The Analysis of Limit Orders Using the Cox Proportional Hazards Model with Independent Competing Risks

by

Colleen Kenney

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE Masters of Arts

APPROVED, THESIS COMMITTEE:

K. B. Ensor, Professor, Chair Statistics

Rudolf Riedi, Associate Professor Statistics

Barbara Ostdiek, Associate Professor Management and Statistics

Houston, Texas

November, 2007
Abstract

Title

by

Colleen Kenney

I apply the Cox proportional hazards model with independent competing risks to study the hazard rates of executed, cancelled, and partially executed limit orders submitted for Microsoft to the Island ECN for one day. The instantaneous probability of execution increases with decreases in the buy order price but increases to the sell order price, increases in volume on the sell side of the market and market activity. The probability of cancellation increases with increases in the liquidity demand and market activity for buy orders, volume on the same side of the market and absolute market activity for sell orders. Finally, the partially executed hazard rate for buy orders increases with increases in price, volume on the opposite side of the market, size, and absolute market activity; for sell orders, the hazard rate increases with increases in the volume on the same side of the market, liquidity demanded, and market activity.
Acknowledgments

I would like to thank my committee, especially Dr. Kathy Ensor for her support during the writing process. Also, I thank Premvir Jain and Darrin Gershman for their preliminary work on the original dataset.

This material is based upon work supported by the National Science Foundation under Grant No. 0240058. I wish to acknowledge the financial support for this research from the National Science Foundation, Division of Mathematical Sciences Vertical Integration of Graduate Education and Research program, the Army Research Office and the Center for Computational Finance and Economic Systems as well as the Department of Statistics.
<table>
<thead>
<tr>
<th></th>
<th>Proportional Hazards Tests for Cancelled Buy Orders</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>Proportional Hazards Test for Partially Executed Buy Orders</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>Proportional Hazards Tests for Executed Sell Orders</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>Proportional Hazards Tests for Cancelled Sell Orders</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>Proportional Hazards Tests for Partially Executed Sell Orders</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>12</td>
<td>Nonlinearity diagnostic plots for executed buy orders</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>Nonlinearity diagnostic plots for cancelled buy orders</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>Nonlinearity diagnostic plots for partially executed buy orders</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>Nonlinearity diagnostic plots for executed sell orders</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>16</td>
<td>Nonlinearity diagnostic plots for cancelled sell orders</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>17</td>
<td>Nonlinearity diagnostic plots for partially executed sell orders</td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>
1 Introduction

In the current analysis, I study the relationship between a limit order's characteristics and market conditions at the time of order submission with the time until a limit order's terminal event in continuous time, whether this event is caused by the order's execution or cancellation. I do so in order to better understand the dynamics of limit orders as well as learn optimal market conditions under which to submit a limit order so that it has a better chance of being executed.

The data used to analyze this relationship is from the Electronic Communications Network (ECN) Island, which is run by NASDAQ. In an ECN, investors place limit orders when buying or selling stocks. A limit order is an order to buy or sell a stock at a specific price or better that also allows the trader to limit the amount of time the order remains outstanding prior to cancellation; while advantageous to reduce price risk, execution is not guaranteed because the market price may surpass the limit order's price prior to order execution. Orders may leave the queue because they are executed, cancelled, or a combination of the two, called partial execution. In the current analysis, trading of Microsoft stocks on October 26, 2006, is studied.

This study utilizes survival analysis to model the time until an order exits the queue. Survival analysis is the study of the failure time distribution of the data, and this distribution can be fully characterized by its hazard function, the instantaneous probability that an order will fail in the next instant given survival up to a certain time. Here, the "failure" of an order is taken to be the last execution or cancellation taken on an order submitted to the queue. Under the competing risks framework, I look at the execution, cancellation, or partial execution of an order as events competing for an order. By utilizing this method, I am able to separately model the instantaneous failure rates for each type of failure to better understand the dynamics of the queue.

Specifically, the Cox proportional hazards model with independent competing risks is applied to Microsoft orders added to the queue between 10 a.m. and 4 p.m. on
October 26, 2006, in order to study the failure rates of limit order executions, cancellations, and partial executions. In this type of model, the log hazard function is modeled as a log-linear model of the covariates, and the baseline hazard function is chosen arbitrarily, as it does not affect the estimation of the coefficients vector.

The study of the relationship between limit order characteristics as well as the current conditions in the market with limit order dynamics has become popular with the recent availability of limit order data from electronic markets. Chakrabarty et. al. (forthcoming) utilize the competing risk methodology for cancelled and executed limit orders submitted on an ECN to develop a model for the hazard rates; they find that times until order exit from the limit order queue are best modeled by the Weibull proportional hazards model with independent competing risks. Lo et. al. (2002) treat cancellations as censored observations but models limit order executions with Accelerated Failure Time models. Tyurin (forthcoming) applies the competing risk methodology for buy and sell orders and semi-parametric estimation to analyze order flow and price formation in the Reuters D2000-2 brokerage system. Bisiére and Kamionka (2000) model order flow using competing risks. Hollifield et. al. (2004) model order submission strategies based on the trader’s valuation of the stock. Cited here is only a modest list of the literature.

The remainder of the paper is outlined as follows. In section 2, the data and covariates used for this study are described. A background of the survival analysis methods utilized to analyze the data is presented in section 3. The exploratory data analysis results found prior to application of Cox proportional hazards model with independent competing risks are summarized in section 4; the results of the model are summarized in section 5. The conclusion and future work initiatives can be found in section 6.
2 Data

2.1 Data Set

The data used in this analysis was obtained for Island, an ECN owned by NASDAQ. Online access to the archive of log files was granted to Rice University’s Statistic Department. These log files consist of the information sent to professional traders using the Island ECN.

In the original dataset, the variables for each order included the time (in milliseconds after midnight) of order submission; a factor indicating whether the order was an addition to the order book, a cancellation, an execution, or an execution of a hidden order; the ID number of the order; the price (in dollars); the number of shares; and a factor indicating whether the order was a buy or sell order. For this analysis, I look at additions to the order book and subsequent actions taken on such orders for Microsoft stocks traded on October 26, 2006. I chose this stock due to the high trading volume; looking at multiple days of trading as well as other stocks are left for future research.

Following the work of Chakrabarty et al. (forthcoming), I removed outliers in the data as follows; I do this because I believe orders with these characteristics follow different dynamics than I would like to study in the present analysis. First, I look at trades added to the order book between 10 a.m. and 4 p.m. EST. The “failure time” of an order is assumed as the time at which the last action is taken on the order. I apply right censoring to these data if the last action taken on the order occurred after 4:00 p.m. or in greater than or equal to 10 minutes after order submission; I apply left censoring to the data if the last action taken occurred in less than or equal to 2 seconds following order submission. Also, I exclude orders whose price is greater than $0.25 from the midquote price, which is the 98th percentile of the difference between the price and the midquote price. For buy and sell orders, respectively, there were 50,983 and 50,680 added orders remaining in the order book after applying these
2.2 Covariates

To examine the relationship between the time until last action of an order and the limit order characteristics as well as the conditions present in the order book, I define several covariates. These variables highlight the dynamics of limit order executions and cancellations. Understanding these characteristics results in improved limit order submission strategies.

Let $P_l$ denote the limit order price, $P_b$ and $P_s$ denote the best buy and sell prices, respectively, at the time an order is submitted, and $S_b$ and $S_s$ denote the number of shares available at the best buy and sell prices. Define the midquote price by $P_q = \frac{1}{2} (P_b + P_s)$. Following the work of Lo et al. (2002) and Chakrabarty et al. (forthcoming), the following variables were considered for buy limit orders:

\[
\begin{align*}
MQLP &= P_q - P_l \\
BSID &= \begin{cases} 
1 & \text{if previous trade was a sell trade} \\
-1 & \text{if previous trade was a buy trade}
\end{cases} \\
MKD_1 &= \begin{cases} 
\ln (S_b) (1 + 100 (P_b - P_l)) & \text{if } P_l \leq P_b \\
0 & \text{if } P_l > P_b
\end{cases} \\
MKD_2 &= \begin{cases} 
\ln (S_s) (1 + 100 (P_s - P_l)) & \text{if } P_l \leq P_s \\
0 & \text{if } P_l > P_s
\end{cases} \\
SZSD &= \begin{cases} 
\ln (S_l) (1 + 100 (P_s - P_l)) & \text{if } P_l < P_s \\
\ln (S_l - S_s) & \text{if } P_l = P_s \text{ and } S_l > S_s \\
0 & \text{otherwise}
\end{cases} \\
STKV &= \frac{\text{# of trades in the last half-hour}}{\text{# of trades in the last hour}} \\
TURN &= \ln (\text{# of trades in the last hour})
\end{align*}
\]
Similar covariates are considered for sell orders, with

\[
MQLP = P_t - P_q
\]

\[
MKD1 = \begin{cases} 
\ln(S_s) (1 + 100 (P_t - P_s)) & \text{if } P_t \geq P_s \\
0 & \text{if } P_t < P_s 
\end{cases}
\]

\[
MKD2 = \begin{cases} 
\ln(S_b) (1 + 100 (P_t - P_b)) & \text{if } P_t \geq P_b \\
0 & \text{if } P_t < P_b 
\end{cases}
\]

\[
SZSD = \begin{cases} 
\ln(S_l) (1 + 100 (P_t - P_b)) & \text{if } P_t > P_b \\
\ln(S_l - S_b) & \text{if } P_t = P_b \text{ and } S_l > S_b \\
0 & \text{otherwise} 
\end{cases}
\]

The covariates for each order are defined at the time of order submission and thus are assumed to be fixed rather than time-varying. \textit{MQLP} measures the distance between the midquote price and the limit order price. \textit{BSID} indicates which side of the market the most recent transaction was initiated from, regardless of whether it was executed or cancelled. \textit{MKD1} estimates the minimum number of shares with higher priority on the same side of the market, scaled by the distance between the limit order price and the best price, while \textit{MKD2} estimates the liquidity demanded on the opposite side of the market. \textit{SZSD} measures the liquidity demanded by the limit order scaled by the distance between the limit order price and the best price on the opposite side of the market. Both \textit{STKV} and \textit{TURN} are measures of trading activity; \textit{STKV} estimates short term shifts in trading activity, while \textit{TURN} measures the absolute recent trading activity. Each of these variables was created using the data from the log file provided by Island.

3 Background

Here, I present a brief review of the survival analysis methods used in this study;
more detail of these methods can be found in Kalbfleisch and Prentice (2002), Klein and Moeschberger (2003), and Lee and Wang (2003).

Survival analysis focuses on studying the failure time distribution of data. In our application, the failure time is taken as the time at which an order exits the queue, i.e., the time an order is fully executed and/or cancelled. This distribution can be fully characterized by the hazard function, which gives the instantaneous probability that an order remaining in the queue at least $t$ seconds leaves the queue in the next instant; from this, we are able to derive both the survival function and the probability density function of the failure times.

Consider data consisting of $(T_i, \delta_i, x_i(t))$, $i = 1, \ldots, n$, where $T_i$ is the (continuous) time until exit of the $i$th order from the queue, $\delta_i$ is an indicator equal to 1 if the order has left the queue and 0 if the data are right-censored, and $x_i(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{ip}(t))$ is the covariate vector of the $i$th order at time $t$. Of interest is estimation of the distribution of the hazard function, $h(t|x)$, defined by

$$h(t|x) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i \leq t + \Delta t | T_i \geq t, x_i)}{\Delta t} \quad (1)$$

and its relationship with the covariates vector $x$.

A commonly used model to estimate this relationship is the Cox Proportional Hazards model, which is assumed to be of the form

$$h(t|x) = h_0(t) \exp (\beta^T x) \quad (2)$$

here, the baseline hazard function $h_0(t)$ is chosen arbitrarily and the coefficients vector $\beta$ is estimated from the data using the method of maximum partial likelihoods described below.

While it does assume the presence of proportional hazards, unlike parametric models, this semi-parametric model allows us to estimate the survival distribution without specifying the form of the baseline hazard function. The model also offers flexibility,
with extensions that include allowance of stratified and time-varying covariates.

Suppose we have \( D \) distinct times until exit \( t_1 < t_2 < \ldots < t_D \), and let \( x_{(i)k} \) denote the \( k \)th covariate of the order which exited the queue at time \( t_i \). At each time \( t_i \), let \( R(t_i) \) denote the set of orders in the queue, otherwise called the risk set. To estimate the coefficients vector \( \beta \), the method of maximum partial likelihoods is used; assuming no tied survival times, the partial likelihood function is given by

\[
L(\beta) = \prod_{i=1}^{D} \frac{\exp \left( \sum_{k=1}^{p} \beta_k x_{(i)k} \right)}{\sum_{l \in R(t_i)} \exp \left( \sum_{j=1}^{p} \beta_j x_{jl} \right)} = \prod_{i=1}^{D} \frac{\exp \left( \beta^T x_i \right)}{\sum_{l \in R(t_i)} \exp \left( \beta^T x_l \right)}.
\]

Here, the numerator estimates the probability that an individual order exits the queue at time \( t_i \) given survival to time \( t_i \), while the denominator estimates the probability that any one order exits the queue at time \( t_i \) given survival to time \( t_i \).

However, tied survival times are oftentimes present in data, the current application included. To account for this, we can approximate the partial likelihood function with one presented by Efron (1977). As before, suppose we have \( D \) distinct times until exit \( t_1 < t_2 < \ldots < t_D \) and the risk set \( R(t_i) \) at each time \( t_i \). Let \( d_i \) denote the number of orders which exit the queue at time \( t_i \) and \( D_i \) denote the set of all such orders. Let \( s_i = \sum_{j \in D_i} x_j \), that is, the sum of the vectors \( x_j \) of the orders which exit the queue at time \( t_i \). Then we can approximate the partial likelihood by

\[
L_E(\beta) = \prod_{i=1}^{D} \frac{\exp \left( \beta^T s_i \right)}{\prod_{j=1}^{d_i} \left[ \sum_{l \in R(t_i)} \exp \left( \beta^T x_l \right) \right]^{d_i} \sum_{l \in D_i} \exp \left( \beta^T x_l \right)}.
\]

Upon finding the partial likelihood function, we can perform inference as we would with a normal likelihood. In particular, we can find the estimates \( \hat{\beta} \) by solving the equations \( \frac{d\ln L(\beta)}{d\beta} = 0 \) and \( Cov(\hat{\beta}) \) by computing \( \left( \frac{d\ln L(\hat{\beta})}{d\beta^T d\beta} \right)^{-1} \), both using the Newton-Raphson iterated procedure.

One extension of the Cox Proportional Hazards model allows for multiple strata. For the stratified Cox model, the strata divide orders into disjoint groups, each of
which has a distinct baseline hazard function but common values for their coefficients vectors. For the $k$th stratum, the hazard function is given by

$$h_k(t | x) = h_{0k}(t) \exp(\beta^T x)$$

and the overall likelihood function is the sum of each stratum's likelihood function, that is,

$$L(\beta) = \sum_{k=1}^{K} L_k(\beta),$$

where $L_k(\beta)$ is the partial likelihood function for the $k$th stratum. This extension provides a means of applying the proportional hazards model if, for example, the proportional hazards assumption is violated for one or more of the covariates and these covariates take only a few values.

We might also consider the different ways in which orders leave the queue and model each of these differently. Orders leave the queue due to cancellation, execution, or a combination of the two, called partial execution, and we might regard these modes of exit as competing risks.

Competing risk analysis primarily studies the relationship between the covariates and the failure rates of specific failure types. While it is also of interest to study the relationship between failure types as well as the risk of one type after removing other failure types, estimation of these quantities is not straightforward.

Suppose there are $J = 1, ..., m$ types of exit from the queue and that only one can occur for each order. Then the cause- or type-specific hazard function is defined by

$$h_j(t | x) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t, J = j | T \geq t, x)}{\Delta t},$$

and, in the case when only one cause of failure can occur, the overall hazard function is the sum of the cause-specific hazard functions,
\[ h(t | x) = \sum_{j=1}^{m} h_j(t | x). \]  

(8)

Hence, using the competing risks methodology allows for study of the cause-specific failure rates and their impact on the overall hazard function.

Also suppose that each pair of covariate vectors \( x_i \) and \( x_j, i, j = 1, \ldots, n \), have proportional hazards. Then the relationship between the covariates and the cause-specific hazard function can be estimated by

\[ h_j(t | x) = h_{0j}(t) \exp \left( \beta^T x \right). \]  

(9)

The partial likelihood in this case becomes

\[ L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{k_j} \frac{\exp \left( \beta_j^T x_{ji} \right)}{\sum_{l \in R(t_i)} \exp \left( \beta_j^T x_l \right)} \]  

(10)

where \( k_j \) is the number of observations for the \( j \)th exit mode, \( j \in J \). Again, since more than one event may occur at a given time \( t_i \), \( i = 1, \ldots, n \), we can adapt the Efron approximation of the partial likelihood function in order to estimate the covariates vector \( \beta \).

4 Exploratory Data Analysis

Table 1 summarizes the observations for buy and sell orders after the filters outlined in Section 2.1 were applied. Buy and sell orders each make up about half of the dataset, with 50,983 and 50,680 observations, respectively. 18.0% and 81.0% of buy orders are executed and cancelled, respectively; 17.5% and 81.5% of sell orders are executed and cancelled. The remaining 1% for each of the two subsets are orders that are partially executed. The proportion of cancelled orders is similar to those reported by Hasbrouck and Saar (2005).

Table 2 provides the mean, median, and standard deviation of the time until last
action for buy and sell orders. Both completed and cancelled buy and sell orders on average take a little over a minute to be completed with median times between about 20 and 30 seconds. Partially executed orders’ mean and median times until completion are slightly longer, with average times of about 89 and 87 seconds and median times of about 45 and 41 seconds, respectively, for buy and sell orders.

Table 1: Summary of Orders

<table>
<thead>
<tr>
<th></th>
<th>Buy Orders</th>
<th>Sell Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td># of observations</td>
<td>50,983</td>
<td>50,680</td>
</tr>
<tr>
<td>% executed</td>
<td>18.0%</td>
<td>17.5%</td>
</tr>
<tr>
<td>% cancelled</td>
<td>81.0%</td>
<td>81.5%</td>
</tr>
<tr>
<td>% partially executed</td>
<td>1.0%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Table 2: Summary of Times

<table>
<thead>
<tr>
<th></th>
<th>Buy Orders</th>
<th>Sell Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executions</td>
<td>64.25</td>
<td>68.63</td>
</tr>
<tr>
<td>Cancellations</td>
<td>65.60</td>
<td>63.98</td>
</tr>
<tr>
<td>Partial Executions</td>
<td>88.58</td>
<td>87.23</td>
</tr>
</tbody>
</table>

Heat plots were also generated to better understand limit orders’ time until exit from the queue; the plots can be found in figure 1. For executed, cancelled, and partially executed orders, respectively, 70%, 68%, and 57% were completed in less than 60 seconds, and 14%, 15%, and 19% were completed between one and two minutes; note that orders completed in 2 seconds or less as well as in more than 10 minutes were excluded from this analysis. Interestingly, the cancelled orders plots show short-lived submissions that form horizontal lines; this may be due to traders attempting to influence the prices on both sides of the market.

5 Model and Results

5.1 Buy Orders

For added buy limit orders, the Newton-Raphson iterated procedure was applied to (4) for each of executed, cancelled, and partially executed orders. Upon obtaining
Figure 1: Heat Plots by Order Type
The following are heat plots for each of the failure types for both buy and sell orders. The spectrum runs from red to violet, with orders surviving around 2 seconds in red and orders surviving around 10 minutes in violet.
estimates for the coefficients vector $\beta$ for the covariates described in section 2.2, a Wald test was performed to identify the statistically significant covariates at the 0.05 level. The three models are given by

$$h_{\text{Executed}}(t|x) = \sum_k h_{0,\text{Executed},k}(t) \exp(\beta_{\text{Executed}}^T (MQLP + BSID + MKD1 + MKD2 + STKV + TURN))$$

$$h_{\text{Cancelled}}(t|x) = \sum_k h_{0,\text{Cancelled},k}(t) \exp(\beta_{\text{Cancelled}}^T (MQLP + BSID + MKD2 + SZSD + STKV + TURN))$$

$$h_{\text{Partial}}(t|x) = \sum_k h_{0,\text{Partial},k}(t) \exp(\beta_{\text{Partial}}^T (MKD1 + MKD2 + SZSD + TURN))$$

Here, I stratified on the bid-ask spread, which takes values 0.01, 0.02, and 0.03, in order to account for the variable and at the same time deal with its tendency to vary with time. Table 3 summarizes the estimated coefficients for the Cox proportional hazards model for each failure type for buy limit orders. Recall that $MQLP$ measures the distance between the midquote price and the limit order price. $BSID$ is an indicator equal to 1 if the most recent transaction was initiated from the sell side of the market and -1 if it was initiated from the buy side. $MKD1$ estimates the number of shares with higher priority on the same side of the market, $MKD2$ estimates the liquidity demanded on the opposite side of the market, and $SZSD$ measures the liquidity demanded by the limit order. $STKV$ estimates short term shifts in trading activity, while $TURN$ estimates absolute recent trading activity. Coefficients statistically significant at the 0.05 level based on the Wald test are identified with asterisks. Note that taking the exponential of the coefficients results in the multiplicative effect each coefficient has on the hazard function.

The coefficients for $MQLP$ are significantly negative for cancelled orders but significantly positive for executed orders. Therefore, increasing $MQLP$ and hence either decreasing the buy limit price or experiencing an increase in the midquote price are
Table 3: Model Estimates for Buy Orders

<table>
<thead>
<tr>
<th></th>
<th>Executed Orders</th>
<th>Cancelled Orders</th>
<th>Partially Executed Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-statistic</td>
<td>Coefficient</td>
</tr>
<tr>
<td>MQLP</td>
<td>0.32</td>
<td>4.42*</td>
<td>-0.05</td>
</tr>
<tr>
<td>BSID</td>
<td>-0.04</td>
<td>-3.13*</td>
<td>-0.07</td>
</tr>
<tr>
<td>MKD1</td>
<td>-0.03</td>
<td>-6.97*</td>
<td>0.001</td>
</tr>
<tr>
<td>MKD2</td>
<td>0.16</td>
<td>7.49*</td>
<td>0.03</td>
</tr>
<tr>
<td>SZSD</td>
<td>-0.0004</td>
<td>-0.09</td>
<td>0.005</td>
</tr>
<tr>
<td>STKV</td>
<td>0.50</td>
<td>9.08*</td>
<td>-0.05</td>
</tr>
<tr>
<td>TURN</td>
<td>0.33</td>
<td>18.14*</td>
<td>0.13</td>
</tr>
</tbody>
</table>

associated with an increased execution hazard rate but a decreased cancellation hazard rate. The coefficient for BSID is negative for both executions and cancellations but is not significant for partially executed orders.

MKD1 is significant and negative for both executed and partially executed orders. Therefore, as expected, the execution and partial execution hazard rates decrease as the number of shares with higher priority on the buy side of the market increases. Similarly, as the market depth on the opposite side of the market increases, which is captured by MKD2, the execution, partial execution, and cancellation hazard rates increase as indicated by the positive coefficients for these three variables.

For both cancellations and partial executions, the coefficients for SZSD are slightly positive; therefore, the hazard rate is increased by factors of 1.01 and 1.08, respectively, with the size of the buy limit order when the limit order price is less than the sell price or when the limit order price is equal to the sell price but the limit order size is greater than the market depth on the sell side of the market. The fact that the hazard rates are not very sensitive to the size of the limit order is consistent with other literature.

Shifts in relative activity, as indicated by STKV, have a positive effect on the execution hazard but a negative effect on the cancellation hazard. Therefore, on average, increases in the relative activity lead to increases in the execution hazard rate but decreases in the cancellation hazard rate. TURN, however, is positive for the execution, cancellation, and partial execution hazard functions, indicating that
shifts in the absolute activity in the queue result in increased hazard rates for each of the three events by factors of 1.397, 1.153, and 1.40, respectively.

Some of these results are similar to the results found in Chakrabarty et al., but quite a few differ. In particular, they found that $MQLP$ for executed orders was negative, $MKD2$ for cancelled orders was negative, $SZSD$ for both executed and cancelled orders was positive, and $STKV$ for cancelled orders was positive.

5.2 Sell Orders

The procedure applied for buy limit orders was repeated for added sell orders. Once again, the Newton-Raphson iterated procedure was applied to the partial likelihood for sell limit orders, and the Wald test was performed to identify significant covariates. The three models are given by

$$h_{\text{Executed}}(t|x) = \sum_k h_{0,\text{Executed},k}(t) \exp\left(\beta^T_{\text{Executed}} (MQLP + MKD1 + MKD2 + TURN)\right)$$

$$h_{\text{Cancelled}}(t|x) = \sum_k h_{0,\text{Cancelled},k}(t) \exp\left(\beta^T_{\text{Cancelled}} (MQLP + BSID + MKD1 + MKD2 + SZSD + STKV + TURN)\right)$$

$$h_{\text{Partial}}(t|x) = \sum_k h_{0,\text{Partial},k}(t) \exp\left(\beta^T_{\text{Partial}} (MQLP + BSID + MKD1 + MKD2 + SZSD + TURN)\right)$$

Once again, I stratified on the bid-ask spread, which takes values 0.01 and 0.02. The coefficient for the model for sell orders can be found in Table 4. As before, coefficients that were not significant at the 0.05 level have been excluded from the model.

The positive coefficients for the executed and partially executed $MQLP$ variable indicate that increasing the distance between the limit price and the midquote price increases the hazard rate. Also, the instantaneous probability of cancellation is decreased when this distance is increased as indicated by the negative coefficients for $MQLP$. The results for executed and cancelled orders are similar to those found for buy orders. $BSID$ was significant and positive for cancelled orders but negative for
partially executed orders.

The slightly positive coefficients for $MKD1$ for the three cause-specific hazard functions indicate that the hazard rates increase modestly as the market grows in depth on the same side of the market. However, as the market depth on the opposite side of the market increases, the hazard rates decrease for the three failure types as suggested by the negative coefficients for $MKD2$.

$SZSD$ is negative for cancelled orders but positive for partially executed orders. Thus, as the liquidity demanded by the limit order increases, the instantaneous probability of partial execution increases but that of cancellation decreases.

The hazards rates for executed, cancelled, and partially executed orders are increased by 44%, 13%, and 47%, respectively, as the absolute trading activity increases, as measured by the variable $TURN$. However, the relative trading activity measure $STKV$ is significant only for cancelled orders; the coefficient is negative, indicating that the hazard rate decreases by a factor of 0.88 with increases to relative trading activity.

<table>
<thead>
<tr>
<th>Table 4: Model Estimates for Sell Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executed Orders</td>
</tr>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>MQLP</td>
</tr>
<tr>
<td>BSID</td>
</tr>
<tr>
<td>MKD1</td>
</tr>
<tr>
<td>MKD2</td>
</tr>
<tr>
<td>SZSD</td>
</tr>
<tr>
<td>STKV</td>
</tr>
<tr>
<td>TURN</td>
</tr>
</tbody>
</table>

5.3 Goodness-of-Fit

To assess the goodness-of-fit for the models specified for buy and sell orders, plots of the Cumulative Hazard Function of the Cox-Snell residuals were used. To understand the idea behind these plots, consider the survival times for each of the orders, $T_i$, and the associated survivorship functions $S_i(t)$. Since
\[ S(t) = P(T \geq t) = U \]

we have that

\[ T = S^{-1}(U) \]

so that \( U \) is a Uniform(0, 1) random variable. Taking the negative of its natural logarithm, therefore, results in a unit exponential random variable. However, we also have that \( H(t) = -\ln S(t) \); hence, the estimated cumulative hazard function for each order at the time of failure or censoring should behave like a unit exponential random variable if the model is correctly specified.

Figure 2: Cumulative Hazards of Cox-Snell Residuals for Buy Orders

Figures 2 and 3 plot the Cumulative Hazard Functions of the Cox-Snell residuals for the buy and sell order models, respectively. While the model for buy orders provides a better fit for the data, both models could be improved through the addition of and modification to the variables used to fit the data.

In particular, the inclusion of time-dependent variables might improve the models, as indicated by plots 6 through 11, which are diagnostic plots to check for the presence
of proportional hazards; in these figures, the scaled Schoenfeld residuals for each variable are shown. For executed and cancelled orders, these plots indicate that the model might be improved through the inclusion of variables that capture the overabundance of actions taken a fixed time intervals. For example, figure 4 shows the fitted least squares line of the scaled Schoenfeld residuals along with the residuals for the cancelled buy orders $SZSD$ variable. The residuals plot shows increased activity at about 10 seconds and 3 minutes, and this activity greatly influences the fitted least squares line. This phenomenon occurs for quite a few variables. As another example, the residuals for the $STKV$ variable of executed buy orders are well-distributed (figure 5), but we again see a peak in the fitted least squares line for the residuals of this variable. On the other hand, figures 8 and 11, which show the residuals for partially executed orders, indicate that the proportional hazards assumption is approximately met for the statistically significant variables.

The presence of nonlinearity, or incorrect specification of the functional form of the parameters, is also a concern when assessing goodness-of-fit for a model. We can check for nonlinearity by plotting the covariates against the martingale residuals, as shown in figures 12 through 17 in the appendix. These plots indicate that, in general,
nonlinearity among the covariates is only minor; most deviations from linearity are driven by few observations.

6 Conclusion

In the analysis of Microsoft stocks traded on October 26, 2006, on the Island ECN, the Cox proportional hazards model with independent competing risks was applied to study the times until execution, cancellation, and partial execution of the limit orders. For executed orders, I find that the execution hazard rate on the buy side of the market increases when there is a decrease in price, there is an increase in market depth on the opposite side of the market, or there is an increase in trading activity; for sell orders, an increase in the execution hazard rate occurs when there is an increase in the limit order price relative to the midquote price, increased depth on the same side of the market, or an increase in the absolute trading activity. The cancellation hazard rate for buy orders increases when there is an increase in market depth on the opposite side of the market, an increase in the liquidity demanded by the limit order, and an increase in absolute trading activity and for sell orders when there is an increase in market depth on the same side of the market and an increase in absolute
trading activity. Finally, the instantaneous probability of partial execution increases for buy orders with increases in market depth on the opposite side of the market, increases in the liquidity demanded by the limit orders, and increases in absolute trading activity and for sell orders with increases in the limit order price, increases to the market depth on the same side of the market, increases in the liquidity demanded by the limit orders, and increases in absolute trading activity for sell orders.

Future work related to this study includes the inclusion of time-dependent variables in the model as well as comparing the Cox proportional hazards model with other models to explore the best fit for the data. I would also like to study multiple days across multiple stocks in addition to looking at multiple ECNs.

7 References


Chakrabarty, Bidisha, Zhaohui Han, and Xiaoyong Zheng, 2006, A Competing Risk Analysis of Executions and Cancellations in a Limit Order Market, Working Paper,
Indiana University.


A.1 Figures

Figure 6: Proportional Hazards Test for Executed Buy Orders
Diagnostic Plots showing the scaled Schoenfeld residuals for each variable significant among executed orders. For each variable, the top plot shows the fitted least squares line with the residuals suppressed; the plot with the residuals is on the bottom. Figures 7 through 11 show the same proportional hazards tests for cancelled and partially executed buy orders and executed, cancelled, and partially executed sell orders.
Figure 7: Proportional Hazards Tests for Cancelled Buy Orders

- MOLP for Cancelled Buy Orders
- BBS for Cancelled Buy Orders
- MEO for Cancelled Buy Orders
- MCOJ for Cancelled Buy Orders
- S250 for Cancelled Buy Orders
- STKV for Cancelled Buy Orders
- TURN for Cancelled Buy Orders
Figure 8: Proportional Hazards Test for Partially Executed Buy Orders
Figure 9: Proportional Hazards Tests for Executed Sell Orders
Figure 10: Proportional Hazards Tests for Cancelled Sell Orders

MQLP for Cancelled Sell Orders

MK01 for Cancelled Sell Orders

MKQ2 for Cancelled Sell Orders

S2SD for Cancelled Sell Orders

STKU for Cancelled Sell Orders

TURN for Cancelled Sell Orders
Figure 11: Proportional Hazards Tests for Partially Executed Sell Orders

MQLP for Partially Executed Sell Orders

BMD for Partially Executed Sell Orders

MWD1 for Partially Executed Sell Orders

MWD2 for Partially Executed Sell Orders

STHD for Partially Executed Sell Orders

STHV for Partially Executed Sell Orders

TWM for Partially Executed Sell Orders
Figure 12: Nonlinearity diagnostic plots for executed buy orders
Plots of the martingale residuals with fitted local linear regression line for cancelled buy orders. Similar plots for executed and partially executed buy orders and executed, cancelled, and partially executed sell orders can be found in figures 13 through 11.
Figure 13: Nonlinearity diagnostic plots for cancelled buy orders

Figure 14: Nonlinearity diagnostic plots for partially executed buy orders
Figure 15: Nonlinearity diagnostic plots for executed sell orders

Figure 16: Nonlinearity diagnostic plots for cancelled sell orders
Figure 17: Nonlinearity diagnostic plots for partially executed sell orders