation) for the outstanding buy limit orders.

Since $SZSD$ is a measure of liquidity demanded by the buy limit order trader scaled by its price aggressiveness, we expect the coefficient of that variable on time-to-first-execution of the buy limit order to be negative. This is exactly the case as we observe the covariate effect for $SZSD$ associated with the execution risk to be negative (-0.05). Interestingly, the effect of $SZSD$ on time-to-cancellation is also negative, but its absolute value is much larger (-0.20) with a much larger t-ratio. This can be viewed as evidence of the fleeting character of aggressively priced buy limit orders; such orders are more likely to be executed but also more likely to be cancelled before the seller arrives.

The absolute values of coefficients corresponding to the last two covariates ($STKV$ and $TURN$) are much larger for the execution risk than for the cancellation risk. All coefficients are significantly negative, suggesting the negative effect of both relative and absolute levels of market activity on the time-to-execution and time-to-cancellation of buy limit orders. This indicates that the increase of trading activity causes more buy limit orders to be picked up by aggressive sellers and also induces the buy limit order traders to cancel their previously submitted limit orders more urgently.

The four rightmost columns of Table 3 report the parameter estimates of the generalized gamma accelerated failure time model for sell orders, along with their t-statistics. The effects of most covariates on the time-to-execution and time-to-cancellation of sell limit orders are in agreement with the effects of the corresponding covariates for buy limit orders.

B.2. Weibull and Cox Proportional Hazard Models with Independent Risks

The conclusions based on the estimates of the Weibull and Cox proportional hazard models (Tables 4 and 5) are generally in agreement with the interpretation of the estimated parameters of the generalized gamma AFT model. The difference between the two models is that the proportional hazard models focus on the hazard rates of competing risks while the generalized gamma AFT model focuses on the risk-specific durations, so the signs of the coefficients of all variables in the proportional hazard models are opposite to those in the generalized gamma AFT model for buy orders, as expected. Discrepancies between the signs on the covariate effects for sell limit orders are predominantly confined to the coefficients that appear to be small and economically insignificant.[25]

We emphasize once again that the coefficients on the variable $MQLP$ characterizing the price aggressiveness of the limit order are large, highly economically and statistically significant, and have opposite signs for buy (-0.97) and sell ($+1.14$) limit order executions. The effect of the price aggressiveness on the cancellation hazard rates are still significant but much smaller in magnitude for buy (-0.06) and sell ($+0.03$) limit orders. This clearly indicates that the hazard rates of buy and sell order executions and cancellations tend to increase with the limit price aggressiveness, but the execution risk is much more sensitive to changes in the limit order price. This also indicates that the instantaneous odds that an order will be terminated by cancellation rather than execution tend to increase for less aggressively priced orders.

Once again, we find that the effects of the past buy/sell indicator variable $BSID$ are only economically and statistically significant for buy order execution risks. The same qualitative conclusion generally applies to the effects of variables $MKD1$ and $MKD2$ on execution and cancellation hazards; the hazard rates of executions appear to be much more sensitive to the variation of those variables than the hazard rates of cancellations. Parlour (1998) shows that if the market depth is greater on the buy side (variable $MKD1$), then the probability of execution of a subsidiary limit buy order is smaller. However, if the market depth is thicker on the sell side (variable $MKD2$), then the probability of execution of a subsidiary limit order is greater. Our results provide some support for this prediction. The reasoning is as follows. Suppose a trader enters the market and observes that there are so many limit sell orders on the book that she becomes reluctant to submit a sell limit order since it has a low chance of being executed. Instead, this trader may jump the queue by bettering the current book or submitting a marketable sell order to cross the best opposing price. Thus, the execution probability of buy orders on the book will increase.

The only covariate whose variation has a substantial effect on the cancellation risk and negligible effect on the execution risk is $SZSD$, which measures the liquidity demanded by limit orders scaled by the limit order price aggressiveness. This can be viewed as an evidence

[25] Due to the large sample size, small and economically insignificant coefficients are often statistically significant. Therefore, one should be cautious about making inferences based on their t-statistics. Also see our discussion of the estimates for the competing risks model with gamma frailty below.

that, other things being the same, large limit orders tend to be cancelled much sooner than small limit orders. The size effect magnifies when a large limit order is submitted further away from the best quote on the opposite side of the market. In other words, small order traders are willing to wait longer than large order traders before cancelling their orders, especially if the limit order has been submitted close to the best quote on the opposite side of the market.

According to our model, the limit order size does not seem to affect the probability of its execution. This is likely to be a consequence of our treatment of partial executions (see our discussion at the beginning of Section III for details). We treat all executions – partial and complete – similarly, interpreting the time-to-first-fill of a limit order as its execution time. As a result, aggressive traders demanding only a small amount of liquidity, would still be filling the limit orders, no matter whether they are large or small, just taking as much liquidity as necessary and leaving the rest of the limit order unfilled.

Again, the magnitude of the coefficients on the last two covariates ($STKV$ and $TURN$) generally tends to be larger for execution risk than for cancellation risk. Both coefficients are significantly positive, suggesting the positive effect of relative and absolute levels of past trading activity on the hazard rates of executions and cancellations.

Finally, the estimates of γ_1 and γ_2 are both smaller than unity, indicating that the two competing risks hazard both decline as the limit order stays on the book longer. Since γ_1 tends to be much larger than γ_2, the baseline hazard rate tends to decline much faster for cancellation risk. We interpret this as evidence of the fleeting character of limit orders, which tend to be cancelled much more frequently early in their lifetimes. Surprisingly, this effect is substantial even after we excluded from the analysis the limit orders executed or cancelled within the first two seconds of their lifetime.

B.3. Weibull Proportional Hazard Model with Gamma Frailty

The estimation results for the Weibull competing risks model with gamma frailty are presented in Table 6. Since this model focuses on estimating the hazard rates, we expect the coefficients in Table 6 to have similar signs to the coefficients in Tables 4 and 5 discussed in subsection B.2. A thorough comparison of those tables reveals that most coefficients indeed tend to have similar signs, which means that the Weibull model with gamma frailty and the model with conditionally independent competing risks might have *qualitatively* similar implications. However, there are several important differences.

Before discussing those differences and similarities, note that the coefficients on the covariates in Table 6 should be interpreted as the sensitivities of hazard rates for cancellation and execution risks *conditional* on the observed covariates *and* an unobserved state variable (frailty), which affects both competing risks multiplicatively. This should be always kept in mind when the signs and magnitude of the coefficients in Table 6 are compared to those from Tables 4 and 5; the estimates reported in Tables 4 and 5 represent the sensitivities of execution and cancellation hazard rates *conditional on the observed covariates only.*

The coefficient on the variable of $MQLP$ in the hazard rates for execution of buy limit orders is significantly negative (-1.19) and larger than the corresponding coefficient (-0.97) in the model with independent Weibull competing risks. Similarly, the magnitude of the coefficient of this variable tends to increase in the gamma frailty model for cancellation rate of buy limit orders (-0.13) relative to the same coefficient in the model with independent competing risks (-0.06). Indeed, conditioning on the unobserved frailty parameter appears to reduce the noise, which in turn tends to boost the hazard rate sensitivities.

The negative signs of the coefficients corresponding to the covariate $BSID$ in the conditional hazard rates of execution and cancellation indicate that if the prior transaction has been on the buy side, a shorter time-to-execution (-0.16) and a slightly shorter time-to-cancellation (-0.04) of a buy limit order are expected. Our results are consistent with previous studies. For example, Lo *et al.* (2002) reports that if the prior transaction was seller-initiated, a longer time-to-execution is expected for buy orders.

The coefficients on the variable of $MKD1$ are significantly negative in the hazard functions of both execution and cancellation of buy limit orders. This indicates that a longer time-to-execution and a longer time-to-cancellation are expected if the proxy for the number of shares having higher priority to execution on the buy side is larger. In contrast, the coefficients on the variable of $MKD1$ tend to be less significantly negative in the hazard functions of sell order execution and tend to be closer to zero in the hazard functions of sell order cancellation. This provides some evidence that a large number of shares having higher priority to execution on the sell side does not tend to increase substantially the time-to-cancellation of a lower priority sell limit order, suggesting that sell limit order traders tend to be more patient.

On the other hand, the positive signs of the estimated coefficients of $MKD2$ in the models for buy and sell order execution ($+0.24$ and $+0.29$, respectively), indicate that the greater is the depth on the opposite side of the market, the shorter is the expected time-to-execution, and the effect is stronger for sell limit order execution. At the same time, the estimated coefficient of $MKD2$ for the cancellation risk is essentially close to zero for buy limit orders, and negative (-0.10) for sell limit orders, indicating that the large depth on the opposite side of the market tends to discourage sell limit order cancellations but does not seem to affect buy limit order cancellations.

In contrast to the competing risks model with conditionally independent Weibull competing risks, the coefficients of the variable $SZSD$ become significantly positive for *both* executions and cancellations of limit orders. This indicates that both execution and cancellation risks are increasing with the size of the limit order and the distance between its limit price and the best price on the opposite side. Still, the cancellation rates grow faster with the size of $SZSD$ than execution rates for both sell and buy limit orders suggesting that even though executions tend to be spurred by the larger limit order size, the cancellation rates of such limit orders also tend to increase, and they increase at somewhat higher rate, even after the common frailty factor is taken into consideration.

We also observe, once again, that the coefficients of the last two covariates ($STKV$ and $TURN$) are much larger for the execution risk than for the cancellation risk. All coefficients are significantly positive, suggesting the positive effect of both relative and absolute levels of market activity on the execution and cancellation hazard rates for buy and sell limit orders.

Finally, the estimates of γ_1 and γ_2 are both significantly larger than unity, which indicates that the hazard rates of the two competing risks (conditional on the covariates *and* an unobserved frailty parameter) are increasing, while the rate of increase appears to be much faster for the conditional baseline hazard rate of execution relative to the conditional baseline hazard rate of cancellation. The unobserved heterogeneity appears to be important, as it affects the size and, occasionally, the sign of the covariate effects. The unobserved covariates that affect proportionally the execution and cancellation hazard rates but are not included as explanatory variables appear to be a probable cause for dependence between the risks of limit order execution and cancellation.

B.4. Comparison Across Stocks and Time Periods

Table 7 summarizes the signs and significance of the estimated covariate effects based on the Weibull proportional hazard model with conditionally independent competing risks for buy and sell limit orders across the alternative stocks and time periods.[26] We estimated our model separately for buy and sell orders in each of the four moderately liquid stocks month-by-month for each month (July–December 2005). Because of the huge number of observations, our estimation for the liquid ticker INTC was performed using the data from a randomly selected business day of each month. Since all our estimates have similar qualitative interpretations, we focus our discussion in this subsection on the stability of the signs of covariate effects across time periods and the robustness of our results across different stocks.

Clearly, most of the covariates are robust in both direction and magnitude across the sample stocks. For example, effects of $MQLP$ is negative (positive) and significant for

[26] The numerical values of estimated coefficients for the alternative stocks and time periods can be requested from the authors.

all stocks and both execution and cancellation risk, for buy (sell) orders. Likewise, the coefficients of $BSID$, $MKD1$, $MKD2$, $TURN$, and $STKV$ are all significant and have expected and consistent direction for both buy and sell order executions, although the sign of these coefficients are more variable and less stable for cancellations.

The only covariate that changes the sign for both executions and cancellations is $SZSD$. We think that the reason this covariate changes sign across stocks and months reflects the asymmetric variation in the depth of the limit order book on the buy and sell sides across months, likely driven by stock-specific information events such as earnings announcements and other news arrivals. A similar explanation can be proposed for the relatively unstable signs of most covariates for the cancellation risk. The detailed exploration of factors underlying this instability is left for future research.

V. Goodness of Fit and Model Selection

We compare the goodness-of-fit of alternative models using the modified cumulative probability plots in combination with the numerical diagnostics (percentile statistics). Our approach parallels the method of Lo *et al.* (2002); however, we explicitly account for the presence of multiple risks and right-censoring.[27]

A. Cumulative Probability Plots for Competing Risks

The idea of probability plots can be described as follows. Denote by

$$\begin{aligned} Q_1(t|t_0,\mathbf{z}) &= \Pr\{T_1 < t, T_2 \geq T_1 | t > t_0, \mathbf{z}\}, \\ Q_2(t|t_0,\mathbf{z}) &= \Pr\{T_2 < t, T_1 \geq T_2 | t > t_0, \mathbf{z}\}, \end{aligned}$$

the true conditional incidence rates of the two competing risks defined for durations $t \in (t_0; t_c(\mathbf{z})]$, where $t_c(\mathbf{z})$ is the deterministic right-censoring time, which may be a function of market conditions $\mathbf{Z} = \mathbf{z}$ prevailing at the time of the limit order submission. For any vector of risk-specific durations (T_1, T_2) and auxiliary uniform $[0; 1]$ random variable S (independent of T_1, T_2, and $\mathbf{Z}$), we construct the random function

$$Q(T_1, T_2, S|t_0, \mathbf{z}) = \begin{cases} Q_1(T_1|t_0,\mathbf{z}) & \text{if } T_1 < \min(T_2, t_c(\mathbf{z})), \\ Q_2(T_2|t_0,\mathbf{z}) & \text{if } T_2 < \min(T_1, t_c(\mathbf{z})), \\ Q_c(t_0,\mathbf{z}) \cdot S & \text{if } t_c(\mathbf{z}) \leq \min(T_1, T_2), \end{cases} \tag{22}$$

[27]Without loss of generality, we can assume that all observations with $t_i \leq t_0$ have been eliminated from the empirical sample of durations $\{t_i\}_{i=1}^n$.

which must be uniformly distributed on $[0;1]$, where

$$Q_c(t_0, \mathbf{z}) = 1 - Q_1(t_c(\mathbf{z})|t_0, \mathbf{z}) - Q_2(t_c(\mathbf{z})|t_0, \mathbf{z})$$

is the conditional probability of right-censoring of the limit order given the information that the order has survived for t_0 seconds since its submission at the market conditions characterized by the covariate vector $\mathbf{z}$. The vector of exogenous covariates $\mathbf{Z} \overset{d}{\sim} \Phi_{\mathbf{Z}}(\cdot)$ and the auxiliary uniform $[0;1]$ random variable S can be integrated out of the expression (22) to obtain the random variable

$$Q(T_1, T_2|t_0) = \iint Q(T_1, T_2, s|t_0, \mathbf{z}) \cdot \mathbf{1}_{[0;1]}(s) ds d\Phi_{\mathbf{Z}}(\mathbf{z})$$

which must be uniformly distributed on $[0;1]$.

Now assume that we have a sample of data $\{(t_i, j_i, \delta_i, \mathbf{z}_i)\}_{i=1}^n$, where $t_i > t_0$, $j_i = 1$ or 2, $\delta_i = 0$ or 1, and $\mathbf{z}_i$ are the observed covariates. Augmenting this sample by an auxiliary sequence of uniformly distributed random numbers $\{s_i\}_{i=1}^n$, and transforming it into the sequence of

$$\pi_i = \pi(t_i, j_i, \delta_i, s_i|t_0, \mathbf{z}_i) = \begin{cases} Q_1(t_i|t_0, \mathbf{z}_i) & \text{if } j_i = 1 \text{ and } \delta_i = 1, \\ Q_2(t_i|t_0, \mathbf{z}_i) & \text{if } j_i = 2 \text{ and } \delta_i = 1, \\ Q_c(t_0, \mathbf{z}_i) \cdot s_i & \quad \text{if } \delta_i = 0, \end{cases} \tag{23}$$

should yield an approximately uniformly distributed random sample on $[0;1]$. If the sample of exogenous covariates $\{\mathbf{z}_i\}_{i=1}^n$ is "representative," so that its empirical distribution function $\widehat{\Phi}_{\mathbf{z}_1,\ldots,\mathbf{z}_n}(\mathbf{z})$ is close to the theoretical cdf $\Phi_{\mathbf{Z}}(\mathbf{z})$, then the sample of random numbers $\{\pi_i\}_{i=1}^n$ obtained by the formula (23) should be approximately uniformly distributed on $[0;1]$. As a result, a valid test of the hypothesis that the sequence of $\{\pi_i\}_{i=1}^n$ is drawn from a uniform $[0;1]$ distribution would serve as a test that $Q_1(t|t_0, \mathbf{z})$ and $Q_2(t|t_0, \mathbf{z})$ indeed provide the true functional forms of the conditional incidence rates for the competing risks in the interval of durations $t \in (t_0; t_c(\mathbf{z})]$.

In practice, the parametric functional forms of the conditional incidence rates $Q_1(t|t_0, \mathbf{z})$ and $Q_2(t|t_0, \mathbf{z})$ are unknown and must be estimated. We use the estimates $\widehat{\mathbf{\Theta}}$ of our model's unknown parameters to obtain proxies for the incidence rates $Q_1(t|t_0, \mathbf{z}_i; \widehat{\mathbf{\Theta}})$ and $Q_2(t|t_0, \mathbf{z}_i; \widehat{\mathbf{\Theta}})$. Then we plug those proxies into formula (23) to obtain the sequence of $\pi_i(\widehat{\mathbf{\Theta}})$ that could be subsequently used to construct a diagnostic plot of our model's goodness-of-fit. If the model is specified correctly, then the estimated conditional incidence rates $Q_1(t|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$ and $Q_2(t|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$ should be close to the true incidence rates $\pi_1(t|t_0, \mathbf{z})$ and $\pi_2(t|t_0, \mathbf{z})$, and the estimated conditional right-censoring probability $Q_c(t_0, \mathbf{z}; \widehat{\mathbf{\Theta}}) = 1 - Q_1(t_c(\mathbf{z})|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}}) - Q_2(t_c(\mathbf{z})|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$ should be close to the true conditional right-censoring probability $Q_c(t_0, \mathbf{z})$. Properties of the sequence of random numbers

$$\widehat{\pi}_i = \begin{cases} Q_1(t_i|t_0, \mathbf{z}_i; \widehat{\mathbf{\Theta}}) & \text{if } j_i = 1 \text{ and } \delta_i = 1, \\ Q_2(t_i|t_0, \mathbf{z}_i; \widehat{\mathbf{\Theta}}) & \text{if } j_i = 2 \text{ and } \delta_i = 1, \\ Q_c(t_0, \mathbf{z}_i; \widehat{\mathbf{\Theta}}) \cdot s_i & \text{if } \delta_i = 0, \end{cases} \tag{24}$$

should be similar to those of the sequence $\{\pi_i\}_{i=1}^n$ obtained by formula (23) from the exact expressions of the conditional rates $Q_1(t|t_0, \mathbf{z})$, $Q_2(t|t_0, \mathbf{z})$, and $Q_c(t_0, \mathbf{z})$. If the model is correctly specified, the sample $\{\widehat{\pi}_i\}_{i=1}^n$ would be approximately uniformly distributed on $[0; 1]$. Equivalently, the probability plot of the sample $\left\{\widehat{\pi}_{(i)}\right\}_{i=1}^n$, obtained from $\{\widehat{\pi}_i\}_{i=1}^n$ by lexicographic ordering (first by $\delta_i \in \{1, 0\}$, then by $j_i \in \{1, 2\}$, and finally by $t_i > t_0$), would be close to a straight line with a unit slope, and the sequence of the dynamic gaps ("scores") $\left\{\widehat{\pi}_{(i)} - \frac{i}{n+1}\right\}_{i=1}^n$ for all values of i would not deviate too far from the origin.[28] Unusually large positive (negative) slopes on the plot of "scores" $\left\{\widehat{\pi}_{(i)} - \frac{i}{n+1}\right\}_{i=1}^n$ against $\left\{\frac{i}{n+1}\right\}_{i=1}^n$ indicate the range of incidence rates (and, by implication, the range of durations and the type of events) where the model tends to overpredict (underpredict) the empirical frequencies of those events.

The same idea can be used to construct diagnostic plots for evaluation of alternative competing risks model performance out-of-sample. The only difference is that the estimates $\widehat{\mathbf{\Theta}}$ of the unknown parameters used to obtain the parametric functional forms of the conditional rates $Q_1(t|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$, $Q_2(t|t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$, and $Q_c(t_0, \mathbf{z}; \widehat{\mathbf{\Theta}})$ are obtained using the training sample, while the sample $\{(t_i, j_i, \delta_i, \mathbf{z}_i)\}_{i=1}^n$ used for construction of the random sequence $\{\widehat{\pi}_i\}_{i=1}^n$ must come out-of-sample, for example, using the limit order data for a different stock or a different time period.

Details of this general algorithm customized for the construction of diagnostic plots for the three alternative competing risks models of this paper appear in Appendix II.

B. Model Selection

Figure 1 displays the diagnostic cumulative probability plot based on the predicted risk-specific incidence rates for the Weibull proportional hazard model with independent competing risks of execution and cancellation for buy limit orders. Superimposed on this plot

[28] In the covariate-free case, the distribution of the random process $\left\{\widehat{\pi}_{([nr])} - r\right\}_{r \in [0;1]}$ must be identical to that of a Brownian bridge. In the more general case with covariates, the distribution of process $\left\{\widehat{\pi}_{([nr])} - r\right\}_{r \in [0;1]}$ would depend on the properties of covariates, even though it would still cluster around the origin if the model is correctly specified.

are similar plots for the gamma frailty model with Weibull baseline hazard functions and for the generalized gamma AFT model. The sample of limit order durations is left-censored at $t_0 = 2$ seconds and right-censored at $t_c = 600$ seconds or the duration corresponding to 4 p.m. EST, whichever occurs earlier. Among the three models, the Weibull proportional hazard with conditionally independent competing risks model demonstrates a much better overall goodness-of-fit performance relative to the other two models (generalized gamma AFT and Weibull PH model with gamma frailty). Similar plots for alternative competing risks models of sell limit order execution and cancellation support the same conclusion.[29]

Figures 2a and 2b provide an even more striking illustration as they show the goodness-of-fit for the risk-specific incidence rates as functions of duration since the limit order arrival (only the duration range between 2 and 120 seconds is shown on the plots). These plots give a more nuanced picture in support of the conclusion that the Weibull proportional hazard specification with conditionally independent competing risks predicts the empirical incidence rates of execution and cancellation events much better than the AFT model with the risk-specific durations modeled by the generalized gamma distribution. Since the generalized gamma AFT model severely overpredicts the risk of execution and underpredicts the risk of cancellation, the inference based on this model should be approached with caution and this model's predictions of limit order execution and cancellation events would generally be unreliable. Both Weibull proportional hazard models (with and without gamma frailty) generally perform much better in fitting the incidence rates of execution events, although the Weibull proportional hazard model with conditionally independent competing risks tends to perform slightly better.

In terms of fitting the incidence rates of cancellation events, the Weibull proportional hazard model with independent competing risks significantly outperforms not only the generalized gamma AFT model, but also the Weibull hazard model with gamma frailty. For example, the Weibull gamma frailty model predicts that a randomly selected buy limit order will be cancelled between 2 and 7 seconds after its submission with probability of approximately 25% (provided that it is not cancelled or executed prior to the two-second duration mark). In contrast, the empirical probability of buy order cancellation within the same time interval is slightly higher than 45% (As a result, the cumulative scrore differential at 7 seconds is -0.2 for this model in Figure 2b). The generalized gamma AFT model fares much worse, predicting only the 10% cancellation probability in the duration interval between 2 and 7 seconds after the limit order submission. The gap between predicted and empirical probabilities of cancellations for the generalized gamma AFT model tends to be even larger for longer durations.

[29] Those plots are available from the authors upon request.

In summary, the generalized gamma accelerated failure time model of competing risks fails miserably, severely overpredicting the execution risk and underpredicting the cancellation risk. The Weibull proportional hazard competing risks model with gamma frailty performs better for the risk of limit order execution, but still fails to adequately capture the cancellation risk. The Weibull proportional hazard model with conditionally independent competing risks demonstrates the best overall performance among the three models considered in this paper, adequately capturing the execution and cancellation risks in-sample and providing satisfactory predictions out-of-sample.[30] Therefore, we believe that the characterization of competing risks provided by the Weibull proportional hazard model with conditionally independent competing risks appears to be more plausible, and the estimates of covariate effects based on this model are more reliable.

VI. Conclusion and Future Research

In this paper we build and estimate three econometric models of limit order execution and cancellation using competing risk analysis and the INET datafeed for five Nasdaq stocks. Of particular importance is our formulation of dependent competing risks models that provides adequate characterization for the times to limit order executions *and cancellations*. We also offer a way to graphically demonstrate the goodness-of-fit of the alternative models. Time-to-execution is found to be more sensitive to the limit price than time-to-cancellation, but time-to-cancellation appears to be more sensitive to the limit size than the time-to-execution (measured as time-to-first-fill). Therefore, a safe way to reduce the time-to-execution with minimal increase in the risk of later cancellation is to increase the price aggressiveness of the submitted limit order. More importantly, we find that if investors want to reduce the expected time-to-execution for their limit orders *without inducing any significant increase in the risk of subsequent cancellation* should submit their orders when the market depth is smaller on the side of their orders or when the market depth is greater on the opposite side of their orders.

Our future research plans include (but are not limited to) the following. First, since the order size (related to the covariate $SZSD$) appears to be an important factor influencing the cancellation risk, and often is used to distinguish retail orders from institutional orders, we plan to compare results by splitting orders according to order size. Similarly, since cancellation strategies used by limit order traders are likely to depend on price aggressiveness (determined by the covariate $MQLP$), it may be advantageous to compare results by splitting orders according to their price aggressiveness. This may reveal some important differences

[30] The out-of-sample diagnostic plots are available from the authors upon request.

in the trading strategies employed by market participants.

Second, introducing time-varying covariates in our models is an important direction for future research. Traders submit their limit orders based on current market conditions and their anticipation of future changes in market conditions, so it is reasonable to expect that the limit order trading strategy is a dynamic process unfolding in real time. Introducing time-varying covariates would allow us to capture the effect of adjustments that many limit order traders are likely to make in response to changes in market conditions over the lifetime of the order.

Finally, the large fraction of cancelled orders suggests that it would be important to consider what a trader does after cancelling an order. For example, if the trader wants to buy a certain number of shares, then we can think of an expected price and time-to-execution trade-off. The typical trader on INET will submit multiple orders to meet his objective and we could use the competing risks model estimates to generate a forecast for the time until completion for a given policy; where a policy may be defined as a distance between the quote and the order price. An interesting question might be to determine the price-time trade-off, which is a potential measure of traders' demand for liquidity.

Appendix I: Technical Derivations

Derivation of (3)

Note that we observe only a sample of realizations of (T, J, δ). We know that if $\delta = 0$, then $T = t \geq t_c$ and if $\delta = 1$, we observe either $(T = T_1 = t,\ J = j = 1)$ or $(T = T_2 = t,\ J = j = 2)$. Furthermore, the left censoring assumption $T_1 > t_0, T_2 > t_0$ is always maintained and for notational simplicity, we use $\Pr(\cdot|t_0)$ and $\Pr(\cdot|T_1 > t_0, T_2 > t_0)$ interchangeably. Therefore, assuming that the censoring rule is exogenous, which is the case of limit orders, the log-likelihood function for a single observation (T, J, δ) takes the form

$$\ln L = (1-\delta)\ln \Pr(\delta = 0|T > t_0) + \delta \ln \Pr\left(T = t, J = j|\delta = 1, T > t_0\right), \qquad \text{(A1)}$$

where

$$\Pr\left(T = t, J = j|\delta = 1, T > t_0\right) = \begin{cases} \Pr(T = T_1 = t|\delta = 1, j = 1, T = T_1 > t_0) \\ \quad \times \Pr(\delta = 1, j = 1|T = T_1 > t_0) \text{ if } j = 1, \\ \Pr(T = T_2 = t|\delta = 1, j = 2, T = T_2 > t_0) \\ \quad \times \Pr(\delta = 1, j = 2|T = T_2 > t_0) \text{ if } j = 2, \end{cases} \qquad \text{(A2)}$$

is the joint probability density for (T, J) conditional on $\delta = 1$. As a result, in order to derive the likelihood function, we just need to obtain expressions for $\Pr(\delta = 0|T > t_0)$ and each of the probabilities on the right-hand side of (A2). We derive each of them in turn.

First,

$$\begin{aligned} \Pr(\delta &= 0|T > t_0) = \Pr(T \geq t_c|T > t_0) \\ &= \Pr(T_1 \geq t_c,\ T_2 \geq t_c|T_1 > t_0,\ T_2 > t_0) = S(t_c, t_c|t_0). \end{aligned} \qquad \text{(A3)}$$

Second,

$$\begin{aligned} \Pr(T &= T_1 = t|\delta = 1, j = 1, T = T_1 > t_0)\Pr(\delta = 1, j = 1|T = T_1 > t_0) \\ &= \frac{\Pr(T = T_1 = t|T = T_1 > t_0)}{\Pr(\delta = 1, j = 1|T = T_1 > t_0)}\Pr(\delta = 1, j = 1|T = T_1 > t_0) \\ &= \Pr(T = T_1 = t|T = T_1 > t_0) \\ &= \Pr\left(T_2 > t\,|\,T_1 = t, T = T_1 > t_0\right)\Pr(T_1 = t|T = T_1 > t_0). \end{aligned} \qquad \text{(A4)}$$

Therefore, in order to derive $\Pr(T = T_1 = t|\delta = 1, j = 1, T = T_1 > t_0)\Pr(\delta = 1, j = 1|T = T_1 > t_0)$, we just need to obtain $\Pr\left(T_2 > t\,|\,T_1 = t, T = T_1 > t_0\right)\Pr(T_1 = t|T = T_1 > t_0)$, which can be obtained as follows

$$
\begin{aligned}
\Pr\left(T_2 > t \,|\, T_1 = t, T = T_1 > t_0\right) \Pr(T_1 &= t|T = T_1 > t_0) \\
&= \lim_{\varepsilon \to 0} \Pr\left(T_2 > t \,|\, t \le T_1 \le t + \varepsilon, T_1 > t_0,\, T_2 > t_0\right) \\
\times \Pr(T_1 &= t|T_1 > t_0,\, T_2 > t_0) \\
&= \lim_{\varepsilon \to 0} \frac{\Pr\left(t + \varepsilon > T_1 > t, T_2 > t|T_1 > t_0,\, T_2 > t_0\right)}{\Pr\left(t + \varepsilon > T_1 > t|T_1 > t_0,\, T_2 > t_0\right)} \\
\times \Pr(T_1 &= t|T_1 > t_0,\, T_2 > t_0)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Pr\left(t + \varepsilon > T_1 > t, T_2 > t|T_1 > t_0,\, T_2 > t_0\right)}{\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Pr\left(t + \varepsilon > T_1 > t|T_1 > t_0,\, T_2 > t_0\right)} \Pr(T_1 = t|T_1 > t_0,\, T_2 > t_0) \\
&= -\left.\frac{\partial S\left(t_1,\, t_2|t_0\right)}{\partial t_1}\right|_{t_1 = t_2 = t} = f\left(t, 1|t_0\right), \qquad \text{(A5)}
\end{aligned}
$$

that is

$$
\begin{aligned}
\Pr(T &= T_1 = t|\delta = 1, j = 1, T = T_1 > t_0) \Pr(\delta = 1, j = 1|T = T_1 > t_0) \\
&= -\left.\frac{\partial S\left(t_1,\, t_2|t_0\right)}{\partial t_1}\right|_{t_1 = t_2 = t} = f\left(t, 1|t_0\right) \qquad \text{(A6)}
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
\Pr(T &= T_2 = t|\delta = 1, j = 2, T = T_2 > t_0) \Pr(\delta = 1, j = 2|T = T_2 > t_0) \\
&= -\left.\frac{\partial S\left(t_1, t_2|t_0\right)}{\partial t_2}\right|_{t_1 = t_2 = t} = f\left(t, 2|t_0\right). \qquad \text{(A7)}
\end{aligned}
$$

Therefore, the conditional log-likelihood function for a sample of n observations is

$$
\ln L_c = \prod_{i=1}^{n} \left((1 - \delta_i) \ln S\left(t_{ci}, t_{ci}|t_0\right) + \delta_i \ln f\left(t_i, j_i|t_0\right)\right) \qquad \text{(A8)}
$$

where

$$
S\left(t_1, t_2|t_0\right) = \Pr\left(T_1 > t_1,\ T_2 > t_2|T_1 > t_0, T_2 > t_0\right), \qquad \text{(A9)}
$$

and

$$
f\left(t,\, j|t_0\right) = -\left.\frac{\partial S\left(t_1, t_2|t_0\right)}{\partial t_j}\right|_{t_1 = t_2 = t}, \qquad j = 1, 2. \qquad \text{(A10)}
$$

Derivation of (5)

$$
\begin{aligned}
S\left(t_1,t_2|t_0\right) &= \Pr(t_c \geq T_1 > t_1,\ t_c \geq T_2 > t_2|t_0) \\
&= \Pr\left(T_1 \leq t_c, T_2 \leq t_c|t_0\right) - \Pr\left(T_1 \leq t_1, T_2 \leq t_c|t_0\right) \\
-F(T_1 &\leq t_c, T_2 \leq t_2|t_0) + F\left(T_1 \leq t_1, T_2 \leq t_2|t_0\right) \\
&= F\left(t_c,t_c|t_0\right) - F\left(t_1,t_c|t_0\right) - F(t_c,t_2|t_0) + F\left(t_1,t_2|t_0\right). \qquad \text{(A11)}
\end{aligned}
$$

Derivation of (8)

Since T_1 and T_2 are independent, given that each of these durations exceeds t_0, we have $S\left(t_1,t_2|t_0\right) = S\left(t_1,t_2|T_1 > t_0, T_2 > t_0\right) = S_1\left(t_1|T_1 > t_0\right)S_2(t_2|T_2 > t_0)$. On the other hand,

$$
\begin{aligned}
h_j\left(t|t_0\right) &= \lim_{\triangle t \to 0} \frac{P\left(t \leq T_j < t + \triangle t\,|\,T_j \geq t, T_j > t_0\right)}{\triangle t} \\
&= \lim_{\triangle t \to 0} \frac{P\left(t \leq T_j < t + \triangle t|T_j > t_0\right)}{\triangle t \Pr(T_j \geq t|T_j > t_0)} \\
&= \frac{f_j(t|t_0)}{S_j(t|t_0)} = -\frac{\partial \ln S_j(t|t_0)}{\partial t} \qquad \text{(A12)}
\end{aligned}
$$

where $j = 1,2,$ $f_j(t|t_0)$ is the marginal density for T_j conditional on $T_j > t_0$ and the last equality follows from the relationship $S_j(t|t_0) = 1 - F_j(t|t_0)$. Integrating both sides of (25), we have

$$
S_j(t|t_0) = \exp\left(-\int_{t_0}^{t} h_j\left(\tau|t_0\right) d\tau\right),\ j = 1,2. \qquad \text{(A13)}
$$

Therefore,

$$
S\left(t_1,t_2|t_0\right) = \exp\left[-\left(\int_{t_0}^{t_1} h_1\left(\tau|t_0\right) d\tau + \int_{t_0}^{t_2} h_2\left(\tau|t_0\right) d\tau\right)\right]. \qquad \text{(A14)}
$$

Derivation of (17)

Since $h_j\left(t|t_0, \mathbf{Z};\, \boldsymbol{\theta}_j\right)$, $j = 1,2$, are independent conditional on $\mathbf{Z} = \mathbf{z}$, then $S_j\left(t|t_0, \mathbf{Z};\, \boldsymbol{\theta}_j\right)$, $j = 1,2$, are also independent conditional on $\mathbf{Z} = \mathbf{z}$. Therefore, we have

$$
\begin{aligned}
S\left(t_1,t_2|t_0,\, \mathbf{z}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) &= S_1\left(t_1|t_0, \mathbf{z}; \boldsymbol{\theta}_1\right) \cdot S_2\left(t_2|t_0, \mathbf{z}; \boldsymbol{\theta}_2\right) \\
&= \exp\left(-\int_{t_0}^{t_1} h_1\left(\tau|t_0, \mathbf{z}; \boldsymbol{\theta}_1\right) d\tau\right) \cdot \exp\left(-\int_{t_0}^{t_2} h_2\left(\tau|t_0, \mathbf{z}; \boldsymbol{\theta}_2\right) d\tau\right) \\
&= \exp\left[-\left(\int_{t_0}^{t_1} \gamma_1 \tau^{\gamma_1 - 1} \exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) d\tau + \int_{t_0}^{t_2} \gamma_2 \tau^{\gamma_2 - 1} \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) d\tau\right)\right] \\
&= \exp\left[-\left(\exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right)(t_1^{\gamma_1} - t_0^{\gamma_1}) + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right)(t_2^{\gamma_2} - t_0^{\gamma_2})\right)\right], \qquad \text{(A15)}
\end{aligned}
$$

where the first equality follows from $S_j(t|t_0, \mathbf{z}) = \exp\left(-\int_{t_0}^{t} h_j(\tau|t_0, \mathbf{z})\, d\tau\right)$, $j = 1, 2$, which is a result in the derivation of (8).

Derivation of (21)

Since $h_j(t|\mathbf{Z}, V; \boldsymbol{\theta}_j)$, $j = 1, 2$, are independent conditional on $\mathbf{Z} = \mathbf{z}$ and $V = v$, then the functions $S_j(t|\mathbf{Z}, V; \boldsymbol{\theta}_j)$, $j = 1, 2$, are also independent conditional on $\mathbf{Z} = \mathbf{z}$ and $V = v$. Therefore, we have

$$\begin{aligned} S(t_1, t_2|\mathbf{z}, v; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= S_1(t_1|\mathbf{z}, v; \boldsymbol{\theta}_1) \cdot S_2(t_2|\mathbf{z}, v; \boldsymbol{\theta}_2) \\ &= \exp\left(-\int_0^{t_1} h_1(\tau|\mathbf{z}, v; \boldsymbol{\theta}_1)\, d\tau\right) \times \exp\left(-\int_0^{t_2} h_2(\tau|\mathbf{z}, v; \boldsymbol{\theta}_2)\, d\tau\right) \\ &= \exp\left[-\left(\int_0^{t_1} \gamma_1 \tau^{\gamma_1 - 1} \exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) d\tau + \int_0^{t_2} \gamma_2 \tau^{\gamma_2 - 1} \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) d\tau\right) v\right] \\ &= \exp\left[-\left(\exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) t_1^{\gamma_1} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) t_2^{\gamma_2}\right) v\right], \end{aligned} \tag{A16}$$

where the first equality follows from $S_j(t|\mathbf{z}) = \exp\left(-\int_{t_0}^{t} h_j(\tau|\mathbf{z})\, d\tau\right)$, $j = 1, 2$, which is in turn a result of the derivation of (8), while the second equality follows from the specification (19).

Since V is not observable, $S(t_1, t_2|\mathbf{z}, v; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ cannot be used for estimation. We have to integrate v out of the above expression to obtain the joint survival function conditional on $\mathbf{Z} = \mathbf{z}$ but unconditional on V. With the assumed gamma density for the random variable V as (20), we have

$$\begin{aligned} S(t_1, t_2|\mathbf{z}; \boldsymbol{\Theta}) &= \int_0^{\infty} S(t_1, t_2|\mathbf{z}, v; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) f(v|\boldsymbol{\varphi})\, dv \\ &= \int_0^{\infty} \exp\left[-\left(\exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) t_1^{\gamma_1} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) t_2^{\gamma_2}\right) v\right] \frac{1}{\tau\Gamma(\alpha)} \left(\frac{v}{\tau}\right)^{\alpha - 1} \exp\left(-\frac{v}{\tau}\right) dv \\ &= \int_0^{\infty} \frac{1}{\tau\Gamma(\alpha)} \left(\frac{v}{\tau}\right)^{\alpha - 1} \exp\left[-\left(\exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) t_1^{\gamma_1} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) t_2^{\gamma_2} + \frac{1}{\tau}\right) v\right] dv \end{aligned} \tag{A17}$$

with parameter vector $\boldsymbol{\Theta} = \left(\tau, \alpha, \gamma_1, \gamma_2, \zeta_1, \zeta_2, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$.

Define $\frac{1}{\eta} = \frac{1}{\tau} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) t_1^{\gamma_1} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) t_2^{\gamma_2}$ and $\frac{1}{\eta_0} = \frac{1}{\tau} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_1\right) t_0^{\gamma_1} + \exp\left(\mathbf{z}'\boldsymbol{\beta}_2\right) t_0^{\gamma_2}$.

Then we can write

$$
\begin{aligned}
S\left(t_1,t_2|t_0,\mathbf{z};\mathbf{\Theta}\right) &= S\left(t_1,t_2|\mathbf{z};\mathbf{\Theta}\right)/S\left(t_0,t_0|\mathbf{z};\mathbf{\Theta}\right)\\
&= \int_0^\infty \frac{1}{\tau\Gamma\left(\alpha\right)}\left(\frac{v}{\tau}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta}\right)dv \Big/ \int_0^\infty \frac{1}{\tau\Gamma\left(\alpha\right)}\left(\frac{v}{\tau}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta_0}\right)dv\\
&= \int_0^\infty \left(\frac{\eta}{\tau}\right)^{\alpha}\frac{1}{\eta\Gamma\left(\alpha\right)}\left(\frac{v}{\eta}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta}\right)dv \Big/ \int_0^\infty \left(\frac{\eta_0}{\tau}\right)^{\alpha}\frac{1}{\eta_0\Gamma\left(\alpha\right)}\left(\frac{v}{\eta_0}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta_0}\right)dv\\
&= \left(\frac{\eta}{\eta_0}\right)^{\alpha}\int_0^\infty \frac{1}{\eta\Gamma\left(\alpha\right)}\left(\frac{v}{\eta}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta}\right)dv \Big/ \int_0^\infty \frac{1}{\eta_0\Gamma\left(\alpha\right)}\left(\frac{v}{\eta_0}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta_0}\right)dv.
\end{aligned}
\tag{A18}
$$

Note that $\frac{1}{\eta\Gamma(\alpha)}\left(\frac{v}{\eta}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta}\right)$ is the gamma density with mean $\eta\alpha$ and variance $\eta^2\alpha$. Therefore, $\int_0^\infty \frac{1}{\eta\Gamma(\alpha)}\left(\frac{v}{\eta}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta}\right)dv = 1$ and, similarly, $\int_0^\infty \frac{1}{\eta_0\Gamma(\alpha)}\left(\frac{v}{\eta_0}\right)^{\alpha-1}\exp\left(-\frac{v}{\eta_0}\right)dv = 1$. As a result, the expression (A18) simplifies as follows

$$
S\left(t_1,t_2|t_0,\mathbf{z};\mathbf{\Theta}\right) = \left(\frac{\eta}{\eta_0}\right)^{\alpha} = \left(\frac{1+\tau(t_0^{\gamma_1}\exp(\mathbf{z}'\boldsymbol{\beta}_1)+t_0^{\gamma_2}\exp(\mathbf{z}'\boldsymbol{\beta}_2))}{1+\tau(t_1^{\gamma_1}\exp(\mathbf{z}'\boldsymbol{\beta}_1)+t_2^{\gamma_2}\exp(\mathbf{z}'\boldsymbol{\beta}_2))}\right)^{\alpha}. \tag{A19}
$$

Appendix II: Construction of Diagnostic Plots

1. Step 1 of the algorithm depends on the functional form of the evaluated model. Therefore, it will be described separately for each of the three alternative competing risks model evaluated in this paper.

   (a) For the *generalized gamma AFT independent competing risks model* with the estimated parameter vector $\widehat{\mathbf{\Theta}} = (\widehat{\boldsymbol{\omega}}_1, \widehat{\boldsymbol{\omega}}_2)$, where $\widehat{\boldsymbol{\omega}}_j$=$(\widehat{\kappa}_j, \widehat{p}_j, \widehat{\boldsymbol{\beta}}_j^{'})'$ characterize the shape of the generalized gamma distributions and the covariate effects ($j = 1, 2$), use the expressions for the risk-specific conditional duration (11) and conditional survival function (12) to generate a pseudo-random sample $\left\{(\widetilde{t}_{1i}, \widetilde{t}_{2i})\right\}$ of risk-specific durations

$$
\widetilde{t}_{ji} = \left(\frac{1}{\widehat{\kappa}_j}\Gamma^{-1}(\widehat{\kappa}_j,\, u_{ji}\Gamma\left(\widehat{\kappa}_j\right))\right)^{1/\widehat{p}_j}\exp(\mathbf{z}_i'\widehat{\boldsymbol{\beta}}_j), \qquad j = 1, 2, \tag{A20}
$$

   where $\{\mathbf{z}_i\}_{i=1}^n$ is the sequence of observed covariates, $\{(u_{1i}, u_{2i})\}_{i=1}^n$ is the sequence of i.i.d. uniformly distributed pseudo-random vectors, and $\Gamma^{-1}(\kappa, y)$ is the inverse gamma function of y (i.e., the solution a of equation $\Gamma(\kappa, a) = y$).

(b) For the *model of conditionally independent Weibull proportional hazard competing risks* with the estimated parameter vector $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)$, where $\widehat{\boldsymbol{\theta}}_j = (\widehat{\gamma}_j, \widehat{\boldsymbol{\beta}}_j^{\prime})^{\prime}$ characterize the shape of the Weibull baseline hazard functions and the covariate effects ($j = 1, 2$), use the expressions for the risk-specific conditional survival function (17) to generate a pseudo-random sample $\{(\widetilde{t}_{1i}, \widetilde{t}_{2i})\}$ of risk-specific durations

$$\widetilde{t}_{ji} = \left(\frac{-\ln(u_{ji})}{\exp(\mathbf{z}_i^{\prime}\widehat{\boldsymbol{\beta}}_j)} \right)^{1/\widehat{\gamma}_j}, \qquad j = 1, 2, \tag{A21}$$

where $\{\mathbf{z}_i\}_{i=1}^n$ is the sequence of observed covariates, and $\{(u_{1i}, u_{2i})\}_{i=1}^n$ is the sequence of i.i.d. uniformly distributed pseudo-random vectors.

(c) For the *gamma frailty model of Weibull proportional hazard competing risks* with the estimated parameter vector $\widehat{\boldsymbol{\Theta}} = (\widehat{\alpha}, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\boldsymbol{\beta}}_1^{\prime}, \widehat{\boldsymbol{\beta}}_2^{\prime})$, first obtain the pseudo-random simulated values of frailty coefficients $\{\widetilde{v}_i\}_{i=1}^n$ by drawing them as i.i.d. gamma$(\widehat{\alpha}, 1)$ pseudo-random numbers, then use the expressions for the risk-specific conditional hazard functions (17) and the simulated frailty coefficients $\{\widetilde{v}_i\}_{i=1}^n$ to generate a pseudo-random sample $\{(\widetilde{t}_{1i}, \widetilde{t}_{2i})\}$ of risk-specific durations

$$\widetilde{t}_{ji} = \left(\frac{-\ln(u_{ji})}{\widetilde{v}_i \exp(\mathbf{z}_i^{\prime}\widehat{\boldsymbol{\beta}}_j)} \right)^{1/\widehat{\gamma}_j}, \qquad j = 1, 2, \tag{A22}$$

where $\{\mathbf{z}_i\}_{i=1}^n$ is the sequence of observed covariates, and $\{(u_{1i}, u_{2i})\}_{i=1}^n$ is the sequence of i.i.d. uniformly distributed pseudo-random vectors.

2. For each value of index $i = 1, \ldots, n$ such that $\widetilde{t}_i \equiv \min(\widetilde{t}_{1i}, \widetilde{t}_{2i}) \leq t_0$, discard this value of $\widetilde{t}_i$ and return to step 1.

3. Using $\widetilde{t}_{ji}$, $j = 1, 2$, obtained at steps 1 and 2, determine the sample $\{(\widetilde{t}_i, \widetilde{j}_i, \widetilde{\delta}_i, \mathbf{z}_i^{\prime})\}_{i=1}^n$, where $\widetilde{t}_i = \min(\widetilde{t}_{1i}, \widetilde{t}_{2i}) > t_0$, $\widetilde{j}_i = \underset{j=1,2}{\arg\min} \widetilde{t}_{ji}$, $\widetilde{\delta}_i = 0$ if $\widetilde{t}_i > t_c(\mathbf{z}_i)$ and $\widetilde{\delta}_i = 1$ if $\widetilde{t}_i \leq t_c(\mathbf{z}_i)$, and $\mathbf{z}_i$ is the observed covariates that pre-determine the right-censoring duration $t_c(\mathbf{z}_i)$. Generate the auxiliary sequence $\{s_i\}_{i=1}^n$ of i.i.d. uniformly distributed pseudo-random numbers.

4. Construct the ordered sequence $\{\widehat{\pi}_{(i)}\}_{i=1}^n$ of the predicted risk-specific incidence rates

by formula (24) to obtain

$$\widehat{\pi}_i = \begin{cases} Q_1(\widetilde{t}_i|t_0;\widehat{\mathbf{\Theta}}) & \text{if } \widetilde{j}_i = 1 \text{ and } \widetilde{\delta}_i = 1, \\ Q_2(\widetilde{t}_i|t_0,\mathbf{z}_i;\widehat{\mathbf{\Theta}}) & \text{if } \widetilde{j}_i = 2 \text{ and } \widetilde{\delta}_i = 1, \\ Q_c(t_0,\mathbf{z}_i;\widehat{\mathbf{\Theta}}) \cdot s_i & \quad\text{if } \widetilde{\delta}_i = 0, \end{cases} \tag{A23}$$

where the incidence rates $Q_j(t|t_0;\mathbf{\Theta})$ for $t > t_0$ and $j = 1, 2$ are approximated by

$$\widetilde{Q}_j(t|t_0;\widehat{\mathbf{\Theta}}) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{\widetilde{t}_k \le t \text{ and } \widetilde{j}_k = j\}, \tag{A24}$$

and the proxy for the right-censoring rate $Q_c(t_0;\mathbf{\Theta})$ can be obtained as

$$\widetilde{Q}_c(t_0;\widehat{\mathbf{\Theta}}) = 1 - \widetilde{Q}_1(\infty|t_0;\widehat{\mathbf{\Theta}}) - \widetilde{Q}_2(\infty|t_0;\widehat{\mathbf{\Theta}}). \tag{A25}$$

References

An, M. Y., 2004, Likelihood based estimation of a proportional hazard competing risk model with grouped duration data, Working paper, Fannie Mae.

Barclay, M. J., and T. Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041–1073.

Battalio, R., J. Greene, and B. Hatch, 2002, Does the order routing decision matter? *Review of Financial Studies* 15, 159–194.

Berg, G. J. van den, 2000, Duration models: specification, identification, and multiple durations, *Handbook of Econometrics*, vol. 5, North-Holland.

Bernoulli, D., 1766, Essai d'une nouvelle analyse de la mortalite causee par la petite verole, *Mem. Math. Phy. Acad. Royal Sciences*, Paris. (English translation "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it" published in *Smallpox Inoculation: An Eighteenth Century Mathematical Controversy*, 1971, Adult Education Department, Nottingham.)

Biais, B., P. Hillion, and C. Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance* 50, 1655–89.

Biais, B., P. Hillion, and C. Spatt, 1999, Price discovery and learning during the preopening period in the Paris Bourse, *Journal of Political Economy* 107, 1218–48.

Bisière, C. and T. Kamionka, 2000, Timing of orders, order aggressiveness and the order book in the Paris Bourse, *Annals d'Économie et de Statistique* 60, 43–72.

Bessembinder, H., 2002, Trade execution costs and market quality after decimalization, *Journal of Financial and Quantitative Analysis* 38, 747–777.

Bierens, H. J. and J. R. Carvalho, 2006, Semi-nonparametric competing risks analysis of recidivism, forthcoming in *Journal of Applied Econometrics*.

Chakrabarty, B., B. Li, V.T. Nguyen, and R. Van Ness, 2005, Trade classification algorithms for electronic communication network trades, Working paper, University of Mississippi.

Chakravarty, S., R. Wood, and R. Van Ness, 2004, Decimals and liquidity: a study of the NYSE, *Journal of Financial Research* 27, 75–94.

Chung, K.H., Chuwonganant, C. and McCormick, D.T, 2003, Order preferencing and market quality on NASDAQ before and after decimalization, *Journal of Financial Economics* 71*,* 581–612.

Chung, K. H., B. Van Ness and R. Van Ness, 1999, Limit orders and the bid-ask spread, *Journal of Financial Economics* 53, 255–287.

Chung, K. H., B. Van Ness and R. Van Ness, 2004, Trading costs and quote clustering on the NYSE and NASDAQ after decimalization, *Journal of Financial Research* 27, 309–328.

Ciochetti, B. A., Y. Deng, G. Lee, J. D. Shilling, and R. Yao, 2003, A proportional hazards model of commercial mortgage default with originator bias, *Journal of Real Estate Finance and Economics* 27, 5–23.

Clayton, D., and J. Cuzick, 1985, Multivariate generalizations of the proportional hazards model, *Journal of the Royal Statistical Society* 148, 82–117.

Cox, D. R., 1972, Regression models and life tables, *Journal of the Royal Statistical Society* 34, 187–220.

Ellul, A., C. W. Holden, P. Jain, and R. Jennings, 2003, Order dynamics: recent evidence from the NYSE, Working paper, Indiana University.

Glosten, L. and L. Harris,1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics* 21, 123–142.

Han, A. and J. Hausman, 1990, Flexible parametric estimation of duration and competing risk models, *Journal of Applied Econometrics* 5, 325–353.

Hasbrouck, J., and G. Saar, 2004, Technology and liquidity provision: the blurring of traditional definitions, Working Paper, New York Univerisity.

Harris, L. and J. Hasbrouck, 1996, Market vs. limit order: the SuperDOT evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 212–231.

Harris, L. and V. Panchapagesan, 2005, The information content of the limit order book: evidence from NYSE specialist trading decisions, *Journal of Financial Markets* 8, 25–67.

Hausman, J., 1978, Specification tests in econometrics, *Econometrica*, 46, 1251–71.

Hautsch, N., 1999, Analyzing the time between trades with a gamma compounded hazard model: an application to LIFFE bund future transactions, Working paper, University of Konstanz.

Heckman, J.J. and B.E. Honoré, 1989, The identifiability of the competing risks model, *Biometrika*, 76, 325–330.

Henker, T. and M. Martens, 2003, Spread decomposition and commonality in liquidity, Working paper, University of New South Wales and Erasmus University.

Hollifield, B., R. Miller, and P. Sandas, 2004, Empirical analysis of limit order markets, *Review of Economic Studies* 71, 1027–2063.

Kalbfleisch, J. D., and R. L. Prentice, 1980, *The Statistical Analysis of Failure Time Data* (John Wiley & Sons, New York).

Katz, L., 1986, Layoffs, recall and the duration of unemployment, Working paper No. 1825, NBER, Cambridge, Massachusetts.

Klein, H. P., and M. L. Moeschberger, 1997, *Survival Analysis Techniques for Censored and Truncated Data* (Springer Press, New York).

Lancaster, T., 1990, *The Econometric Analysis of Transition Data* (Cambridge University Press, Cambridge).

Lo, A. W., A. C. MacKinlay, and J. Zhang, 2002, Econometric models of limit-order executions, *Journal of Financial Economics* 65, 31–71.

Lo, I., and S. Sapp, 2005, Price aggressiveness and quantity: how are they determined in a limit order market, Working paper, Bank of Canada.

McCall, B. P., 1996, Unemployment insurance rules, joblessness, and part-time work, *Econometrica* 64, 647–682.

Odders-White, E., 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3, 259-286.

Parlour, C. A., 1998, Price dynamics in limit order markets, *The Review of Financial Studies* 11, 789–816.

Peterson, M. and E. Sirri, 2002, Order submission strategy and the curious case of marketable limit orders, *Journal of Financial and Quantitative Analysis* 37, 221–241.

Ranaldo, A., 2004, Order aggressiveness in limit order book markets, *Journal of Financial Markets* 7, 53–74.

Rosholm, M. and M. Svarer, 2001, Structurally dependent competing risks, *Economics Letters* 73, 169–173.

Sueyoshi, G., 1992, Semiparametric proportional hazards estimation of competing risks models with time-varying covariates, *Journal of Econometrics* 51, 25–58.

Tyurin, K., 2003, Competing risks and the order flow dynamics in forex electronic brokerage, Working paper, Indiana University.

Yeo, W., 2004, Trading strategies with cancellations, Unpublished paper, Indiana University.

Tables and Figures

Table 1: Descriptive statistics for the sample

This table presents basic numerical characteristics of the four stocks – American Capital Strategies Ltd. (ticker: ACAS), Associated BanCorp (ticker: ASBC), Imclone Systems Inc. (ticker: IMCL), and Career Education Corp (ticker: CECO) – randomly selected for this study, as well as a highly liquid stock, Intel Corp (ticker: INTC). All medians and averages are reported for the period of July–December 2005, unless indicated otherwise.

	ACAS	ASBC	IMCL	CECO	INTC
Average daily # of buy orders	2268	2694	3838	4505	44694
Median buy order size (round lots)	1	1	1	1	3
Average daily # of buy orders executed	208	169	325	406	5547
Average daily # of buy orders cancelled	2061	2525	3512	4100	39146
Average daily # of sell orders	2359	2697	4145	4852	45323
Median sell order size (round lots)	1	1	1	1	3
Average daily # of sell orders executed	236	194	329	420	5715
Average daily # of sell orders cancelled	2123	2506	3816	4432	39518
Average transaction price ($ per share)	37.48	32.02	32.64	36.15	25.41
Median INET transaction size (lots)	1	1	1	1	3
Average daily # of INET transactions	652	363	654	826	11262
Average daily INET trade volume (lots)	685	485	1097	1144	58644
Average NBBO quoted spread ($)	0.022	0.024	0.021	0.021	0.01
Median NBBO depth at bid (lots)	5	4	8	5	260
Median NBBO depth at ask (lots)	4	4	8	4	264
Average monthly realized volatility	8.8%	13.9%	21.2%	21.2%	23.9%
Average weekly realized volatility	13.3%	11.8%	30.9%	32.6%	21.1%
Average daily realized volatility	15.9%	14.4%	44.3%	35.3%	21.3%
Average hourly (10 am–4 pm) volatility	18.9%	16.2%	40.3%	31.2%	22.5%
Price-to-earnings ratio in July, 2005	9.36	12.73	10.25	12.80	16.53
Market capitalization in July, 2005 ($ bn)	4.67	4.15	2.73	2.78	104.71

Table 2: Descriptive statistics of risk-specific durations and covariates for the ACAS stock

This table provides the descriptive statistics for durations (times-to-termination) and covariates used in the competing risks analysis based on buy and sell limit orders for the American Capital Strategies Ltd. (ACAS) stock. The sample includes all INET limit orders submitted between 10 a.m. and 4 p.m. EST in July–December 2005. Limit orders submitted more than \$0.25 away from the bid-ask mid-quote are excluded. Limit orders executed or cancelled within two seconds after their submission are excluded. The upper triangle of the correlation matrix is for Sell orders, and the lower triangle is for Buy orders. All variable definitions are in Section III, B. of the text.

	Variable	Mean	St.Dev.	Skew	Kurt.	1%-tile	50%-tile	99%-tile
Buy orders	Duration	45.57	190.71	40.2	2646	2.08	13.66	425.12
	$MQLP$	3.05	3.09	3.31	18.1	0.5	2.5	19.5
	$BSID$	−0.02	0.97	0.03	1.07	−1	0	1
	$MKD1$	6.17	1.05	−1.14	9.39	4.61	6.23	8.57
	$MKD2$	5.73	0.96	0.40	2.99	4.22	5.76	8.33
	$SZSD$	5.58	0.87	0.39	4.66	4.11	5.51	7.88
	$STKV$	0.53	0.13	0.12	3.05	0.23	0.53	0.85
	$TURN$	11.48	0.55	0.86	5.24	10.42	11.45	13.07
Sell orders	Duration	44.66	194.25	39.2	2396	2.08	13.64	393.80
	$MQLP$	−2.95	3.00	−3.36	18.8	−19	−2	−0.5
	$BSID$	0.08	0.96	−0.16	1.10	−1	1	1
	$MKD1$	6.08	1.15	−1.20	9.66	0	6.22	8.75
	$MKD2$	5.80	0.93	0.21	2.76	4.26	5.86	8.20
	$SZSD$	5.54	0.86	0.51	4.39	4.40	5.46	7.82
	$STKV$	0.53	0.13	0.08	3.03	0.23	0.53	0.85
	$TURN$	11.46	0.55	0.84	5.05	10.41	11.43	13.04

Correlation matrix for Buy and Sell orders

		Sell orders						
		$MQLP$	$BSID$	$MKD1$	$MKD2$	$SZSD$	$STKV$	$TURN$
Buy orders	$MQLP$		0.044	−0.024	0.112	−0.271	−0.062	−0.076
	$BSID$	0.018		−0.043	0.036	−0.043	−0.007	−0.037
	$MKD1$	0.005	0.057		−0.013	0.270	−0.027	−0.017
	$MKD2$	−0.094	−0.018	−0.021		−0.003	−0.049	−0.045
	$SZSD$	0.300	0.031	0.230	−0.018		0.017	0.037
	$STKV$	0.058	−0.016	−0.019	−0.041	0.019		0.026
	$TURN$	0.075	−0.034	−0.018	−0.032	0.017	0.027	

Table 3: Competing risks estimates for the generalized gamma accelerated failure time model

This table provides the fully parametric maximum likelihood estimates of the generalized gamma accelerated failure time (AFT) model with two independent competing risks for American Capital Strategies Ltd. (ACAS) buy and sell limit orders in July 2005 and December 2005. The estimates in the upper panel correspond to the execution risk, while those in the lower panel correspond to the cancellation risk. Parameters κ_1, p_1 are, respectively, the scale and shape parameters of the generalized gamma distribution for the duration time to buy (sell) limit order execution; parameters κ_2, p_2 are, similarly, the scale and shape parameters of the generalized gamma distribution for the duration time to buy (sell) limit order cancellation. The reported log-likelihood numbers are the log-likelihood value per observation.

	Buy Orders				**Sell Orders**			
	July 2005		December 2005		July 2005		December 2005	
Parameter	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
$MQLP$	0.994	39.1	1.071	41.3	−1.141	−48.3	−1.181	−49.1
$BSID$	0.118	5.51	0.130	6.63	−0.023	−1.27	−0.083	−5.10
$MKD1$	0.315	18.2	0.283	18.3	0.238	18.6	0.277	23.9
$MKD2$	−0.214	−10.3	−0.263	−13.6	−0.291	−15.3	−0.322	−19.5
$SZSD$	−0.046	−1.64	−0.032	−1.27	−0.031	−1.31	0.097	4.28
$STKV$	−1.407	−8.74	−0.481	−3.42	−1.005	−7.64	−0.805	−6.79
$TURN$	−0.689	−14.7	−0.658	−14.9	−0.601	−15.6	−0.584	−16.2
Const $\times 10^{-2}$	0.126	21.9	0.121	22.4	0.120	25.2	0.108	23.6
p_1	0.187	6.94	0.151	6.89	0.220	9.85	0.160	8.96
κ_1	11.252	19.06	15.976	27.8	9.272	21.3	17.482	32.3
$MQLP$	0.046	16.1	0.038	13.1	0.014	5.27	−0.060	−21.0
$BSID$	0.018	3.10	0.040	5.41	−0.006	−1.31	−0.011	−1.48
$MKD1$	0.036	5.81	0.088	12.3	−0.021	−4.55	0.035	5.38
$MKD2$	0.012	2.13	−0.002	−0.23	0.041	7.90	0.071	9.79
$SZSD$	−0.196	−28.1	−0.190	−22.2	−0.167	−28.8	−0.287	−33.8
$STKV$	−0.263	−5.92	−0.702	−11.9	−0.272	−7.45	−0.586	−10.4
$TURN$	−0.257	−19.5	−0.304	−17.2	−0.147	−14.1	−0.312	−18.6
Const $\times 10^{-2}$	0.058	34.2	0.070	31.56	0.047	34.3	0.074	35.4
p_2	−0.962	−39.1	−0.463	−30.5	−1.118	−47.6	−0.453	−30.8
κ_2	0.830	15.9	2.618	59.3	0.677	10.7	2.697	64.3
$\ln L$	−4.5629		−4.9062		−4.5438		−4.8733	
# obs.	42683		35759		56311		40336	
# exec.	2924		4124		3946		5836	
# canc.	39507		31391		52141		34176	
# right-cens.	252		244		224		324	

Table 4: Competing risks estimates for the Weibull proportional hazard model

This table gives the fully parametric maximum likelihood estimates for weibull model with two independent competing risks for the American Capital Strategies Ltd. (ACAS) buy and sell limit orders in July 2005 and December 2005. The estimates in the first panel are associated with execution risk and the estimates in the second panel are associated with cancellation risk. The Weibull parameter estimates $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are both smaller than unity, which can be interpreted as an evidence of decreasing baseline hazards for both execution and cancellation risks, after the effect of observed covariates have been factored out. The reported log-likelihood numbers are the log-likelihood value per observation.

	Buy Orders				**Sell Orders**			
	July 2005		December 2005		July 2005		December 2005	
Parameter	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
$MQLP$	−0.967	−35.5	−0.978	−41.9	1.137	43.0	1.149	44.9
$BSID$	−0.123	−6.03	−0.107	−6.36	0.031	1.64	0.076	4.93
$MKD1$	−0.221	−14.0	−0.172	−12.5	−0.143	−10.7	−0.169	−14.8
$MKD2$	0.202	11.0	0.218	13.6	0.266	13.8	0.298	19.1
$SZSD$	0.026	1.01	0.016	0.76	−0.015	−0.64	−0.127	−6.34
$STKV$	1.083	7.15	0.301	2.63	0.833	6.61	0.628	5.71
$TURN$	0.574	13.0	0.541	14.4	0.567	15.1	0.507	15.1
Const $\times 10^{-2}$	−0.104	−19.3	−0.096	−20.9	−0.108	−22.8	−0.090	−20.7
γ_1	0.728	48.8	0.662	53.8	0.759	55.8	0.688	61.4
$MQLP$	−0.061	−30.2	−0.029	−14.8	0.034	15.1	0.049	24.9
$BSID$	−0.013	−2.43	−0.019	−3.26	0.002	0.34	−0.003	−0.47
$MKD1$	−0.024	−4.65	−0.040	−6.99	0.031	7.28	0.001	0.18
$MKD2$	−0.009	−1.68	0.005	0.73	−0.034	−6.78	−0.055	−9.08
$SZSD$	0.246	38.0	0.170	24.2	0.187	33.6	0.262	36.7
$STKV$	0.319	7.97	0.516	11.5	0.299	8.52	0.511	11.6
$TURN$	0.289	25.3	0.280	20.0	0.145	14.8	0.292	21.6
Const $\times 10^{-2}$	−0.052	−34.9	−0.055	−30.8	−0.035	−27.5	−0.060	−34.5
γ_2	0.386	93.1	0.469	103.6	0.388	110.9	0.470	106.0
$\ln L$	−4.4616		−4.8366		−4.4572		−4.7992	
# obs.	42683		35759		56311		40336	
# exec.	2924		4124		3946		5836	
# canc.	39507		31391		52141		34176	
# right-cens.	252		244		224		324	

Table 5: Competing risks estimates for the Cox proportional hazard model

This table reports the semiparametric maximum likelihood estimates for the Cox proportional hazard (CPH) model of independent competing risks for the American Capital Strategies Ltd. (ACAS) buy and sell limit orders in July 2005 and December 2005. The estimates in the first panel pertain to the hazard rate of execution risk, and the estimates in the second panel pertain to the hazard rate of cancellation risk.

	Buy Orders				**Sell Orders**			
	July 2005		December 2005		July 2005		December 2005	
Parameter	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
$MQLP$	−0.967	−41.3	−0.976	−43.8	1.137	51.7	1.146	50.3
$BSID$	−0.123	−6.38	−0.108	−6.60	0.031	1.83	0.077	5.37
$MKD1$	−0.221	−18.2	−0.172	−17.2	−0.143	−15.6	−0.170	−21.3
$MKD2$	0.201	11.3	0.217	13.8	0.267	16.2	0.299	21.0
$SZSD$	0.031	1.19	0.014	0.70	−0.016	−0.71	−0.130	−6.60
$STKV$	1.083	7.58	0.303	2.71	0.830	6.98	0.620	6.11
$TURN$	0.576	13.6	0.541	14.8	0.568	15.9	0.508	16.1
$MQLP$	−0.063	−30.4	−0.029	−14.5	0.036	19.0	0.049	23.7
$BSID$	−0.012	−2.36	−0.019	−3.28	0.001	0.21	−0.003	−0.46
$MKD1$	−0.026	−4.88	−0.040	−7.29	0.031	7.41	0.001	0.22
$MKD2$	−0.008	−1.48	0.005	0.74	−0.033	−6.63	−0.054	−9.07
$SZSD$	0.249	38.8	0.169	24.1	0.188	34.4	0.260	38.5
$STKV$	0.314	7.98	0.519	12.0	0.301	8.71	0.516	12.2
$TURN$	0.287	25.2	0.283	20.8	0.144	14.7	0.295	22.4
# obs.	42683		35759		56311		40336	
# exec.	2924		4124		3946		5836	
# canc.	39507		31391		52141		34176	
# right-cens.	252		244		224		324	

Table 6: Competing risks estimates for the Weibull model with gamma frailty

This table reports the parametric maximum likelihood estimates for the Weibull model with two dependent competing risks for the American Capital Strategies Ltd. (ACAS) buy and sell limit orders in July 2005 and December 2005. The dependence between competing risks is captured by the latent variable (frailty), which is assumed to have the gamma distribution with mean α and variance α. The estimates in the first panel are associated with the execution risk and the estimates in the second panel are associated with the cancellation risk. The Weibull parameter estimates $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are both larger than unity, which can be interpreted as an evidence of increasing baseline hazards for both execution and cancellation risks, after conditioning on the values of observed covariates and the unobserved gamma factor.

	Buy Orders				**Sell Orders**			
	July 2005		December 2005		July 2005		December 2005	
Parameter	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
$MQLP$	−1.189	−36.5	−1.163	−42.2	1.296	42.1	1.412	45.0
$BSID$	−0.155	−6.59	−0.168	−8.56	0.048	2.19	0.103	5.73
$MKD1$	−0.404	−18.2	−0.339	−18.0	−0.261	−13.9	−0.339	−20.4
$MKD2$	0.235	10.8	0.269	14.5	0.290	12.9	0.332	18.5
$SZSD$	0.305	10.2	0.128	5.17	0.290	9.87	0.019	0.76
$STKV$	1.793	10.4	0.997	7.07	1.465	9.07	1.275	9.63
$TURN$	1.110	21.5	0.828	18.5	0.889	19.2	0.795	19.6
Const $\times 10^{-2}$	−0.189	−30.1	−0.149	−27.0	−0.175	−29.5	−0.145	−27.9
γ_1	1.745	47.3	1.374	69.0	2.065	37.2	1.471	77.4
$MQLP$	−0.126	−23.1	−0.073	−18.4	−0.020	−2.09	0.106	28.3
$BSID$	−0.040	−3.42	−0.058	−5.45	0.022	1.87	0.021	2.07
$MKD1$	−0.100	−7.96	−0.130	−11.4	0.029	2.48	−0.062	−6.21
$MKD2$	−0.030	−2.62	0.023	2.17	−0.102	−8.24	−0.072	−7.07
$SZSD$	0.443	31.3	0.266	21.7	0.443	28.0	0.401	32.6
$STKV$	0.780	8.95	1.078	13.3	0.813	9.28	0.954	12.2
$TURN$	0.670	27.1	0.500	20.4	0.433	17.4	0.497	21.1
Const $\times 10^{-2}$	−0.123	−40.6	−0.104	−34.1	−0.103	−31.4	−0.109	−37.5
γ_2	1.358	44.5	1.138	84.7	1.643	31.4	1.160	92.5
α	0.360	22.1	0.555	32.0	0.256	18.7	0.540	35.3
$\ln L$	−5.2407		−5.5767		−5.2322		−5.5319	
# obs.	42683		35759		56311		40336	
# exec.	2924		4124		3946		5836	
# canc.	39507		31391		52141		34176	
# right-cens.	252		244		224		324	

Table 7: Summary of covariate effects for alternative stocks and time periods

These tables provide the summary of signs for the estimated covariate coefficients in the Weibull proportional hazard model with conditionally independent competing risks for each of the six months in July–December 2005. The top panel of each table gives the signs for the covariate sensitivities of execution risk, while the bottom panel of each table gives the signs for the covariate sensitivities of cancellation risk. We assign $+(-)$ sign if the t -statistic of the estimated parameter is larger than 3.29 (smaller than -3.29), which corresponds to the nominal p-value 0.0005 for the one-sided t-test of statistical significance. $=$ stands for a t -statistic between -3.29 and 3.29. Each sign corresponds to one month.

	Buy Orders (July–December 2005)				
Parameter	ACAS	ASBC	IMCL	CECO	INTC
	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12
$MQLP$	− − − − − −	− − − − − −	− − − − − −	− − − − − −	− − − − − −
$BSID$	− − − − − −	− − − = − −	− = − − − −	− − − − − −	− − = = − =
$MKD1$	− − − − − −	− − − − − −	− − − − − −	− − − − − −	− − − − − −
$MKD2$	+ + + + + +	+ + + + + +	+ = + + + +	+ + + + + +	+ + + + = +
$SZSD$	= − = = = =	= = = − − =	+ = = = − +	+ + + = = +	− − − − − =
$STKV$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + = + = +
$TURN$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + = + + +
$MQLP$	− − − − − −	− − − − − −	− − − − − −	− − − − − −	− − − − − −
$BSID$	= = − − = =	− = = = = =	+ = − = + +	= = = = = =	= + = = + =
$MKD1$	− − − − − −	= = = − = +	+ = = − + =	+ − = − = −	= + − = − −
$MKD2$	= − − − − =	= = = − = =	− + − − − −	+ − + = = −	− + − + − −
$SZSD$	+ + + + + +	= − − − − −	= + + = = +	+ + + + = +	− − − − + +
$STKV$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + = + + +
$TURN$	+ + + + + +	+ + + + + +	+ + + + + +	= + + + + +	+ + = + + +

	Sell Orders (July–December 2005)				
Parameter	ACAS	ASBC	IMCL	CECO	INTC
	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12	07 08 09 10 11 12
$MQLP$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +
$BSID$	= + + = = +	+ + + + + +	+ + + = + +	+ + = + + +	= + + = + =
$MKD1$	− − − − − −	− − − − − −	− − − − − −	− − − − − −	− − − − − −
$MKD2$	+ + = + + +	+ + + + + +	+ = + + + +	+ + + + + +	+ + + + + +
$SZSD$	= − = − = −	= = − − = =	+ = = = = =	+ + + + = +	− − − − + =
$STKV$	+ + + = + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +
$TURN$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + = + + +
$MQLP$	+ + + + + +	+ + − + + +	+ + + + + +	+ + + + + +	+ + + + + +
$BSID$	= = = = = =	= = = = = =	= = + + = +	= = − − = −	− − − − = =
$MKD1$	+ − − = = =	= + = + + +	+ + = = = +	= − + = − −	− + − = − =
$MKD2$	− − − − − −	+ = − − − −	− = = − − −	+ − = − − −	= + − = − =
$SZSD$	+ + + + + +	= − − − − −	− + − − + +	+ + + + + +	+ − − = + +
$STKV$	+ + + + + +	+ + + + = +	+ + + + + +	+ + + + + +	+ + + + + +
$TURN$	+ + + + + +	+ + + + + +	+ + + + + +	+ + + + + +	+ + − + + +

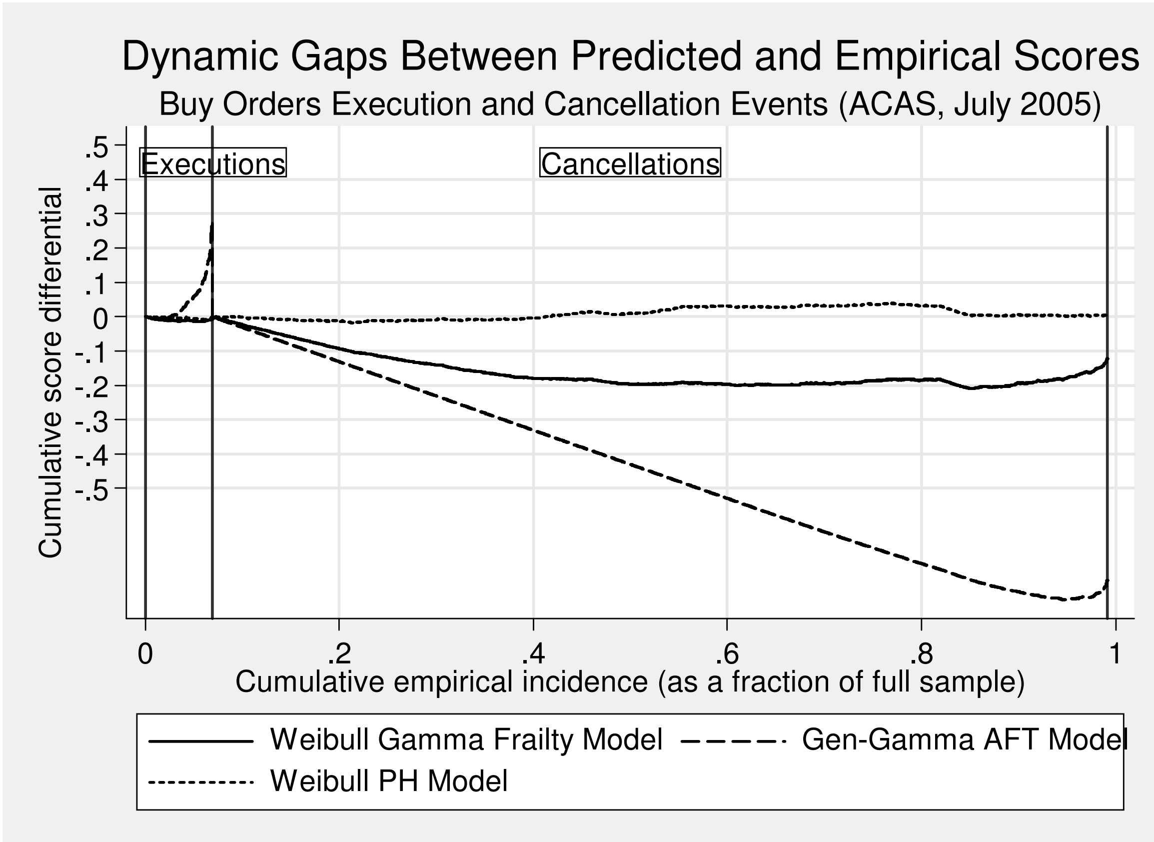

Figure 1: Gaps between realized and predicted frequency scores (incidence rates) of buy order executions and cancellations for ACAS shares. Three models–Weibull PH competing risks, Weibull competing risks with gamma frailty, and generalized gamma AFT competing risks–are compared. Flat gaps to the left of the middle line (separating execution and cancellation events) provide evidence of unbiased predictive performance for limit order execution risk; flat gaps to the right of the middle line provide evidence of unbiased predictive performance for limit order cancellation risk.

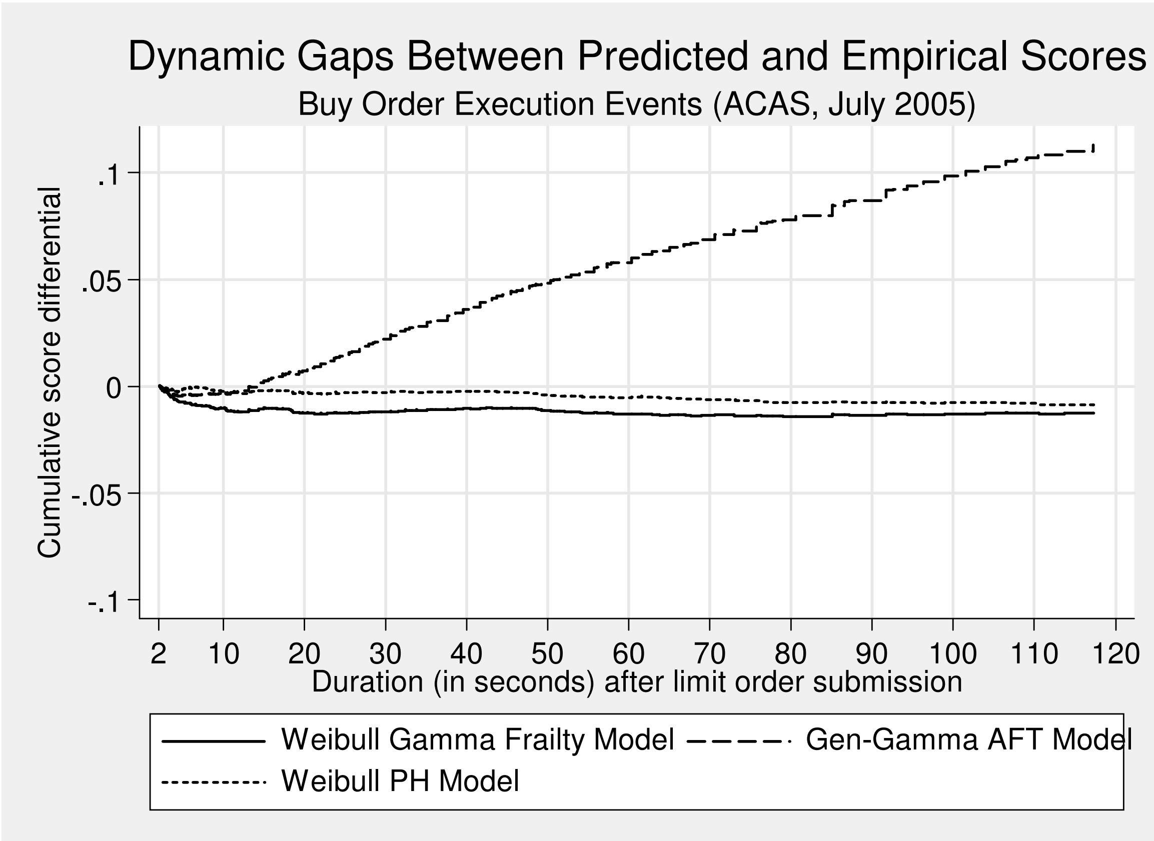

Figure 2a: Dynamic gaps between realized and predicted frequency scores of buy order execution risk for ACAS shares are shown as functions of time elapsed since the limit order arrival. In-sample predictive performance of three alternative competing risks models–Weibull PH, Weibull gamma frailty, and generalized gamma AFT–is compared. The plots are analogous to those of Figure 1, except that they allow to track predictive performance of alternative models for execution risk for different waiting times.

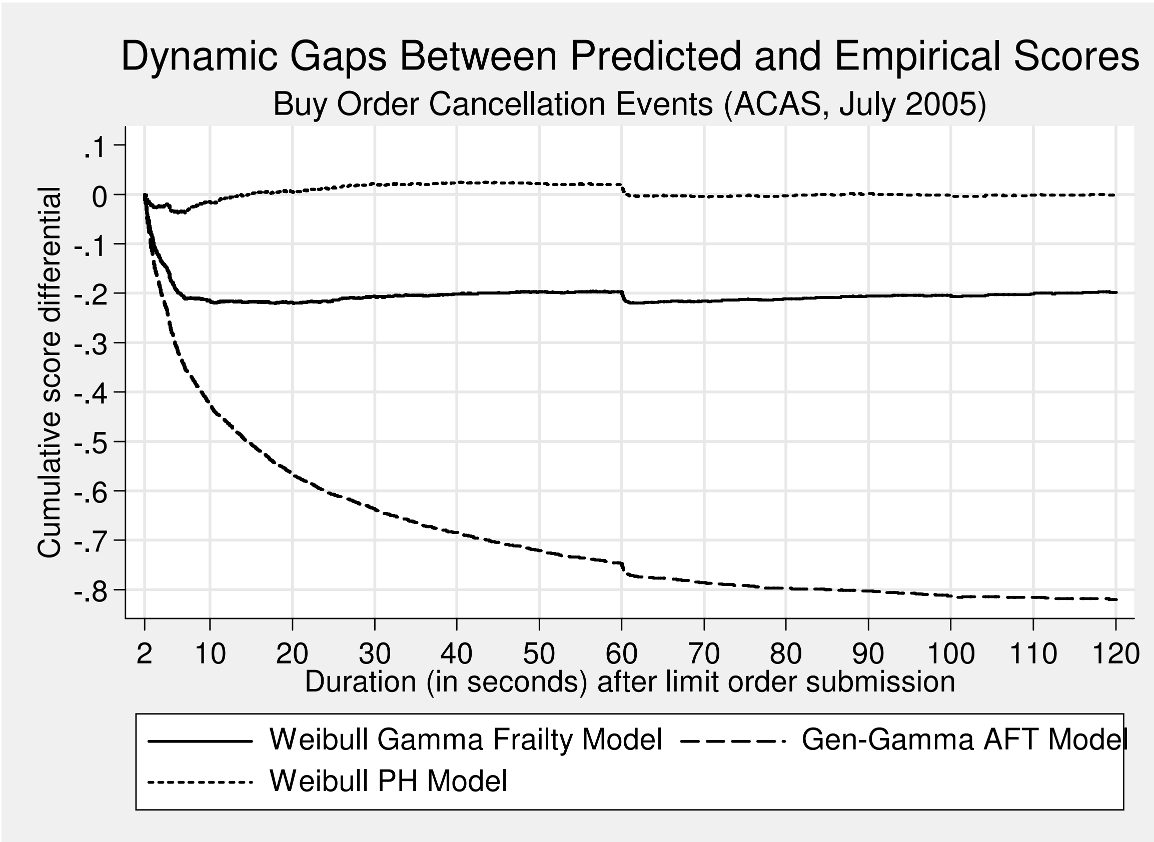

Figure 2b: Dynamic gaps between realized and predicted frequency scores of buy order cancellation risk for ACAS shares are shown as functions of time elapsed since the limit order arrival. In-sample predictive performance of three alternative competing risks models–Weibull PH, Weibull gamma frailty, and generalized gamma AFT–is compared. The plots are analogous to those of Figure 1, except that they allow to track predictive performance of alternative models for cancellation risk for different waiting times.