Essays in Market Microstructure

by

Arnout M. Eikeboom

B.A. (Econometrics), Free University, Amsterdam (1984)
M.A. (Operations Research), Free University, Amsterdam (1986)

Submitted to the Sloan School of Management
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy in Management

at the

Massachusetts Institute of Technology

December 1993

Signature of Author____
Sloan School of Management
December 22, 1993

Certified by____
Andrew W. Lo
Professor of Finance
Thesis Supervisor

Accepted by____
Birger Wernerfelt
Chairman, Ph.D. Program

Essays in Market Microstructure

by

Arnout M. Eikeboom

Submitted to the Sloan School of Management
on December 22, 1993
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Management

Abstract

This thesis contains four essays in Market Microstructure, an area of Finance that is focused on the details of the trading process of securities in asset markets.

Chapter 2 is an attempt to model the dynamics of transaction prices on the NYSE, using a state space modeling approach. Maximum likelihood is employed to estimate the model, and an illustrative forecasting experiment suggests that the resulting predictions of the model are informative.

In Chapter 3, we investigate the independence of sequences of buy and sell orders which result in transactions of individual stocks on the Paris Bourse. Precise transaction data are recovered from market-activity information which is continuously disseminated electronically by the fully automated order execution system (CAC) of the Paris Stock Exchange. Using exact distribution theory for runs we find highly significant *positive* dependence in many daily sequences of buy and sell orders for individual stocks. Likelihood-ratio tests based on Markov chain models confirm the conclusions of the runs tests. That is, for many series the assumption of an independent buy-sell process can be rejected in favor of Markov dependence. Moreover, for trading in some stocks on certain days, Markov dependence can also be rejected in favor of higher-order dependence.

Despite the fact that bid-ask spreads for large NYSE stocks typically take on at most 3 or 4 different values, they exhibit considerable variability in the course of the trading process. In Chapter 4 of this thesis I explore these movements and relate them to economic variables suggested by various theories. The dynamic behavior of the bid-ask spread is shown to be positively related to trading volumes and stock-price volatility in the sense that large trades and volatile prices tend to foreshadow an increase of the bid-ask spread. Transactions executing at the midpoint of the quotes have a minimal effect compared to transactions closer to the bid or ask, a finding consistent with implications of adverse-selection theories. There is also some mild evidence that spreads decrease as the time between trades increases.

Chapter 5 studies the liquidity effects associated with stock splits, focusing on measures of trading activity, bid-ask spreads, and depths. Using transaction data for NYSE companies that split their stocks by 2-for-1 or greater in the two years 1988 and 1991, I show that percentage bid-ask spreads increase significantly after the split and that depths remain unchanged. Thus if liquidity is measured by spread and depth, liquidity decreases after the split. As far as trading activity measures are concerned, I show that for the stocks in my sample the daily number of transactions increases after the split, but total daily dollar volume and split-adjusted share volume remain unchanged. Therefore, dollar volume and split-adjusted share volume of a typical transaction decline significantly after the split. Finally, an application of the ordered probit model reveals that stock splits decrease liquidity in the sense that large trades tend to widen the percentage spread more after the split than before the split.

Thesis Supervisor: Andrew Lo

Title: Professor of Finance

Acknowledgments

One of the greatest privileges of being at MIT is the opportunity to interact with the world's top scholars. I have been fortunate to find three exceptional individuals willing to serve on my thesis committee. Andrew Lo agreed to be my thesis advisor when I switched from the Management Science Area to the Finance Group. His guidance, wisdom, and encouragement have been crucially important to me and I am most thankful for his patience during periods when I was insufficiently focused on finishing my thesis. I had a unique opportunity to learn about continuous-time finance when Professor Robert Merton agreed to hire me as a research assistant for his book. It was extremely satisfying to work for Professor Merton and I feel very privileged to have been so close to this brilliant scholar. Perhaps it was Professor John Cox who planted the first seeds for my final conversion to Finance when I took his options course in my first year at MIT. I still consider this as one of the best courses I have ever taken: Professor Cox's insights and explanations of the subject matter are truly superior.

Prior to joining the Finance Group at Sloan I worked in the area of Operations Research and Statistics. During this time I had a very fruitful and stimulating collaboration with Peter Kempthorne—chapters 2 and 3 of this thesis are coauthored with him. It was a joy to work with someone who is not only exceptionally smart and creative, but always saw the humorous side of any given situation. The fact that I subsequently switched to Finance does not reflect on Peter, but on the fact that my interests had changed.

I would like to acknowledge Professor Zvi Bodie from Boston University for his role in my journey. His candid comments about the field of Finance and the field of Operations Research catalyzed me into thinking seriously about switching from OR to Finance.

I thank Andrew Alford, Craig MacKinlay, and Jean-Luc Vila for helpful comments on various parts of this thesis. John Heaton's interest in computers greatly facilitated a smooth execution of the substantial computational requirements of my work. I am grateful to Murray Teitelbaum of the NYSE for arranging a visit to the Exchange and to NYSE specialists Alan Safir and William Johnson for showing me the trading process from their respective posts. I thank Charles d'Autichamp and his group, *La Direction des Interventions de Marché et de la Trésorerie Ingénierie* of the Banque Nationale de Paris for stimulating discussions and counsel during the research on the Paris Bourse. Research support from the International Financial Services Research Center at MIT and from the Banque Nationale de Paris is also gratefully acknowledged.

I met some of my best friends during my early period at MIT, when I still had an office in building E-40. Achilles Venetoulias, an unusually talented statistician with broad interests in applied problems, made this a very pleasant work environment. His perspective and mature vision on academic and non-academic issues alike have been quite important to me. C.S. Venkatakrishnan, better known as Venkat, initially considered me a strange and secretive character. He only changed his mind and started talking to me after he saw me play Blackjack on a Macintosh. We have been good friends ever since. My most recent debt to him is for his help with the formatting of this thesis. I would also like to acknowledge my Blackjack friends, who enabled me to apply my interest in probability and stochastic processes in a truly exciting way.

Past and current fellow students have contributed a great deal to my enjoyment of being at MIT. In particular, I would like to thank Chyi-Mei Chen, Tim Crack, Olivier Ledoit, Ioannis Pierides, Sung-Hwan Shin, and Jim Walton. Past and current secretaries have also played a special role during my stay at MIT. I have known Debbie Hannon as Merton's secretary at MIT and at Harvard. I am grateful to Rosie Lynn for everything she has done for me.

Various other people at the Sloan School have played special and important roles to me. I distinctly remember the day that I received a phone call in Amsterdam from Jim Orlin regarding my admission to Sloan. I have known Jim as a teacher, a mentor, and a facilitator. His creativity and willingness to help have always been very effective. I am sorry that he did not stay on as Chairman of the Ph.D. program for another couple of months so that he could sign my thesis. The Sloan Ph.D. program has an excellent Coordinator in the person of Sharon Cayley. She manages the Ph.D. program with great flair and efficiency. Perhaps even more important is that she clearly has the students' best interests in mind. I am extremely grateful to John Maglio, who has always been prepared to solve my computer-related problems in a good-humored way. John's importance to the Ph.D. students cannot be overemphasized.

When I was still a student in the Netherlands, my wonderful advisor Professor Henk Tijms encouraged me to pursue Ph.D. studies in the United States. Alexander Rinnooy Kan was also very helpful when I was exploring the various Ph.D. programs in the US. I thank these two individuals for making my MIT adventure possible.

Ashdown House has been my home throughout my stay at MIT. I thank Vernon and Beth Ingram for their substantial contribution in making Ashdown an enjoyable place to live. An extracurricular activity that I greatly enjoyed was classical music—it was an enriching experience to be a member of the MIT Chamber Music Society.

I would like to thank Guillaume for his help and stimulating attitude, and I thank Eva for making my final stretch at MIT so colorful. Last but not least I thank my parents Rogier and Len for their unconditional support and interest in my well-being throughout my years at Sloan, as well as my two sisters Jolijn and Frederike. I am very fortunate to have such a loving and caring family.

To my parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Market Microstructure is an area of Finance that is focused on the details of the trading process of securities in asset markets. As the title "Essays in Market Microstructure" suggests, the goal of this thesis is to increase our understanding of the way markets operate at the microscopic level by examining the nature of the trading and informational processes at the finest level for which data are available.

Data collected from actual markets play an important role in this thesis—the thesis has a distinct empirical flavor. Indeed, one of the reasons that market microstructure is such an exciting field is that detailed transactions data have become available only relatively recently.

The coverage of markets investigated here is fairly broad. Most of the thesis focuses on the New York Stock Exchange (NYSE), but there is also a chapter devoted to the operation of the Paris Stock Exchange. Unlike the NYSE, the Paris Bourse does not have specialists who maintain markets in individual stocks. Moreover, the information available about the state of the market allows a more detailed investigation of many issues of interest to finance scholars.

The chapters comprising this thesis reflect my own evolution at MIT from a student in Operations Research and Statistics to a Finance student. Chapters 2 and 3 were conceived during the former period and it is therefore not surprising that they have a distinct "statistical" flavor. Chapters 4 and 5 were written later and focus more on issues whose financial-economic significance is evident, addressing such topics as the intraday behavior of the bid-ask spread and the liquidity effects associated with stock splits.

A brief outline of the thesis is as follows. Chapter 2 is an attempt to model the dynamics of transaction prices on the NYSE in a manner that explicitly incorporates some important institutional features of the trading process, such as the presence of a bid-ask spread, discreteness of prices, and irregularly-spaced transaction intervals. The framework used here is a so-called state space model, which is widely used in the engineering and statistical literature. We posit the existence of (unobservable) "states," which can be interpreted as the "true price" of a stock that would prevail in a frictionless market. By specifying the probabilistic dependence of the actual (observed) transaction prices on the underlying states and the dynamics of those states, we use maximum likelihood to estimate this state space model for a small sample of stocks. An illustrative forecasting experiment suggests that the resulting predictions of the model are informative.

In Chapter 3 we investigate the independence of sequences of buy and sell orders which result in transactions of individual stocks on the Paris Bourse. Precise transaction data are recovered from market-activity information which is continuously disseminated electronically by the fully automated order execution system (CAC) of the Paris Stock Exchange. Using exact distribution theory for runs we find highly significant *positive* dependence in many daily sequences of buy and sell orders for individual stocks. That is, given that the last order for a particular stock was a buy order, the probability that the next order will also be a buy order is higher than the unconditional probability of a buy order. A similar conclusion holds for sell orders. Likelihood-ratio tests based on Markov chain models confirm the conclusions of the runs tests. That is, for many series the assumption of an independent buy-sell process can be rejected in favor of Markov dependence. Moreover, for trading in some stocks on certain days, Markov dependence can also be rejected in favor of higher-order dependence. Appendix A contains some technical details concerning the runs test.

Despite the fact that bid-ask spreads for large NYSE stocks typically take on at most 3 or 4 different values, they exhibit considerable variability in the course of the trading process. In Chapter 4 of this thesis I explore these movements and relate them to economic variables suggested by various theories. Using transactions data for the 30 stocks included in the Dow Jones Industrial Average in 1988, the dynamic behavior of the bid-ask spread is shown to be positively related to trading volumes and stock-price volatility in the sense that large trades and volatile prices tend to foreshadow an increase of the bid-ask spread. Transactions executing at the midpoint of the quotes have a minimal effect compared to transactions closer to the bid or ask, a finding consistent with implications of adverse-selection theories. There is also some mild evidence that spreads decrease as the time between trades increases.

Predictions generated by the estimation technique used in Chapter 4, the ordered probit model, are compared with predictions produced by Ordinary Least Squares (OLS) in an out-of-sample forecasting experiment of bid-ask spreads for the sample of 30 stocks. If performance is measured by root mean square prediction error, the difference between OLS and ordered probit is insignificant. However, if performance is measured using hit rate, the fraction of spreads forecasted correctly, I show that ordered probit tends to do better than OLS for stocks whose (absolute) bid-ask spread is relatively large. This is explained by noting that for such stocks discreteness and nonlinearity of the spread distribution are important—phenomena which can be captured using ordered probit but not using OLS. A detailed description of the data used in this chapter is given in Appendix B.

Finally, Chapter 5 studies the liquidity effects associated with stock splits, focusing on measures of trading activity, bid-ask spreads, and depths. Using transaction data for NYSE companies that split their stocks by 2-for-1 or greater in the two years 1988 and 1991, I show that percentage bid-ask spreads increase significantly after the split. No clear pattern is found for the direction of change of dollar depths, the dollar value of the number of shares available at the quoted bid and ask prices. Thus if liquidity is measured by spread and depth, liquidity decreases after the split. As far as trading activity measures are concerned, I show that for the stocks in my sample the daily number of transactions increases after the split, but total daily dollar volume and split-adjusted share volume remain unchanged. Therefore, dollar volume and split-adjusted share volume of a typical transaction decline significantly after the split. Finally, an application of the ordered probit model reveals that stock splits decrease liquidity in the sense that large trades tend to widen the percentage spread more after the split than before the split.

In addition to the main findings regarding stock splits and liquidity, I also document two interesting empirical regularities, a buy-sell asymmetry and a property of closing quotes. The buy-sell asymmetry is that transaction volumes and quoted depths are such that buys are smaller relative to ask depths than sells relative to bid depths. In this context, "buys" ("sells") are transactions that were (most likely) initiated by a buyer (seller). As far as closing quotes is concerned, I show in Appendix C that closing spreads are on average narrower than spreads quoted at the very end of the trading day. In other words, measuring a stock's spread using closing quotes induces a downward biased estimate of the real spread.

Of course, there is still much to learn empirically about the trading process in actual markets. As highly detailed and comprehensive datasets continue to become available, existing theories will become more refined and new theories will be developed. It is my hope that the research documented in this thesis is a useful step in the continuing process of increasing our understanding of the intricacies of financial markets.

Chapter 2

Modeling the Dynamics of Transaction Prices

2.1 Introduction

Traditional models for the dynamics of stock prices are discrete-time, univariate and cross-sectional time series based on monthly, weekly, and daily data with continuous-valued sample spaces.[1] With intraday data at the transaction level, these models are inappropriate. We can no longer assume that prices change in a continuous fashion. Instead prices trade at discrete increments of monetary value, a *tick*, which is \$0.125 for the New York Stock Exchange (NYSE). Also, stocks trade at unequally-spaced intervals of time which can vary considerably over the trading day.

To model the dynamics of stock prices at the transaction level it is critical to understand the basic nature of the trading process. The NYSE is a continuous agency auction market with a specialist designated by the Exchange for each stock. This means that (i) trading takes place continuously in time, (ii) trading on the floor of the exchange is limited to exchange members who trade for themselves or as agents for non-members, i.e., institutions and private investors, and (iii) each stock has a specialist or designated agent. The specialist supplies liquidity and information to the market by announcing cur-

[1]This chapter is reproduced from the paper "Modeling the Dynamics of Transaction Prices on the NYSE in a Bayesian State-Space Filtering Framework," coauthored with Peter J. Kempthorne. This paper also appeared in *Computing Science and Statistics* (1992), C. Page and R. LePage, eds., Springer-Verlag, New York.

rent buying (*bid*) and selling (*ask*) prices (*quotes*) to the market. These prices are good for limited numbers of shares (also announced by the specialist). These shares offered or demanded are either limit orders left with the specialist by other agents (floor traders) or reflect orders by the specialist for her own account. The latter instance occurs when the specialist fulfills her obligation to the Exchange by reacting to a temporary shortage of either buyers or sellers and trading against the trend of the market.[2]

The Bayesian state-space filtering framework, well known in the engineering and statistical literature, is suited to modeling and predicting transaction-price time series. Kitagawa (1987) discusses the general framework we apply here. The state space models we introduce in section 2.2 explicitly allow for discrete outcomes and irregularly-spaced observation intervals.

2.2 State Space Model

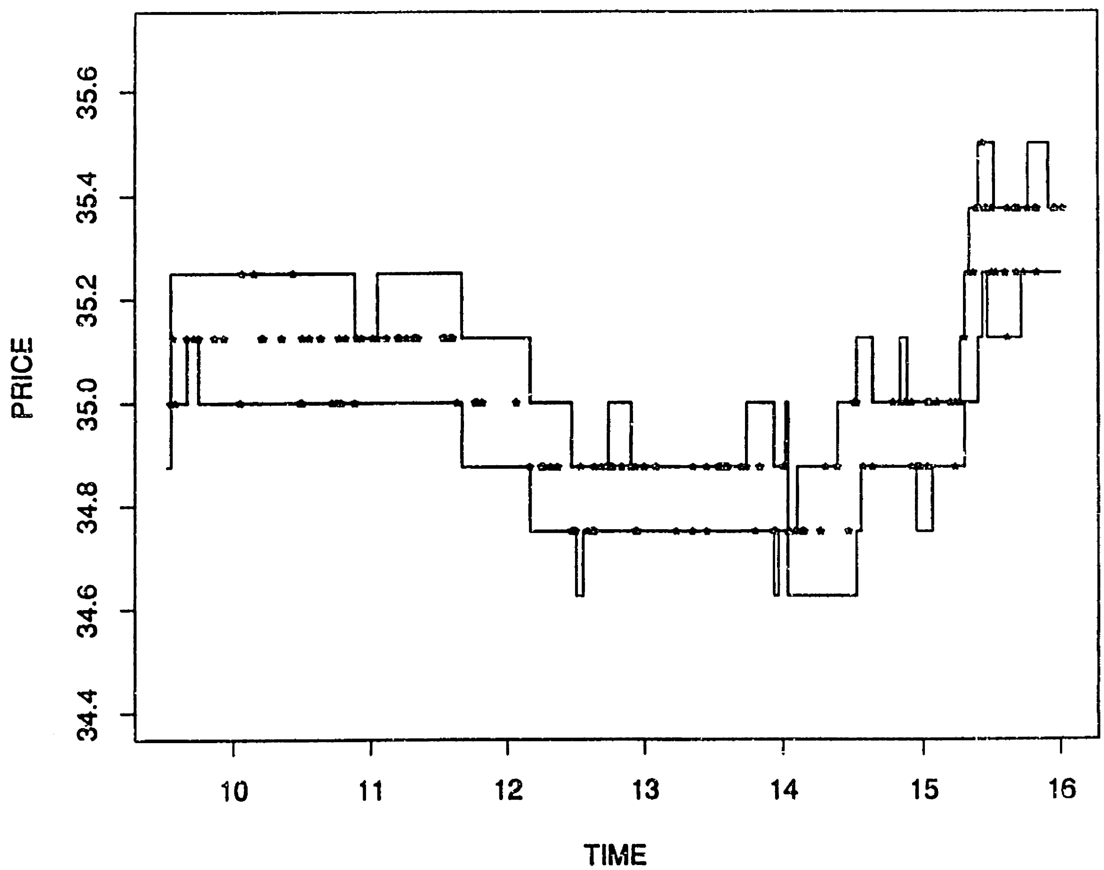
Suppose there are N transactions in a particular stock on a given day. Then we define, for $j = 1, \ldots, N$, the following variables:

$$
\begin{aligned}
P_j &= \text{Price of the } j\text{th transaction} \\
T_j &= \text{Time of the } j\text{th transaction} \\
A_j &= \text{Ask quote just before the } j\text{th transaction} \\
B_j &= \text{Bid quote just before the } j\text{th transaction.}
\end{aligned}
$$

Here the ask quote is the price at which the specialist is willing to sell and the bid quote is the price at which she is willing to buy one share. Naturally, $A_j > B_j$. The difference $A_j - B_j$ between the ask price and the bid price is called the *spread*. Typically, the transaction price lies between the quotes, i.e., $B_j \leq P_j \leq A_j$. See Figure 2.1, a plot of the transaction data for Southwestern Bell Company on April 22, 1988.

When market buy and sell orders arrive randomly, execution prices tend to bounce between the ask and bid prices, a phenomenon described as *price jiggling* by Working

[2]For a general discussion of the operation of the New York Stock Exchange, see the New York Stock Exchange Annual Report 1989, pp 14–25.



(Note: asterisks indicate transactions and lines connect bid prices and ask prices)

Figure 2.1: Transaction Data SBC (4/22/88)

in the 1950s. Such price reversals lead to negative autocorrelation in the time series of observed stock market prices; see Niederhoffer and Osborne (1966), Garman (1976), and Roll (1984).

We formulate a state space model that explicitly incorporates the presence of the bid-ask spread and its effect on the trading process of a given stock. Suppose the transactions for a given day of a stock are at prices $P_1, P_2, \ldots, P_N$ and times $T_1 \leq T_2 \leq \cdots \leq T_N$, respectively. We posit the existence of (unobservable) "states" $\mu_1, \mu_2, \ldots, \mu_N$, where

$$\mu_j = \text{"True price" of the stock just before the } j\text{th transaction.}$$

One can think of the "true price" of a stock as the price that would prevail in a frictionless market. See Kempthorne and Marsh (1990) for additional discussion motivating the model and its financial-economic interpretation.

To complete the specification of a state-space model for the transaction prices, we must specify: (i) the probabilistic dependence of every P_j on μ_j, and (ii) the dynamics of the states μ_j. These are the topics of the next two subsections.

2.2.1 Observation-State Relation

We model the true price μ_j as the median parameter of a probability distribution F_j. For convenience, we assume that F_j is a lognormal distribution with a non-changing variance parameter σ, the price volatility. We can think of F_j as the distribution of reservation prices of active market participants in a market with transaction costs.

If P_j^* denotes the reservation price of the market participant initiating the jth transaction, it is reasonable to assume that $P_j^* \geq P_j$ when $P_j = A_j$. Otherwise, there would be no rational incentive to pay A_j, the price of a market buy order. Similarly, we assume that $P_j^* \leq P_j$ when $P_j = B_j$. In these two cases, we interpret the jth transaction as having been initiated by a buyer and a seller, respectively. When the spread is more than one tick, the transaction price can be between the quotes. Then, we make no assumption whether it was initiated by a buyer or by a seller. While we shall not pursue the issue here, it is common practice to associate buying or selling interests as motivating trades between the quotes applying a "tick rule." The sign of the last non-zero price change is used to determine whether trades are "buys" or "sells."

Assuming that all transactions occur at either the bid price or at the ask price or at the midquote, we thus have a multinomial conditional distribution $L(P_j \mid \mu_j, \sigma)$ defined in terms of the lognormal F_j as follows:

Event	Likelihood: $L(P_j \mid \mu_j, \sigma)$
$P_j = A_j$	$F_j(P_j^* \geq P_j)$
$P_j = B_j$	$F_j(P_j^* \leq P_j)$
$P_j = (A_j + B_j)/2$	$F_j(B_j < P_j^* < A_j)$

This completes the specification of the observation-state relation.

2.2.2 State Dynamics of μ_j

We now address the issue of the dynamics of the underlying states over time. In the context of the NYSE, we assume that $\mu_1, \mu_2, \ldots, \mu_N$ are the values of a continuous-time process $\{\mu(t),\ 9{:}30\text{am} \leq t \leq 4{:}00\text{pm}\}$ over the trading day at the transaction times T_j so that $\mu_j = \mu(T_j)$.

We further assume that $\mu(\cdot)$ is a Brownian motion process on some time scale $\tau(\cdot)$. Thus $\mu(t)$ is a subordinated stochastic process and $\tau(\cdot)$ is the "operational" time scale. Subordinated stochastic processes are not new to the finance literature; see references and discussion in Kempthorne and Marsh (1990).

Formally, we define the dynamics of a new process $d\mu^*$ as a geometric Brownian motion on the log scale:

$$d \log \mu^*(\tau) = \lambda\, dW(\tau),$$

where dW are the increments of a standard Wiener process, i.e., $W(\tau') - W(\tau) \approx N(0, \tau' - \tau)$ for $\tau' \geq \tau$. The volatility parameter λ measures the degree of variability of the process $\mu^*(\cdot)$ and is referred to as the "process volatility." The connection between the processes $\mu(\cdot)$ and $\mu^*(\cdot)$ is established by specifying $\mu(T_j) = \mu^*[\tau(T_j)]$.

We consider three choices for the operational time scale: (i) Minutes, $\tau(T_j) = T_j$,

(ii) Transactions, $\tau(T_j) = j$, and (iii) Share Volume, $\tau(T_j) = \sum_{i \leq j} V_i$, where V_i is the volume associated with the ith transaction.

2.3 Model Specification

A complete specification of the state-space model calls for (i) the identification of the optimal operational time scale τ and (ii) estimation of the parameters λ and σ. While a fully Bayesian approach would provide (posterior) probability distributions for these quantities, we apply Good's Type-II maximum-likelihood approach, specifying the model through optimization of the marginal likelihood function.

For a given operational time scale τ, the (marginal) likelihood function of the model can be expressed as

$$L(\lambda, \sigma) = \int_{\mu_1} \cdots \int_{\mu_N} \prod_{j=1}^{N} L(P_j \mid \mu_j, \sigma)\, \pi(\mu_1, \ldots, \mu_N \mid \lambda)\, d\mu_1 \cdots d\mu_N,$$

where $L(\cdot \mid \mu_j, \sigma)$ is given in subsection 2.2.1 and $\pi(\cdot \mid \lambda)$ is the conditional distribution of the μ's given the state volatility λ.

The integral specified above looks forbidding. However, thanks to the Markov property of the stochastic process for the value distribution, this N-dimensional integral simplifies to an iterated sequence of N 1-dimensional integrals. Thus computation of the likelihood function is entirely feasible; see, e.g., Kitagawa (1987). While evaluating the likelihood function for given values of λ and σ, we can calculate several interrelated density functions for the value distribution. For $j = 1, \ldots, N$, we have the following densities:

- $p_j(\mu_j) = \pi(\mu_j \mid P_1, \ldots, P_{j-1})$
  (one-step-ahead prediction density)

- $f_j(\mu_j) = \pi(\mu_j \mid P_1, \ldots, P_{j-1}, P_j)$
  (filtered density)

- $s_j(\mu_j) = \pi(\mu_j \mid P_1, \ldots, P_{j-1}, P_j, \ldots, P_N)$,
  (smoothed density)

where $f_0(\cdot)$ is some starting density. Thus the filtered density "filters" in the current observation P_j starting from the corresponding one-step-ahead prediction density and the smoothed density considers the value distribution conditional upon all—including future—observations. The iterative process of calculating these prior and posterior distribution densities is called "state-space filtering." Figure 2.2 is a perspective plot of the filtered densities during the last half of the trading day for SBC on April 22, 1988.

Maximizing the likelihood involves performing, for each operational time scale, a simple two-dimensional nonlinear optimization for which efficient quasi-Newton methods are available.

In the next section we present some numerical results of the model. Prediction-based performance evaluation is then taken up in section 2.5.

2.4 Numerical Results

In Tables 2.1 and 2.2 we present the maximum likelihood results for one day of trading (April 22, 1988) for two stocks, Southwestern Bell Company (SBC) and Ford.

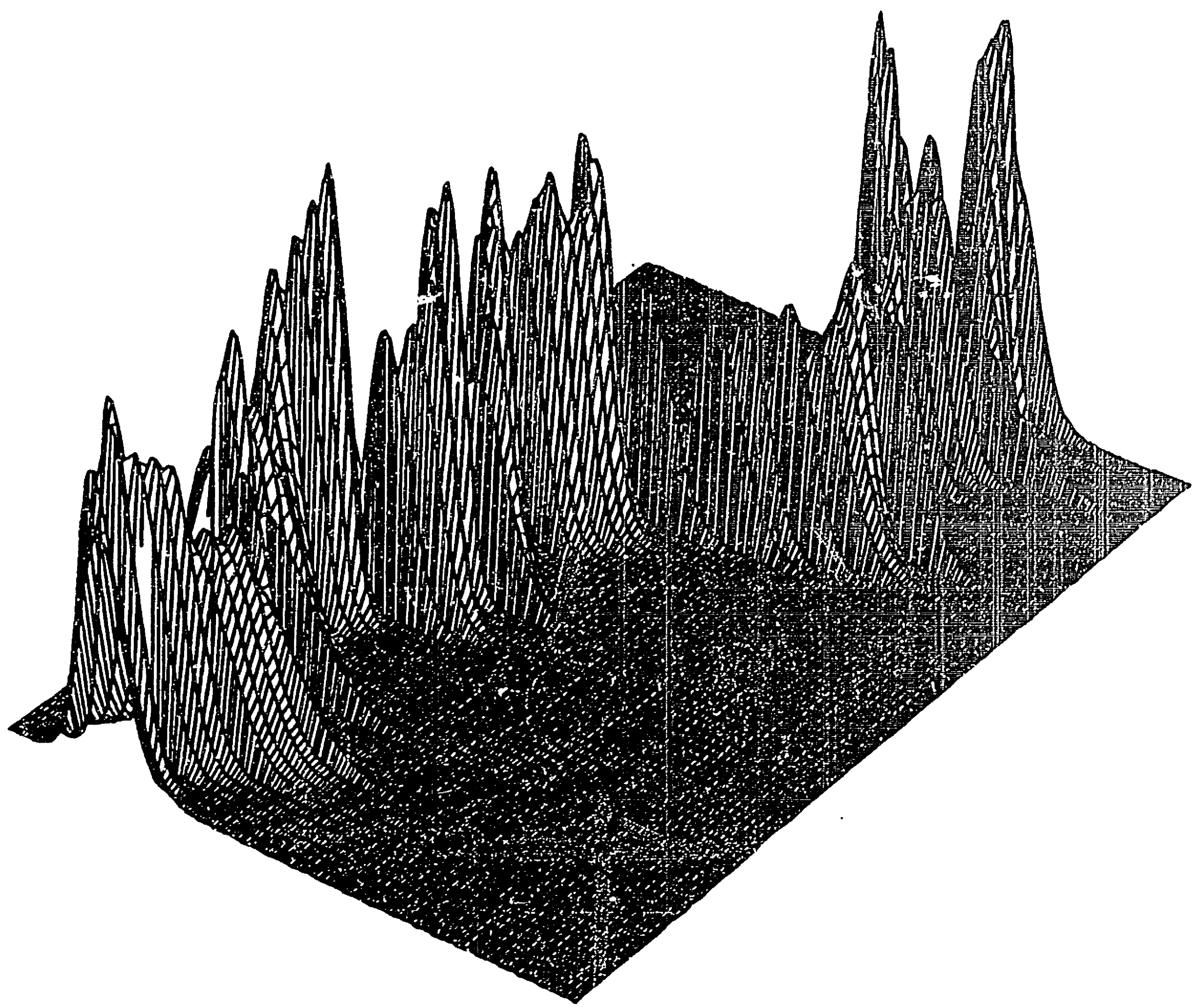
Time Scale (τ)	Log-likelihood	λ (%)	σ (%)
Minutes	−142.89	1.57	0.168
Shares	−136.91	1.46	0.151
Transactions	−141.30	1.39	0.161

Table 2.1: Likelihood Results SBC

Time Scale (τ)	Log-likelihood	λ (%)	σ (%)
Minutes	−259.67	2.20	0.179
Shares	−252.25	1.22	0.172
Transactions	−254.58	1.36	0.160

Table 2.2: Likelihood Results Ford

From these tables we see that the log-likelihood value for the share volume time scale is the highest of all three time scales, both for SBC (−136.91) and for Ford (−252.25). Also,



Axes: price (left to right), transaction index 82–162 (front to back)

Figure 2.2: Filtered Densities of SBC's "True Price"

the minutes time scale has the lowest value for the log-likelihood and the transaction time scale lies in between for each stock. We consider these differences significant. Assuming the uncertainty in the volatility parameters is negligible, and that each time scale is equally likely, a priori, the odds are about 400 to 1 that the share volume time scale is better than the minutes time scale for SBC, since $\exp(142.89 - 136.91) \approx 400$. For Ford, the corresponding odds are considerably more than 1000 to 1, since $\exp(259.67 - 252.25) \approx 1670$.[3]

Tables 2.1 and 2.2 also give the maximum-likelihood estimates of the value volatility σ and the process volatility λ. The estimates of σ are comparable across time scales and the two stocks—they measure the standard deviation of the (instantaneous) distribution of the logarithm of reservation prices for the stock, noting that variation on the logarithm scale is essentially equivalent to variation on a percent scale. For example, the fitted instantaneous dispersion in the reservation prices of active market agents for SBC is 0.168% under the minutes time scale. It is smallest under the share volume time scale, 0.151%. For Ford, the estimates of dispersion in reservation prices on the log/percent scale are of similar magnitudes—between 0.1% and 0.2%.

The estimates for λ are also comparable across operational time scale and stock. For each stock and operational time scale, λ measures the standard deviation of the daily change in the "true price" on the log/percent scale. The results indicate that the daily dispersion in "true price" was between 1% and 2% for each stock, regardless of the assumed time scale. For each stock, the minutes time scale was least parsimonious in terms of yielding the highest estimate of daily volatility of the underlying "true price."

2.5 Performance Illustration

Finally, we briefly discuss the predictive power of the state space models. Using the filtered density at a given time during the trading day, we can calculate the model's estimate $\hat{P}$

[3]Formal application of likelihood ratio tests is not possible because the models under the different time scales are not nested. Formal application of Bayes tests would require specifying a joint prior distribution for the volatility parameters under each time scale and evaluating the likelihood function at all points in the support of that distribution.

of the "true price" as the average value of the relevant filtered density. To measure how cheap or expensive the ask price is at the given time, we define the "ask premium" as the ask price minus the model's estimate $\hat{P}$.

On the small sample of stocks examined thus far, the ask premium discriminates well those times when the ask price is an offer to sell at a discount, that is when the "true" ask premium is low and subsequent prices might be expected to rise. Figure 2.3 illustrates this predictive power of the ask premium for the same SBC data displayed in Figure 2.1, using the share-volume maximum-likelihood specification of the filtered densities in the state space model. Plotted are subsequent changes in the ask price (one, two, or three steps ahead) against the ask premium (estimated from the corresponding filtered densities). The distribution of changes in the ask price clearly varies depending on the ask premium. The average ask premium is about one-half the typical spread of \$0.125. When the ask premium drops below this mean level, indicating that the market price to buy the stock is discounted according to the model, subsequent ask prices tended to be higher and *never* declined.

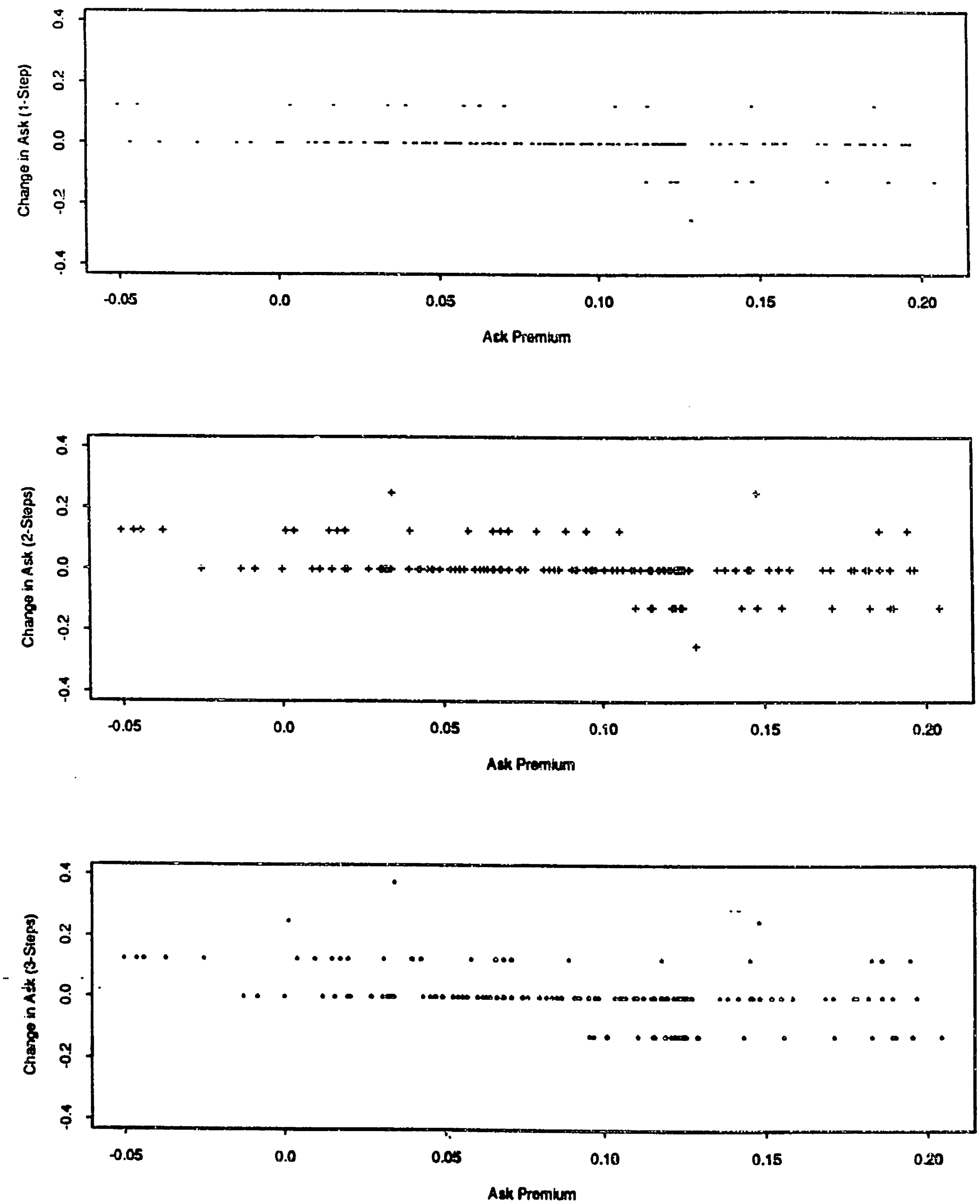


Figure 2.3: Ask-Price Changes versus Ask Premium SBC

Chapter 3

Order Dependence on the Paris Bourse

3.1 Introduction

This chapter explores dependence properties of the flow of market orders to buy or sell stocks on the Paris Bourse.[1] To analyze the issue of randomness of order flow, we focus on the *side* of orders crossing the market—buy side versus sell side. Specifically, we investigate the following questions. First, given the relative frequency of buy orders and sell orders, is it the case that buys (sells) tend to be followed by buys (sells) or is the order flow totally random?[2] Second, if there is evidence of dependence in the side of orders entering the market, is the dependence simple in nature, i.e., consistent with a Markov process, or complex?

Our empirical findings soundly reject the hypothesis of independence of order flow for daily sequences of transaction orders for several stocks traded on the Paris Bourse.

We employ two statistical tests to reach our conclusions. First, a runs test uncovers a strong positive dependence in the daily sequences of buy and sell orders. To analyze whether this dependence is "simple" or "complex," we then investigate the degree of dependence using the framework of Markov chain models. This second test indicates that

[1]This chapter is a revised version of IFSRC Working Paper No. 153-91, Sloan School of Management, MIT, coauthored with Peter J. Kempthorne.

[2]Note that we do *not* hypothesize that buy orders and sell orders are equally likely events. Although not reported here, we have found that this more stringent hypothesis can be rejected without any doubt.

the dependence is often of degree higher than one. Here the degree of dependence measures how much separation between two orders is necessary before they are independent.

While many empirical studies of stock market efficiency look at dependence in transaction prices or returns, there are several reasons for investigating dependence in the sides of successive orders. First, the side of an order is the force consummating a transaction and it determines the price at which the transaction occurs. Therefore, understanding the nature and fundamental properties of the process describing the sides of subsequent transactions is a first step that logically precedes analysis of those other questions.

A second reason for studying order dependence is the importance of random order flow assumptions in the market microstructure literature. Orders are usually among the building blocks of theories of market microstructure and their stochastic properties imply the resulting predictions of the theory. For example, in deriving his formula for the effective bid-ask spread, Roll (1984) explicitly assumes independence in the side of successive market orders. Our study contradicts this assumption for the Paris Bourse.

Finally, results of an order-dependence study can be of direct significance to real-time trading strategies. Suppose buy orders tend to follow buy orders with high probability and similarly for sell orders in a particular market. A trader wanting to sell a certain number of shares should consider the tradeoff of selling immediately at the bid price or issuing a limit sell order at the ask or between the bid and ask. If it is known that the previous trade was a market buy order, it may well be advantageous for this trader to issue a limit order because there is a higher likelihood of the next market order being a "buy."

Stylized models of securities markets often adopt the hypothesis that the pattern of buy and sell orders is random. This hypothesis assumes that any information on previous orders should be of no use for predicting the side of the next order. Many accept the plausibility of this hypothesis and in fact use it to explain the negative autocorrelation in successive price changes as a natural consequence of random order flow in the presence of a bid-ask spread. See Niederhoffer and Osborne (1966) for an early empirical demonstration of this phenomenon and Roll (1984) for an analytic demonstration in a simplified setting.

We strongly reject the hypothesis of independence of order flow for daily sequences of transaction orders for several stocks traded on the Paris Bourse. Our results complement studies of transaction prices of stocks traded on the NYSE which address the hypothesis of independence of transactions. Garbade and Lieber (1977) find that over very short intervals of time (less than 5 or 10 minutes), transactions tend to cluster on a particular side of the market. They conjecture that this is a consequence of market procedures on the NYSE. Using estimates of bid and ask prices, Smidt (1979) presents evidence that changes in market price (defined as the average of the bid price and the ask price) are not independent. More recently, Hasbrouck and Ho (1987) find a significant positive first-order autocorrelation of a series of buy-sell indicators. Unlike the earlier studies, their data include the reported quotes.

Our finding of dependence in orders on the Paris Bourse is particularly significant in light of the contrasting market structure and market procedures of the New York Stock Exchange vis-à-vis the Paris Bourse. First, order fragmentation and aggregation exist on the NYSE complicating the identification of individual orders. These issues are nonexistent on the Paris Bourse as a consequence of the automated order execution system (CAC system, see section 3.2) that allows the precise determination of each individual order.[3]

Second, there is no ambiguity about the side (buy or sell) of any given order on the Paris Bourse. This is not the case on the NYSE where not only transactions can occur between the quotes, but in addition the reported quotes are not always the true quotes in effect at the time of a transaction. A final important difference is the presence of specialists and market makers on the NYSE who seek to maintain orderly markets with non-volatile price changes while minimizing the risk of large long or short positions.[4] Designated market makers (specialists) have only a minimal inventory role in the fully automated Paris Bourse. The net effect of all these market considerations is that our conclusion of order dependence on the Paris Bourse is a strong and fundamental conclusion that cannot be explained by fuzzy data or other ambiguities. The underlying causes for order

[3]Order fragmentation refers to the splitting of an order into several pieces whose execution are reported as separate transactions. Order aggregation happens when a specialist stops a group of successive market orders and reports a single trade for the group at a possibly better price.

[4]See Smidt (1979) for an elaboration of this point.

dependence and its consequences warrant further investigation.

The outline of the balance of this chapter is as follows. Section 3.2 summarizes some relevant features of the Paris Stock Exchange. Section 3.3 gives a brief description of the data used in this study and some examples of the way the order flow process is reconstructed from the individual transaction and order book messages. Positive dependence in orders is established with runs tests in section 3.4. We use the concept of m-dependent processes in section 3.5 to assess whether the order dependence is Markovian or of higher degree. Section 3.6 summarizes our results. Finally, Appendix A gives the details of the runs tests.

3.2 The CAC System

While the New York Stock Exchange and the Paris Stock Exchange are both continuous markets, the Paris Bourse is fully automated. On June 23, 1986, a new quotation and transaction system was introduced, called "CAC" (*Cotation Assisteé en Continu*). In this continuous quotation system all orders are centralized in an "information booklet" of quotes (a central computer). While the organization of the market is centralized, the functioning is not. There is no longer a physical location where the quotes are set or where trading takes place. New orders arriving to the system are handled on a first-come-first-served basis. This ensures a transparent and liquid market where buyers and sellers confront each other directly. Execution of an order takes about 5 seconds, significantly less than for an "open-outcry" system.

A trading day in a particular stock consists of a "pre-opening" from 9am until 10am, the opening at 10am, and a quotation period from 10am until 5pm. During the preopening, orders can be entered into the system without provoking transactions. The preopening allows market participants to "test" the market without making actual trades. At the opening all securities are opened simultaneously and automatically. As is common for this type of "clearing house" transaction, the opening price is defined as the price that maximizes the number of shares traded.[5] Orders that enter the system during the quo-

[5]See Mendelson (1982) for analysis of the clearing house market mechanism.

tation period are executed as soon as an order at the opposite side of the market allows it. In addition to *price priority*, the CAC system maintains *time priority* of orders. Thus within a class of limit orders with the same limit price, an order cannot be executed until all orders that were entered previously have been executed in full (first-come-first-served principle).

The CAC system disseminates information on market conditions for each security at several levels: (a) the *market summary* contains the best bid and the best ask, including the number of securities offered at these prices, (b) the *market limits* consist of the five best buy and the five best sell order prices, including the volume at each price as well as the number of distinct orders comprising that volume, and (c) the *last transaction*, which just gives the last trade's price and volume.

For our purposes, the market limits are of primary importance because this level of information enables us to disentangle the entire stream of orders arriving at the market. Several types of orders can be submitted to the CAC system, including market orders, limit orders, hidden orders, and applications. Market orders submitted prior to the opening are filled not at all, in part or in full—depending on market conditions—at the opening price. After the opening, a market order becomes a limit order with the price of the best ask (in case of a buy order) or of the best bid (in case of a sell order). Applications are special orders which occur when a trader simultaneously submits a buy order and a sell order for the same volume at the same price which is at or between the best bid and and ask prices that prevail at the time of submission. The application orders are not subject to the ordinary priority rules and often serve a role similar to block trades in U.S. markets. Since it is not possible to discern whether an application was initiated by a buyer or seller, we exclude this type of transaction order from our analysis. This will not affect our results noticeably, since the number of applications is quite small compared to the number of market orders.[6] A hidden order is a special kind of limit order whose presence on the order book remains invisible until the "ordinary" limit orders at the hidden order's limit price are exhausted. Examples of the various orders are given at the end of section 3.3.

[6] When measured in volume, this is no longer true since applications are typically associated with much larger volumes than the average market order.

A well-known pitfall with using transaction-level data is the possibility of "order fragmentation." Order fragmentation occurs when a single incoming market order is split up to match multiple individual limit orders that are present on the limit order book. Interpreting such matched orders as distinct trades would induce a meaningless positive dependence in the observed transaction sequence. From the Paris CAC messages, we can infer with certainty what constitutes a single trade and thus eliminate the possibility of order fragmentation. Also, we can unambiguously tell whether a given trade occurred at the buy side or at the sell side of the market. Meaningful tests of order dependence are thus possible.

3.3 Data

The data examined in this study consist of all transaction data on five stocks traded on the Paris Bourse for each of the ten trading days from January 29, 1990 until February 9, 1990. The stocks included in this sample are Elf Aquitaine (oil company), BSN (food products and packaging company), Generale des Eaux (water distribution company), Michelin (tire maker), and Peugeot (car maker). With the exception of Michelin, these stocks are among the highest-capitalization stocks traded on the Paris Stock Exchange. These five companies combined accounted for 21 6 percent of the value-weighted CAC-40 Index at the time of the sample. In Table 3.1, we list the percentage index weights of these five stocks and their average market capitalizations during the sample period. We also include a two-letter abbreviation for each stock for future reference.

Stock	Abbreviation	CAC-40 Index Weight	Market Capitalization
Elf Aquitaine	AQ	6.6%	FF 60 Billion
BSN	BN	4.6%	FF 39 Billion
Generale des Eaux	EX	4.5%	FF 38 Billion
Peugeot	PG	4.2%	FF 36 Billion
Michelin	ML	1.7%	FF 13 Billion

Table 3.1: Sample of Stocks

The complete listing of all CAC-messages broadcasted on transactions and market limits (see section 3.2) were collected for each of these stocks on each of the ten trading days. Our order extraction algorithm, validated with a separately developed algorithm at the Banque Nationale de Paris, processes these data and classifies each transaction as either a buy or as a sell.[7] An attractive feature of the Paris Stock Exchange is that there is no ambiguity about the side of the market of a transaction.[8] This is in sharp contrast to the New York Stock Exchange, where a significant fraction of all trades is executed at a price between the quotes.

I now illustrate the various orders that can occur on the Paris Bourse by presenting selected CAC messages on transactions and market limits for Elf Aquitaine (AQ) from February 5, 1990. Some of these orders result in transactions (market orders and applications), whereas others just change the contents of the order book (limit orders, cancelled orders, and hidden orders). As mentioned earlier, we can infer exactly what kind of order was submitted from the information provided by the CAC messages.

There are two types of CAC messages: transaction messages (labeled 'T') and order book messages (labeled 'O'). The format of these messages is as follows:

- Transaction Message:

  T t I P V,

  where t is the time stamp (seconds after 10am), I is the most recent value of the CAC-40 Index, P is the transaction price (in French Francs), and V is the transaction volume (number of shares).

- Order Book Message

  O t I

[7]The terms "buy" and "sell" may be considered ambiguous, since every transaction involves a buyer and a seller. By a "buy" transaction we mean a transaction *initiated* by a buyer and similarly for a "sell" transaction.

[8]See section 3.2 for a discussion of one special type of transaction called *applications* that constitutes an exception to this rule.

B_1	B_2	B_3	B_4	B_5	A_1	A_2	A_3	A_4	A_5
V_{B_1}	V_{B_2}	V_{B_3}	V_{B_4}	V_{B_5}	V_{A_1}	V_{A_2}	V_{A_3}	V_{A_4}	V_{A_5}
N_{B_1}	N_{B_2}	N_{B_3}	N_{B_4}	N_{B_5}	N_{A_1}	N_{A_2}	N_{A_3}	N_{A_4}	N_{A_5}

In addition to the time stamp t and the CAC-40 Index level I, an order book message consists of the five best bid prices $B_1 > B_2 > \cdots > B_5$ and the five best ask prices $A_1 < A_2 < \cdots < A_5$. The "market" bid is therefore given by B_1 and the "market" ask is given by A_1. Also given for each of the bid (ask) prices B_i (A_i) is the total depth V_{B_i} (V_{A_i}), i.e., the number of shares available at that price, and the total number of distinct orders N_{B_i} (N_{A_i}) comprising that share volume.

Market Order

The first order to be illustrated is the arrival of a market buy order for 1000 shares. The first 700 shares transact at A_1 = FF598, whereas the remaining 300 shares transact at A_2 = FF599. Note that a new price level (FF605) appears when the supply at the (previous) best level reaches zero.

O 436.96 1938.03

596	595	594	592	591	598	599	600	602	604
2800	1800	1100	100	1500	700	700	4300	500	1000
2	1	2	1	2	1	1	15	1	2

T 437.51 1938.03 700 598

T 437.62 1938.03 300 599

O 438.17 1938.03

596	595	594	592	591	599	600	602	604	605
2800	1800	1100	100	1500	400	4300	500	1000	6500
2	1	2	1	2	1	15	1	2	2

Application

In the example below, the first transaction message corresponds to an application (for

1000 shares at FF596) and the second transaction message corresponds to a market sell order (for 100 shares at FF595). Note that applications do not change the contents of the order book.

O 2786.02 1930.55

595	594	593	592	591	596	597	598	599	600
900	800	1200	1100	1000	500	4200	2200	1600	1000
1	2	3	2	1	1	4	7	4	2

T 2829.74 1930.61 1000 596

T 2833.25 1930.61 100 595

O 2833.69 1930.61

595	594	593	592	591	596	597	598	599	600
800	800	1200	1100	1000	500	4200	2200	1600	1000
1	2	3	2	1	1	4	7	4	2

Limit Order

Immediately following the previous order, a limit sell order is submitted for 2000 shares at a price of FF599, as illustrated below. Note that the number of orders at that price increases by one.

O 2833.69 1930.61

595	594	593	592	591	596	597	598	599	600
800	800	1200	1100	1000	500	4200	2200	1600	1000
1	2	3	2	1	1	4	7	4	2

O 2898.34 1929.92

595	594	593	592	591	596	597	598	599	600
800	800	1200	1100	1000	500	4200	2200	3600	1000
1	2	3	2	1	1	4	7	5	2

Hidden Order

Finally an example of a hidden order. Initially, there is a limit buy order for 100 shares at a price of FF596. A transaction (market sell order) is then reported for 1500 shares at a price of FF596. In the absence of hidden orders, the new depth at price FF596 is thus −1400 shares. The subsequent order book shows 1000 shares at this price. Thus a hidden buy order for 2400 shares at a price of FF596 has been detected.

O 3799.78 1930.71

596	595	594	593	592	597	598	599	600	602
100	1000	2500	2200	1200	4200	5700	5100	1000	800
1	1	3	4	3	4	8	7	2	2

T 3799.56 1930.71 1500 596

O 3799.78 1930.71

596	595	594	593	592	597	598	599	600	602
1000	1000	2500	2200	1200	4200	5700	5100	1000	800
1	1	3	4	3	4	8	7	2	2

3.4 Runs Test Analysis

3.4.1 Application of Runs Tests

A powerful test for investigating serial dependence in a dichotomous sequence is the *runs* test. Suppose we have the sequence $BSBBSSS$ consisting of three buy orders (B) and four sell orders (S). This particular sequence has four runs, where a run is defined as a maximal subsequence of like kind. The lengths of these four runs are 1, 1, 2, and 3, respectively. The number of runs in a given sequence is equal to one plus the number of conjunctions of unlike neighbors in that sequence. The idea behind the runs test is that if the number of runs departs significantly from the number to be expected from a totally random sequence, then we can reject an hypothesis of independence.

If there are N trades on a given day divided between b buys and s sells (so that $N = b + s$), the null hypothesis is that the B's and S's are permuted randomly, i.e., that there is no "clustering" of trades at either the buy side or the sell side. The test statistic is the total number of runs, R. Its exact distribution is easily derived using elementary probability theory, see Feller (1968). See Appendix A for a brief derivation and the key results that are used in this study.

3.4.2 Empirical Results

Table 3.2 summarizes the results for the runs test for the first week of our sample. For each stock-day combination, we list the number of buys b, the number of sells s, and the number of runs observed in the sequence of market orders, r. We also give the expected value $\mu(R)$, the standard deviation $\sigma(R)$, and the p-value for the significance of observing r runs under the null hypothesis that the sequence is totally random. The corresponding results for the second week are in Table 3.3. In these two tables (as well as in subsequent tables), we use a two-letter abbreviation for each stock (see Table 3.1) and we index the actual calendar dates from 1 to 10.

From Tables 3.2 and 3.3, we see that in 47 out of 50 cases the observed number of runs r is smaller than the expected number of runs $\mu(R)$ under the assumption of independence. Also, the discrepancy between $\mu(R)$ and the observed value r is often quite large in these 47 cases, when scaled by $\sigma(R)$. Only for the particular cases of the 4th, 8th, and 9th days of trading in BSN the number of observed runs is greater than expected under independence.

The p-value scales the significance of these results: it gives the probability of observing at most r runs in a random sequence consisting of b buys and s sells, i.e., the value of $P\{R \leq r \mid b, s\}$. Whereas we could use the first two moments to calculate a reasonably accurate normal approximation to this probability, the p-values are computed using exact distribution theory. Appendix A discusses the details of this calculation.

Using a cutoff of 0.05, we see from Tables 3.2 and 3.3 that 32 of the 50 cases are statistically significant. Also, 23 of these 32 cases are highly statistically significant (falling below a cutoff of 0.01). This evidence of positive order dependence is fully compelling upon noting that in 9 cases the p-value is even smaller than 0.0005.

The positive dependence among orders on the Paris Bourse complements the finding

Stock	Day	b	s	r	$\mu(R)$	$\sigma(R)$	p
AQ	1	112	114	100	114.0	7.50	0.036
AQ	2	67	116	70	85.9	6.26	0.007
AQ	3	79	63	56	71.1	5.86	0.006
AQ	4	113	121	95	117.9	7.62	0.002
AQ	5	167	135	123	150.3	8.58	0.001
BN	1	116	105	102	111.2	7.40	0.119
BN	2	121	101	109	111.1	7.37	0.414
BN	3	118	93	86	105.0	7.14	0.005
BN	4	128	78	99	97.9	6.74	0.592
BN	5	158	73	95	100.9	6.55	0.209
EX	1	88	59	48	71.6	5.80	0.000
EX	2	74	65	46	70.2	5.85	0.000
EX	3	91	70	60	80.1	6.22	0.001
EX	4	80	49	52	61.8	5.33	0.041
EX	5	144	59	65	84.7	5.85	0.001
PG	1	128	86	90	103.9	7.01	0.028
PG	2	137	101	107	117.3	7.52	0.097
PG	3	94	114	92	104.0	7.13	0.053
PG	4	81	83	79	83.0	6.38	0.292
PG	5	101	76	72	87.7	6.50	0.010
ML	1	100	80	61	89.9	6.61	0.000
ML	2	75	90	78	82.8	6.35	0.248
ML	3	132	117	106	125.0	7.85	0.009
ML	4	104	78	77	90.1	6.59	0.028
ML	5	151	109	108	127.6	7.84	0.007

Table 3.2: Runs Test Results for First Week

For each stock-day combination, the number of buys b, the number of sells s, and the number of runs r is recorded. The values of $\mu(R)$ and $\sigma(R)$ are the mean and the standard deviation of the number of runs under the null hypothesis that the b buys and s sells are permuted randomly. The value of p gives the probability of observing at most r runs under the independence assumption, i.e., $p = P\{R \leq r \mid b, s\}$. If $p < 0.0005$, the value 0.000 is inserted in the table.

Stock	Day	b	s	r	$\mu(R)$	$\sigma(R)$	p
AQ	6	107	150	97	125.9	7.78	0.000
AQ	7	122	155	104	137.5	8.19	0.000
AQ	8	88	149	93	111.6	7.17	0.006
AQ	9	78	82	71	81.0	6.30	0.067
AQ	10	85	80	60	83.4	6.40	0.000
BN	6	126	89	92	105.3	7.10	0.036
BN	7	97	85	89	91.6	6.70	0.377
BN	8	91	129	113	107.7	7.18	0.790
BN	9	75	70	76	73.4	5.99	0.697
BN	10	80	79	79	80.5	6.28	0.437
EX	6	112	71	84	87.9	6.40	0.296
EX	7	53	76	51	63.4	5.48	0.015
EX	8	60	77	50	68.4	5.74	0.001
EX	9	78	66	66	72.5	5.94	0.156
EX	10	47	67	53	56.2	5.15	0.297
PG	6	110	102	104	106.8	7.25	0.373
PG	7	141	160	131	150.9	8.63	0.012
PG	8	267	221	199	242.8	10.94	0.000
PG	9	237	131	125	169.7	8.78	0.000
PG	10	218	135	137	167.7	8.86	0.000
ML	6	280	212	221	242.3	10.87	0.028
ML	7	337	198	220	250.4	10.77	0.003
ML	8	282	192	196	229.5	10.48	0.001
ML	9	515	223	304	312.2	11.45	0.247
ML	10	415	119	170	186.0	7.99	0.027

Table 3.3: Runs Test Results for Second Week

For each stock-day combination, the number of buys b, the number of sells s, and the number of runs r is recorded. The values of $\mu(R)$ and $\sigma(R)$ are the mean and the standard deviation of the number of runs under the null hypothesis that the b buys and s sells are permuted randomly. The value of p gives the probability of observing at most r runs under the independence assumption, i.e., $p = P\{R \leq r \mid b, s\}$. If $p < 0.0005$, the value 0.000 is inserted in the table.

of positive dependence in orders on the New York Stock Exchange by Hasbrouck and Ho (1987). However, the runs-test method outlined above is more sensitive and powerful than their test based on aggregating across 684 NYSE stocks and across 42 days. They calculate the average autocorrelation of the series of buy-sell indicators and find that there is a market-wide positive autocorrelation in buy and sell orders. It should be pointed out that their results are not on a stock-by-stock basis nor do they reveal whether the dependence is significant over short periods of time. That is, they do not demonstrate a statistical significance of order dependence on individual stocks of periods of one day or less. Also, the significance of an autocorrelation is based on a null hypothesis of no autocorrelation and a null distribution based upon normal approximations. Any test based upon the autocorrelation likely uses the same null distribution regardless of the proportion of buy orders and sell orders in the sample. Our test is more sensitive because we use exact distribution theory for the test statistic. This test statistic is appropriate for transaction data of individual stocks, where the transaction data can be of arbitrary length and can have arbitrary proportions of buy and sell orders.

We should remark that a two-sided test for dependence—either positive or negative—could have been performed rather than this test for *positive* dependence. That test would lead to merely doubling all p-values and would not affect the conclusions. Moreover, tests for positive dependence can be justified. For example, we noted above that positive dependence has been verified to exist on the New York Stock Exchange, see Hasbrouck and Ho (1987). Unlike the Markov chain tests presented in section 3.5, the positive dependence hypothesis is a one-sided alternative for the distribution of the number of runs.

We conclude from the runs test analysis that for many series the assumption of a random order flow can be rejected. Moreover, since the number of runs appears to be lower than expected for the case of a random order flow, the results indicate that there is a *positive* dependence. Buys are more likely following buys, and similarly for sells.

3.5 m-Dependence Analysis

3.5.1 m-Dependent Processes

When the runs test suggests that a given sequence displays dependence, what is the nature of this dependence? A general class of dependent processes consists of *m-dependent processes*, where $m \geq 0$ is an integer. A discrete-time process is m-dependent if an observation's distribution is fully specified by the immediately preceding m observations. Thus, any two observations are independent as long as they are more than m units apart. In this terminology an independent sequence is 0-dependent.

To formalize the notion of dependence, consider modeling the series of buy and sell orders in a particular stock on a given day as a two-state, discrete-time stochastic process. The state space I of the process is defined by $I = \{B, S\}$, where the states B and S denote a buy and a sell order, respectively. The process is observed at the times that transactions occur. Let the opening transaction occur at time $t = 0$. If there are N transactions in the stock on this day, then we observe the process at times t for $t = 1, \ldots, N$.[9] The state of the system at time t is denoted by the random variable $X(t)$.

For a given day of trading in a particular stock, we observe the value of N and the realizations of the state $X(t)$ for $t = 1, \ldots, N$. It is important to realize that for the Paris Bourse, there is no ambiguity about the side of the market for any given transaction (see section 3.2).

We define the probabilities $p_i(t)$, $p_{ij}(t)$, and $p_{ijk}(t)$ as follows:

$$\begin{aligned} p_i(t) &= P\{X(t) = i\}, \\ p_{ij}(t) &= P\{X(t) = j \mid X(t-1) = i\}, \quad \text{and} \\ p_{ijk}(t) &= P\{X(t) = k \mid X(t-1) = j, X(t-2) = i\}. \end{aligned}$$

Thus $p_i(t)$ is just the (unconditional) probability of being in state i at time t, while $p_{ij}(t)$ and $p_{ijk}(t)$ are conditional probabilities, conditioning on the previous state and on the previous two states, respectively.

[9]Note that the time variable t is a transaction index and that it does not correspond to clock time.

For simplicity, we will assume that the process $X(t)$ is *stationary*, i.e., that the transition probabilities are independent of t. Thus for a stationary process, we can write $p_i(t) = p_i$, $p_{ij}(t) = p_{ij}$, and $p_{ijk}(t) = p_{ijk}$ for all t. Furthermore, we assume that the stochastic process $X(t)$ is an m-dependent process with either: (a) $m = 0$: the chain is an independent process and is thus fully characterized by the probabilities p_i; (b) $m = 1$: the chain is Markovian and is thus fully characterized by the probabilities p_{ij}; or (c) $m = 2$: the chain has a second-order dependence and is thus fully characterized by the probabilities p_{ijk}.

3.5.2 Empirical Results

Let H_m denote the hypothesis that $\{X(t)\}$ is m-dependent, where $m = 0, 1$, or 2. The tests of m-dependence require the maximum-likelihood estimates of the transition probabilities of $\{X(t)\}$ under the hypotheses H_0, H_1, and H_2. Assuming stationarity of the process, the maximum-likelihood estimates under the various hypotheses are given by:

$$H_0 : \hat{p}_i = n_i \Big/ \sum_l n_l, \quad H_1 : \hat{p}_{ij} = n_{ij} \Big/ \sum_l n_{il}, \quad \text{and} \quad H_2 : \hat{p}_{ijk} = n_{ijk} \Big/ \sum_l n_{ijl},$$

where n_i is the number of times that state i has been visited, n_{ij} is the number of times that state i was immediately followed by state j and similarly for n_{ijk}. Note that by assumption, $\sum_l n_l = N$, $\sum_{il} n_{il} = N - 1$, and $\sum_{ijl} n_{ijl} = N - 2$.

Suppose we want to test H_0 against H_1. Anderson and Goodman (1957) show that under the null hypothesis, the quantity $-2\log(\lambda_{01})$ has an asymptotic χ^2 distribution with one degree of freedom, where the likelihood ratio λ_{01} is defined by

$$\lambda_{01} = \prod_{i,j} \left(\frac{\hat{p}_j}{\hat{p}_{ij}} \right)^{n_{ij}} .$$

Generally, for testing u-dependence versus v-dependence ($u < v$), the quantity $-2\log(\lambda_{uv})$ has an asymptotic χ^2 distribution with $2^v - 2^u$ degrees of freedom, where the likelihood ratio λ_{uv} is defined in a similar fashion as detailed above. For example, when testing H_1 against H_2, the likelihood ratio λ_{12} is given by

$$\lambda_{12} = \prod_{i,j,k} \left(\frac{\hat{p}_{jk}}{\hat{p}_{ijk}} \right)^{n_{ijk}} .$$

For each stock and each of the ten trading days, we first test H_0 versus H_1, independence versus Markovian dependence. Table 3.4 summarizes the testing results. The column labeled "H_0 vs. H_1" gives the approximate p-values for this test, based upon the asymptotic chi-squared distribution of the likelihood ratio test statistic. Using a cutoff level of 5%, we would reject the model of an independent buy-sell process in favor of the Markov process in exactly half of all cases (25 out of 50).

To illustrate the practical significance of the test of H_0 (independence) versus H_1 (Markov dependence), Table 3.4 also gives the estimated probability ratios

$$\hat{\rho}_B = \frac{\hat{p}_{BB}}{\hat{p}_B} \qquad \text{and} \qquad \hat{\rho}_S = \frac{\hat{p}_{SS}}{\hat{p}_S}$$

for all 50 stock-days. As p_B was defined as the probability of a buy order ($p_B = P\{X(t) = B\}$) and p_{BB} as the probability of a buy order given that the previous order was a buy order ($p_{BB} = P\{X(t) = B \mid X(t-1) = B\}$), the probability ratio ρ_B is a direct, interpretable measure of dependence. For a positively-dependent process, we will have $p_{BB} > p_B$ and $p_{SS} > p_S$ so that $\rho_B > 1$ and $\rho_S > 1$. For a negatively-dependent process, the inequalities will be reversed. In the watershed case of an independent process, we will have equality of conditional and unconditional probabilities and the ratios ρ_B and ρ_S are equal to unity.

We see from Table 3.4 that for Elf Aquitaine (AQ), the conditional probability of a buy following a buy is on average 22% higher than the unconditional probability of a buy. The corresponding number for a sell is 18%. Positive dependence is illustrated by observing that the empirical probability ratios exceed 1.0 for 47 out of 50 (94%) stock-days.[10]

In Table 3.5 we finally report the complete results of testing H_1 versus H_2, Markovian dependence versus 2-dependence.[11] Again using a cutoff of 5%, we would reject the assumption of a Markov process in favor of a 2-dependent process in 9 out of 50 cases (or 9 out of 25 cases where the independence assumption was rejected). A generalization of the probability ratios discussed above can also be used to make this test more directly interpretable. For example, we report in Table 3.5 the values of $\hat{\rho}_{BB} = \hat{p}_{BBB}/\hat{p}_{BB}$ and $\hat{\rho}_{SS} = \hat{p}_{SSS}/\hat{p}_{SS}$.

[10] As was the case with the runs test analysis, the three exceptions are all with BSN.

[11] Note that this test is most meaningful for those cases where the independent process (H_0) was rejected.

Stock	Day	H_0 vs. H_1	$\hat{\rho}_B$	$\hat{\rho}_S$	Stock	Day	H_0 vs. H_1	$\hat{\rho}_B$	$\hat{\rho}_S$
AQ	1	0.071	1.12	1.12	AQ	6	0.000	1.30	1.16
AQ	2	0.014	1.30	1.11	AQ	7	0.000	1.30	1.20
AQ	3	0.013	1.18	1.25	AQ	8	0.011	1.29	1.10
AQ	4	0.003	1.20	1.18	AQ	9	0.134	1.12	1.12
AQ	5	0.002	1.14	1.23	AQ	10	0.000	1.26	1.31
BN	1	0.237	1.07	1.09	BN	6	0.071	1.08	1.18
BN	2	0.813	1.01	1.02	BN	7	0.741	1.02	1.03
BN	3	0.009	1.14	1.23	BN	8	0.432	0.93	0.96
BN	4	0.773	0.99	0.96	BN	9	0.602	0.95	0.96
BN	5	0.391	1.02	1.13	BN	10	0.860	1.02	1.01
EX	1	0.000	1.21	1.50	EX	6	0.595	1.02	1.07
EX	2	0.000	1.31	1.38	EX	7	0.031	1.26	1.14
EX	3	0.002	1.20	1.31	EX	8	0.002	1.35	1.20
EX	4	0.083	1.09	1.26	EX	9	0.311	1.07	1.11
EX	5	0.001	1.09	1.57	EX	10	0.571	1.08	1.03
PG	1	0.057	1.09	1.19	PG	6	0.741	1.02	1.03
PG	2	0.188	1.06	1.12	PG	7	0.024	1.15	1.11
PG	3	0.105	1.13	1.10	PG	8	0.000	1.15	1.22
PG	4	0.580	1.05	1.04	PG	9	0.000	1.14	1.48
PG	5	0.019	1.14	1.23	PG	10	0.001	1.11	1.29
ML	1	0.000	1.25	1.41	ML	6	0.057	1.07	1.11
ML	2	0.493	1.06	1.05	ML	7	0.006	1.07	1.20
ML	3	0.018	1.13	1.17	ML	8	0.002	1.10	1.21
ML	4	0.053	1.10	1.20	ML	9	0.504	1.01	1.06
ML	5	0.015	1.11	1.20	ML	10	0.060	1.03	1.28

Table 3.4: Probability Ratios and Significance Levels for Test of H_0 versus H_1

H_m is the hypothesis that $\{X(t)\}$, the buy-sell process for a given stock, is m-dependent. This table reports the results of testing independence (H_0) versus Markov dependence (H_1) for all stock-day combinations. The column labeled "H_0 vs. H_1" gives the approximate p-values for this test, based upon the asymptotic chi-squared distribution of the likelihood ratio test statistic. Also reported are the estimated probability ratios $\hat{\rho}_B = \hat{p}_{BB}/\hat{p}_B$ and $\hat{\rho}_S = \hat{p}_{SS}/\hat{p}_S$. Values of $\hat{\rho}_B$ and $\hat{\rho}_S$ exceeding 1 illustrate positive dependence.

Stock	Day	H_1 vs. H_2	$\hat{\rho}_{BB}$	$\hat{\rho}_{SS}$	Stock	Day	H_1 vs. H_2	$\hat{\rho}_{BB}$	$\hat{\rho}_{SS}$
AQ	1	0.647	0.93	0.97	AQ	6	0.004	1.24	1.05
AQ	2	0.593	1.11	1.01	AQ	7	0.011	1.20	1.00
AQ	3	0.214	1.10	1.08	AQ	8	0.704	1.10	1.00
AQ	4	0.008	1.18	1.10	AQ	9	0.456	1.09	1.08
AQ	5	0.253	1.05	1.08	AQ	10	0.768	1.01	1.04
BN	1	0.069	1.15	1.11	BN	6	0.124	1.10	1.09
BN	2	0.422	0.96	1.14	BN	7	0.276	1.10	1.11
BN	3	0.599	0.99	1.09	BN	8	0.167	1.04	1.14
BN	4	0.147	1.09	0.79	BN	9	0.156	0.88	1.21
BN	5	0.531	1.02	1.19	BN	10	0.658	0.90	1.03
EX	1	0.136	1.07	1.04	EX	6	0.183	1.07	1.21
EX	2	0.038	1.04	1.18	EX	7	0.576	1.14	1.02
EX	3	0.000	1.18	1.18	EX	8	0.025	1.24	1.05
EX	4	0.296	0.93	1.00	EX	9	0.081	1.19	0.96
EX	5	0.482	1.02	1.13	EX	10	0.741	0.96	1.07
PG	1	0.050	1.06	1.23	PG	6	0.114	1.11	1.17
PG	2	0.116	1.12	1.05	PG	7	0.784	1.05	1.00
PG	3	0.868	0.94	1.00	PG	8	0.054	1.08	1.05
PG	4	0.414	1.06	1.14	PG	9	0.001	1.08	1.13
PG	5	0.771	1.03	0.95	PG	10	0.184	1.01	1.16
ML	1	0.348	1.05	0.96	ML	6	0.226	1.06	1.07
ML	2	0.928	0.98	0.98	ML	7	0.010	1.05	1.20
ML	3	0.842	1.02	1.04	ML	8	0.015	1.07	1.15
ML	4	0.787	1.02	1.07	ML	9	0.093	1.02	1.25
ML	5	0.159	1.04	1.15	ML	10	0.113	1.02	1.24

Table 3.5: Probability Ratios and Significance Levels for Test of H_1 versus H_2

H_m is the hypothesis that $\{X(t)\}$, the buy-sell process for a given stock, is m-dependent. This table reports the results of testing Markov dependence (H_1) versus 2-dependence (H_2) for all stock-day combinations. The column labeled "H_1 vs. H_2" gives the approximate p-values for this test, based upon the asymptotic chi-squared distribution of the likelihood ratio test statistic. Also reported are the estimated probability ratios $\hat{\rho}_{BB} = \hat{p}_{BBB}/\hat{p}_{BB}$ and $\hat{\rho}_{SS} = \hat{p}_{SSS}/\hat{p}_{SS}$. Values of $\hat{\rho}_{BB}$ and $\hat{\rho}_{SS}$ exceeding 1 illustrate positive dependence.

Thus we have established both the presence of (positive) dependence and gained some insight in the basic nature of this dependence. Tests for higher-order dependence are infeasible with the shorter sequences. Pooling together of different days would lead to problems of non-stationarity. We therefore do not pursue here the issue of estimating the precise degree of dependence m for individual sequences.

3.6 Summary

We have studied the hypothesis that separate transactions in individual stocks traded on the Paris Bourse are independent events with respect to whether they were initiated by a buyer or a seller. By considering a sample representing five of the major stocks traded on the Paris Stock Exchange over a 10-day trading period, we find strong evidence of positive dependence among orders. That is, given that the last order for a particular stock was a buy order, the probability that the next order will also be a buy order is higher than the unconditional probability of a buy order. A similar conclusion holds for sell orders. Likelihood-ratio tests of m-dependent processes confirm the conclusion of order dependence. For many series the assumption of an independent buy-sell process can be rejected in favor of Markov dependence. Moreover, for trading in some stocks on certain days, Markov dependence can also be rejected in favor of higher-order dependence.

In earlier studies, order dependence has been shown to exist on the New York Stock Exchange. However, it has been suggested that this is a consequence of the market procedures on the NYSE which can result in fragmentation of orders. Also, there is an inherent ambiguity in classifying market orders as buy or sell orders on the NYSE. Since our results on the Paris Bourse are not compromised by any ambiguity (order fragmentation or order classification), our conclusion of order dependence is a significant new finding.

Having established order dependence on the Paris Bourse, many questions arise. What is the structure of the resulting dependencies in price movements and returns? Is it stationary or transitory? Do outstanding limit orders have a tendency to be "sticky," resulting in a current order book which is stale? If so, limit orders could be picked off by sophisticated traders when good news hits the market before the old limit orders are cancelled, giving an explanation for order dependence. Another interesting question is whether current order flow can be used to predict future order flow and future returns. We defer investigation of these challenging questions to further research.

Chapter 4

The Dynamics of the Bid-Ask Spread

4.1 Introduction

The bid-ask spread of an exchange-traded security is widely recognized as an important measure of market liquidity: stocks with high bid-ask spreads are relatively less liquid and stocks with low bid-ask spreads are relatively more liquid. Since investors typically buy a security at the ask price and sell at the bid price, the size of the spread is directly relevant for them since it influences the return obtainable from the security. As part of the quickly expanding field of market microstructure, the bid-ask spread has recently received a lot of attention in the academic community. The interest in the spread is not confined to academia: the New York Stock Exchange explicitly monitors the behavior of bid-ask spreads quoted by its specialists.[1]

The purpose of this chapter is to examine how the bid-ask spread changes on a quote-by-quote basis over the course of the trading day and how this relates to existing theories of the bid-ask spread. By focusing on the dynamic process between successive quote revisions, I seek to capture the impact on bid-ask spreads of the information revealed through the trading process, such as trading volumes, stock-price volatility, and the timing of trades.

[1] For example, the NYSE Fact Book 1991 states in the section titled "Market performance" the following: "The quotation spread between bid and asked was 1/4 point or less in 84.5% of NYSE quotes, up from 81.5% in the previous year."

The main contribution of my study is that it increases our understanding of the determinants of liquidity at the micro level. I analyze empirically the significance of several theoretical determinants of spreads, most of which have not been analyzed before in a quote-by-quote study of the spread for individual stocks. Also, the methodology used in this chapter, the ordered probit model, explicitly takes the discreteness property of bid-ask spreads into account. The bid-ask spread of stocks traded on U.S. exchanges is measured in ticks, where a tick is an eighth of a dollar. Spreads usually assume only very few different values.[2]

The importance of studying the dynamics of the bid-ask spread transcends the identification of empirical properties of the spread in terms of its relation to other economic variables or the relative importance of these variables. Several arguments are outlined below.

First, by studying the dynamics of the bid-ask spread we can get a better insight in the liquidity characteristics of a dealer market and assess the consequences of the specialist system. Knowing what factors determine the cost of liquidity may help to minimize those costs. This has in turn implications for the ongoing debate about market structure and efficient market making: issues of regulation versus competition—the monopolistic specialist system versus a competitive dealer system or automated trading. The basic question in that debate is identifying the best market structure for obtaining the desired liquidity characteristics.

Second, information about the dynamics of the bid-ask spread is important for index arbitrage and program-trading strategies. The success of any strategy that makes use of tiny price differences between different baskets of securities will be influenced by the bid-ask spread. If the strategy calls for trading securities on a scale such that the spread will be affected, a gain that might have materialized for a minimum-size transaction will soon be wiped out.

Third, if we believe that information asymmetries do exist in the market, knowledge about the dynamic properties of the spread is important for liquidity traders since they

[2] Indeed, the proportion of all spreads equal to 1, 2, or 3 ticks for the stocks in my sample averages 98.3%.

will on average lose money in situations with asymmetric information. The behavior of the spread can indicate whether adverse selection is likely to be important.

Fourth, the dynamic properties of the spread can be linked to the cost of equity for firms. More liquid securities, as measured by the size and dynamics of the spread, are more valuable. Hence the dynamics of the spread are important for the issuing company since its cost of capital is affected. Can the company increase the liquidity of its shares and decrease its cost of capital? One application is the analysis of the liquidity effects of stock splits, a topic addressed in Chapter 5 of this thesis.

Finally, I mention that knowing the properties of the bid-ask spread can help to understand the behavior of transaction prices, the primary focus of research in numerous studies. Consider, for example, the model by Roll (1984) and the sizable literature that extends his basic model—see Harris (1990b) for additional references—to estimate the "effective" bid-ask spread from observations of transaction prices only. This entire literature assumes a constant bid-ask spread. Knowledge of the actual dynamic properties of the bid-ask spread can be used to assess whether assuming a constant bid-ask spread is a good approximation. More-elaborate models may be appropriate when this is not the case.

Using 1988 data for a sample consisting of the 30 stocks included in the Dow Jones Industrial Average, my empirical results are consistent with the theoretical prediction that high trading volumes are associated with a subsequent widening of the bid-ask spread. The effect of trades occurring at the midpoint of the bid-ask spread is negligible in comparison to trades occurring at (or close) to the quotes. This is interpreted as evidence in support of the notion that trades initiated by market participants willing to pay (part of) the bid-ask spread have a different information content from trades that are most likely liquidity trades crossing at the midquote, a finding consistent with the ideas of the adverse-selection theory. The effect of lagged volume on spreads appears to be negligible. In other words, the information contained in the trading process is processed very quickly. Also, it appears that purchases affect the spread in most cases more than sales, but the economic significance of this effect is fairly small.

Another economic variable whose theoretical importance is supported in my study is stock-price volatility. Since specialists cannot always have a zero net position in the stocks for which they make a market, they bear the risk associated with unfavorable movements of the stock price. It is therefore plausible that the spread, their reward for bearing this risk, is somehow related to price volatility. My empirical results suggest that this is indeed the case. I find strong evidence that spreads tend to increase when prices are volatile.

Other issues examined in this study are time-of-day effects, the effects of competition, and the relation between the spread and the time between trades. As far as time-of-day effects is concerned, I find no additional increase of the spread at the beginning of the trading day after taking trading volume, volatility, etc. into account. On the other hand, the end-of-day increase of the spread is still measurable after taking account of the other economic variables. The competitive effect on NYSE spreads of trading away from the NYSE is statistically significant, but appears to be economically marginal. Finally, I find some evidence in support of the notion that spreads decrease the longer the time between trades.

One caveat about this study is the following. Many papers have stressed the importance of the market-maker's inventory in setting the quotes. I do not have the data necessary to include the inventory variable in my analysis and the analysis here has therefore nothing to say about the empirical importance of specialist inventory for the behavior of the bid-ask spread.

The organization of this chapter is as follows. Section 4.2 describes various theories of the bid-ask spread and their implications. Section 4.3 presents the model used in this study and the specification used to estimate this model. Section 4.4 contains the results. The issue of forecasting is taken up in section 4.5 and section 4.6 concludes this chapter. A detailed discussion of the data used in this study is presented in Appendix B.

4.2 Theories of the Bid-Ask Spread

The classical view of the spread in a dealer market is that it reflects the price of immediate execution: rather than waiting for a party wishing to take the opposite side of a transaction at a more favorable price, a trader can transact immediately with the dealer. By standing ready to execute these trades, the dealer is exposed to the risk of unfavorable price movements and the bid-ask spread is considered the reward for bearing this risk. Demsetz (1968) was the first to analyze the bid-ask spread along these lines in a formal demand and supply framework. Following Demsetz, Garman (1976) and other researchers have elaborated on the various risk factors (costs) facing the dealer and various models have been developed in which the dealer's quote-setting behavior—and thus the bid-ask spread—results from optimizing behavior.

Stoll (1978) derives the dealer's cost function in a one-period model, focusing on holding costs, the price risk and opportunity cost of holding securities. In addition to the dealer's attitude towards risk, the determinants of the bid-ask spread in his model are the dollar size of the transaction and the variance of return of the stock being traded. Ho and Stoll (1981) derive a multiperiod extension of Stoll (1978) with essentially the same results. Assuming that the dealer's risk aversion does not change over short periods of time, these early papers suggest the importance of dollar volumes and stock-price volatility.

There are essentially two types of bid-ask spread theories that have developed in the market microstructure literature: inventory-based theories and adverse-selection theories. The former theory emphasizes the dependence of bid and ask prices on the dealer inventory. The latter theory is based on the consequences of information asymmetries between different types of traders. In adverse-selection models, the bid-ask spread arises as compensation for the dealer's expected losses to insiders.

The implications of dealer inventory for bid-ask spreads in inventory-based theories are by no means uniform. For example, the spread is independent of inventory in the models by Stoll (1978) and Ho and Stoll (1981). On the other hand, the inventory position does matter in Amihud and Mendelson (1980), who explicitly derive the behavior of the bid-ask spread for the special case of linear demand and supply functions. Similarly, O'Hara and Oldfield (1986) emphasize the important role of inventory in affecting both the placement and the size of the spread. Unfortunately, I cannot include dealer inventory in my empirical analysis since I do not have the necessary data on specialist inventory and specialist participation.[3]

The basic idea of the adverse-selection theories was contained in Bagehot (1971). Formal developments are Copeland and Galai (1983) and the seminal paper by Glosten and Milgrom (1985).[4] An important innovation in this branch of the literature is made by Easley and O'Hara (1987). In their model, transaction volume is correlated with information and the spread widens for large transactions. Note that this result applies to the simultaneous occurrence of trades and quotes: *different* quotes are applicable for different trading volumes. In actual markets, we only observe a *single* quote, which may or may not be revised after a transaction. This quote technically only applies to volumes not exceeding the quoted depth. Thus in my empirical work, the closest approximation to studying the effect predicted by Easley and O'Hara (1987) is to consider whether the spread widens immediately following large transactions.

Note that both Stoll (1978) and Easley and O'Hara (1987) suggest a positive relation between trading volumes and bid-ask spreads, albeit for different reasons: pure adverse selection in Easley and O'Hara (1987) versus compensation for departures from the dealer's optimal portfolio in Stoll (1978). As discussed in the section to follow, I try to distinguish between these two hypotheses by partitioning trading volumes into three separate categories: trades occurring at the midquote, trades occurring at (or close to) the ask, and trades occurring at (or close to) the bid.

This partitioning of trading volumes also enables us to consider the interesting hypothesis by Allen and Gorton (1991). They argue that liquidity sales are more likely than liquidity purchases and as a consequence, the bid price moves less in response to a sale than the ask price in response to a purchase. They also argue that it is easier to exploit good news than bad news, so that there is again more information in a purchase than

[3]But see Madhavan and Smidt (1991) for a rare study of specialist pricing that does incorporate data on these variables.

[4]See also Kyle (1985) for closely related work that does not explicitly focus on the spread.

a sale. I will investigate empirically whether the spread reacts differently to purchases compared to sales.

A whole literature has developed that attempts to decompose the bid-ask spread into its different components. See, for example, George et al. (1991), Glosten (1987), Glosten and Harris (1988), and Stoll (1989). Some of those papers explicitly assume that the bid-ask spread for a given stock is constant. It would be an interesting side result of my study if I can establish that the dynamic behavior of the spread is too important to neglect in analyses of this type.

A recent theory by Easley and O'Hara (1992) elaborates on the importance of the timing of trades. In a model with event uncertainty, they show that the time between trades affects spreads. Specifically, spreads are shown to decrease as the time between trades increases. The reason is that in their model the absence of trades in some time interval is more likely to occur if the information event does not occur. Thus the adverse-selection problem is less severe and the spread is adjusted accordingly.

Volatility has been shown to be an important determinant of the bid-ask spread in several theoretical models. I already mentioned the contributions by Stoll (1978) and Ho and Stoll (1981). Easley and O'Hara (1987) prove in their adverse-selection model that under certain conditions the spread increases with increased variance of the value of the asset [Proposition 6, part (5)]. Copeland and Galai (1983) derive the same result in their model, using the theory of option pricing. It seems plausible that even in the absence of adverse selection, a general risk-return argument suggests that volatility matters for the behavior of the spread.

Finally, I briefly mention some other empirical contributions to the bid-ask spread literature. The initial contributions focused on the cross-sectional determinants of bid-ask spreads, i.e., why different securities have different spreads. For example, see Demsetz (1968), Tiniç (1972), and Tiniç and West (1972). Determinants considered are order flow variables, stock price levels, stock price variability, and competition variables such as the number of market makers and the number of markets where a security is listed. These studies are based on small data sets with a single or at most very few observations of the spread per stock (typically at the close). A more recent study is Venkatesh and Chiang (1986), who study the behavior of spreads prior to earnings and dividend announcements. With the recent availability of detailed transactions data, more comprehensive studies have been undertaken that also focus on spread behavior. Recent examples are Harris (1992), who does a sophisticated cross-sectional analysis of the relative spread and McInish and Wood (1992), who estimate a mixed time-series/cross-sectional regression model for the relative spread. Rather than using their data observation-by-observation, they use 30-minute aggregate values for all variables. See also Hasbrouck (1991), who does a linear autoregression of spreads on (lagged) trade volumes.

4.3 Model and Specification

This section contains a brief summary of the model used in this study, the ordered probit model, and a discussion of its specification to investigate the dynamics of the bid-ask spread.

4.3.1 The Ordered Probit Model

The ordered probit model is a statistical model applicable to situations where the dependent variable has a discrete number of possible outcomes that have a natural ordering. The model was recently applied in the context of market microstructure research by Hausman, Lo, and MacKinlay (1992), hereafter HLM (1992), who estimate the conditional distribution of trade-to-trade price changes for a large sample of stocks. Since HLM (1992) describe the ordered probit model in detail, I only include a brief description of the application of the model in the current context and I refer to HLM (1992) for further discussion and references.

Suppose we have a collection of n bid-ask quotes at times $t_1, \ldots, t_n$. The quote at time t_k consists of a bid price $B(t_k)$ and an ask price $A(t_k)$, where $A(t_k) > B(t_k)$. For ease of notation, let $B_k \equiv B(t_k)$ and $A_k \equiv A(t_k)$. Then the bid-ask spread S_k is given by $S_k \equiv A_k - B_k$. As noted in the introduction, S_k is an integer multiple of a tick and

$S_k > 0$.

Let S_k^* be a continuous random variable such that

$$S_k^* = X_k'\beta + \epsilon_k, \tag{4.1}$$

where $E[\epsilon_k \mid X_k] = 0$ and ϵ_k is a collection of independent normal random variables, say $N(0, \sigma_k^2)$. X_k is a vector of predetermined explanatory variables.

The connection between the observed discrete spread variable S_k and the unobserved continuous variable S_k^* is established by assuming that the following holds:

$$S_k = \begin{cases} s_1 & \text{if } S_k^* \in (\alpha_0, \alpha_1] \\ s_2 & \text{if } S_k^* \in (\alpha_1, \alpha_2] \\ \vdots & \qquad \vdots \\ s_m & \text{if } S_k^* \in (\alpha_{m-1}, \alpha_m], \end{cases} \tag{4.2}$$

where (i) $\alpha_0 \equiv -\infty$, $\alpha_m \equiv +\infty$, and $\alpha_1 < \cdots < \alpha_{m-1}$ are parameters to be estimated; and (ii) the values s_j are the discrete possible outcomes for the spread S_k, i.e., 1 tick, 2 ticks, etc.

Although the spread can in principle assume a countably infinite number of values, the state space dimension m has to be finite in order to make estimation of the model possible. Also, in practice the spread only assumes a small number of different values—usually from 1 to 4 ticks ($\$\frac{1}{8}$ to $\$\frac{1}{2}$) for a typical stock.

For a given value of m, I specify the values of s_j as follows: $s_j = j$ for $j = 1, \ldots, m-1$ and $s_m = \{k \mid k \geq m\}$, all in units of a tick. In other words, the jth state corresponds to a bid-ask spread of j ticks except for the mth state, which corresponds to a bid-ask spread of m ticks or more.[5]

The conditional probability that $S_k = s_j$ given X_k can be easily evaluated. From equations (4.1) and (4.2) and the distributional assumption concerning ϵ_k, we have for $j = 1, \ldots, m$,

$$P\{S_k = s_j \mid X_k\} \;=\; P\{S_k^* \in (\alpha_{j-1}, \alpha_j] \mid X_k\}$$

[5]Different choices for the state space of the spread variables are obviously possible. For example, an alternative for a high-priced stock with "high" absolute spread levels is to let $s_j = \{2j-1, 2j\}$ for $j = 1, \ldots, m-1$ and $s_m = \{k \mid k \geq 2m-1\}$, again both denominated in ticks.

$$= \mathrm{P}\{X_k'\beta + \epsilon_k \in (\alpha_{j-1}, \alpha_j] \mid X_k\}$$
$$= \Phi\left(\frac{\alpha_j - X_k'\beta}{\sigma_k}\right) - \Phi\left(\frac{\alpha_{j-1} - X_k'\beta}{\sigma_k}\right), \qquad (4.3)$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. This concludes the discussion of the ordered probit model.

4.3.2 The Specification

There are several issues to be addressed concerning the specification: the explanatory variables X_k in equation (4.1), the variance term σ_k^2 associated with ϵ_k in equation (4.1), and the dimension m of the state space in equation (4.2).

As far as the variance σ_k^2 is concerned, I make the simplifying assumption of homoskedasticity, i.e., $\sigma_k = \sigma$ for all k. To resolve the identification problem associated with the likelihood function (see HLM, 1992), this common value of σ is set equal to unity.

The choice of the state space dimension m for a given stock is guided by the empirical distribution of spreads. For example, if all spreads are equal to 1 tick, 2 ticks, or 3 ticks, letting $m = 3$ would be appropriate and only the two partition boundaries α_1 and α_2 would need to be estimated. Table B.6 in Appendix B suggests that all stocks included in the Dow Jones Industrial Average in 1988 should allow the estimation of a 4-state model, i.e., $m = 4$, and that most stocks should allow the estimation of a 5-state model, i.e., $m = 5$.[6] Here the first four states correspond to a bid-ask spread of 1, 2, 3, and 4 ticks, respectively, and the fifth state corresponds to a bid-ask spread of 5 ticks or more.

Finally, the explanatory variables X_k need to be specified. The first variable to be included is the lagged value of the bid-ask spread. The exchange requires specialists to maintain price continuity which arguably implies that bid and ask prices are "sticky" and do not move around very much. Therefore, it seems sensible to include the lagged bid-ask spread S_{k-1} as an explanatory variable. High values of S_{k-1} are expected to be associated with high values of S_k.

It was noted in section 4.2 that some prominent theories suggest that transaction

[6] An obvious exception is Navistar (NAV), which has no spread observations exceeding 4 ticks. Thus the largest possible state space dimension for Navistar is $m = 4$.

volumes matter, so I include transaction-volume variables. To distinguish whether buy (sell) volumes, i.e., volumes associated with transactions that were classified as buys (sells), have different effects compared to sell (buy) volumes and compared to volumes associated with transactions that were classified as indeterminate (neutral), I include these variables separately. Volumes are expected to affect the bid-ask spread positively in the sense that higher volumes are associated with higher levels of the bid-ask spread. In the context of the adverse-selection theory, larger volumes are an indication of more activity by informed traders. It was suggested by Glosten and Milgrom (1985) that lagged volume may also matter, so I include this variable as well. To facilitate comparisons across stocks that are trading at different price levels, all volume variables used in this study are actually *dollar volumes* as opposed to *share volumes.*

The dollar-volume variables are included to capture one important aspect of the intensity of the trading process. Two additional characteristics of the trading process are included, namely the number of transactions between quote revisions and the length of time between subsequent quote revisions. The latter variable is included to allow for clock-time effects. It is perhaps plausible that the time between trades enters negatively, since the arrival of new limit orders may tend to narrow the spread and more new orders can be expected in a longer time interval. One argument to include the number of transactions between quote revisions follows along the lines of Demsetz (1968), who reasoned that the fundamental force to reduce the spread is the frequency of transacting. The number of trades between quote revisions is an intraday measure of transaction activity. Another reason for including these two variables is that Easley and O'Hara (1992) argue that the time between trades matters for the behavior of the bid-ask spread. This issue will be investigated in more detail in subsection 4.4.2.

Another set of variables is included to capture a second very important aspect of the trading process, the *volatility.* A trade-based measure of stock-price volatility is problematic because of bid-ask bounce, see McInish and Wood (1992). However, a quote-based measure of stock-price volatility can readily be constructed. Recognizing that quotes move up and down with transaction prices, movements of the quotes obviously capture the movement of transaction prices. I take the change in midquote as a risk or volatility measure. Specifically, define the measure δ_k by

$$\delta_k \equiv |(A_k + B_k)/2 - (A_{k-1} + B_{k-1})/2|,$$

where A_i and B_i are the ask price and bid price of the ith quote observation. In other words, the measure δ_k is simply the absolute change in the midquote between quote revisions $k-1$ and k. Since the measure δ_k is determined contemporaneously with the spread S_k, I include the lagged values δ_{k-1} and δ_{k-2}. It is expected that these measures of stock-price volatility are positively related to the bid-ask spread, since higher volatility makes the stock riskier for the specialist and therefore would prompt him to increase the immediacy premium, i.e., the bid-ask spread.

As mentioned before, the specialist's quote consists not only of bid and ask prices, but also of the number of shares available at the bid, the *bid depth*, and the number of shares available at the ask, the *ask depth*. I include lagged values of both the bid depth and the ask depth. Intuitively, a low depth at the ask side will on average precede an upward revision of the ask price and since bid and ask prices do not necessarily move in tandem, the spread will on average increase. Conversely, if the ask depth is very high, the spread is not likely to increase and it may well decrease on average. A similar story applies to the depth at the bid side so that the prediction is that higher (lower) values of lagged depths tend to be associated with lower (higher) values of current bid-ask spreads.[7]

The next two variables are intended to capture known time-of-day effects of bid-ask spreads. McInish and Wood (1992) document that for a large sample of NYSE stocks, spreads have a reverse J-shape pattern over the course of the trading day: spreads are large at the beginning and end of the trading day compared to the middle of the trading day. See Figure 4.1 for a plot of the equally-weighted percentage spread by trading minute for my sample of 30 stocks. This plot was constructed using the method described in McInish and Wood (1992). Incidentally, Brock and Kleidon (1992) provide a theoretical explanation of this time-of-day effect that focuses on the importance of non-trading periods. To capture

[7]Lee, Mucklow, and Ready (1993) were the first to emphasize the importance of the depth as another aspect of market liquidity.

this time-of-day effect and to see whether the observed pattern holds for individual stocks, I include two indicator variables: one for the first hour of the trading day and one for the last hour of the trading day. If the reverse J-shape pattern described above holds for individual stocks, both of these indicator variables will be positively related to the bid-ask spread.

Finally, I include a competition variable. Recall that the dynamics of the spread studied here apply to the spread prevailing on the NYSE. Trading of stocks listed on the NYSE can occur on the regional exchanges and also on the OTC market. These markets, therefore, offer direct competition to NYSE specialists, who indeed keep track of competing quotes originating from the other exchanges. My measure of competition is a very simple one: an indicator variable that takes the value 1 if one or more trades occurred away from the NYSE since the previous NYSE quote and 0, otherwise. This variable is expected to be negatively related to the (NYSE) bid-ask spread, since NYSE specialists would want to tighten their spread to avoid a diversion of trading volume away from the NYSE floor.

My specification for equation (4.1) is thus given by the following:[8]

$$\begin{aligned} S_k^* = {} & \beta_1 S_{k-1} + \beta_2\,\mathrm{BDOL}_k + \beta_3\,\mathrm{SDOL}_k + \beta_4\,\mathrm{NDOL}_k + \beta_5\,\mathrm{DOL}_{k-1} + \\ & \beta_6\,\mathrm{DEPTHB}_{k-1} + \beta_7\,\mathrm{DEPTHA}_{k-1} + \beta_8\,\mathrm{NTRANS}_k + \beta_9\,\mathrm{DTIME}_k + \\ & \beta_{10}\,\mathrm{RISK}_{k-1} + \beta_{11}\,\mathrm{RISK}_{k-2} + \beta_{12}\,\mathrm{HOUR1}_k + \beta_{13}\,\mathrm{HOURN}_k + \\ & \beta_{14}\,\mathrm{OFFEXCH}_k + \epsilon_k, \end{aligned} \tag{4.4}$$

where the variables are defined as follows:[9]

[8] One variable that is *not* included in my specification is the absolute (contemporaneous or lagged) price level. Although it is plausible that the spread increases with the price, this effect is not measurable on a quote-by-quote basis.

[9] To eliminate the effects of extreme outliers, I truncate the dollar-volume variables for each stock at the 99.5th percentile of its empirical dollar-volume distribution. Similarly, I truncate the volatility measures at 8 ticks for all stocks.

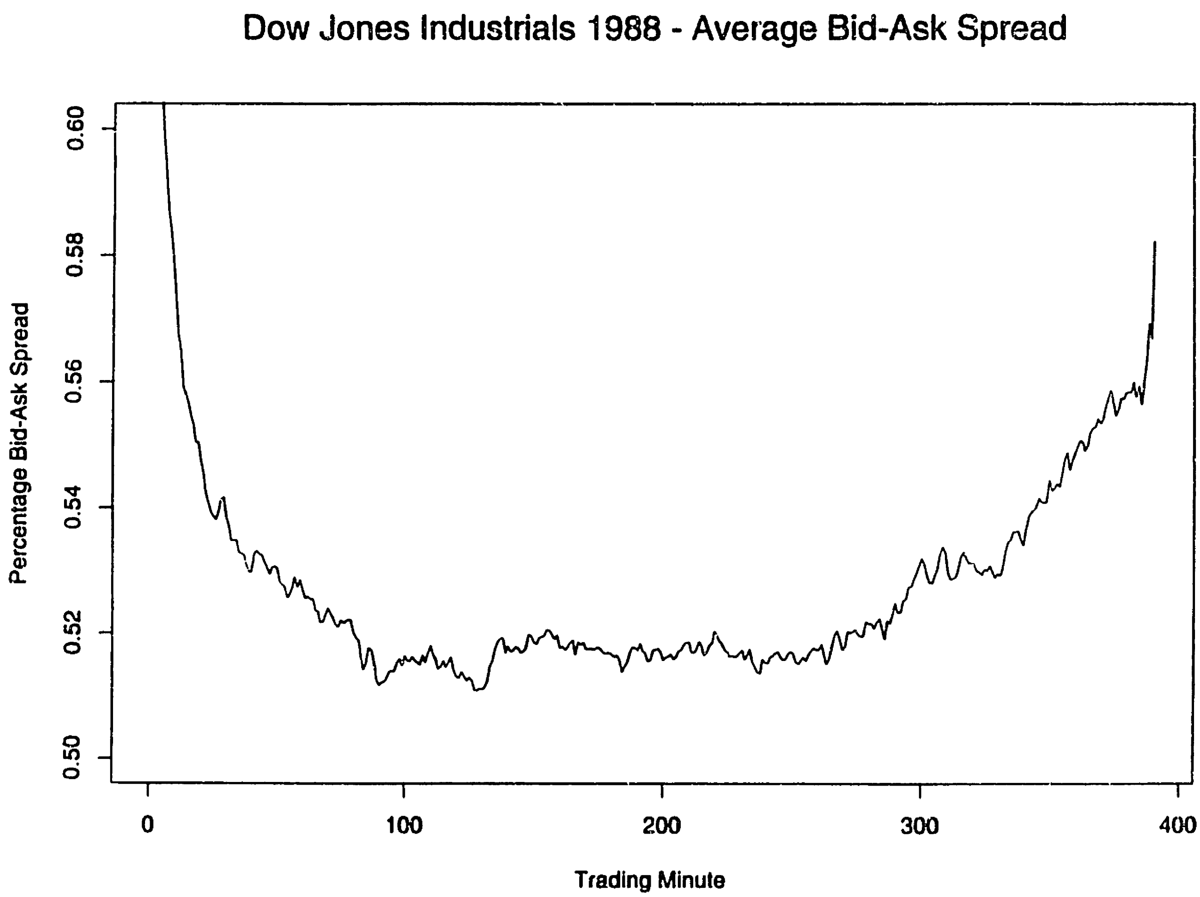


Figure 4.1: Minute-by-Minute Average Bid-Ask Spread

S_{k-1}	The lag of the bid-ask spread, denominated in ticks.
BDOL_k	Natural logarithm of dollar volume of *buy* transactions between quote revisions $k-1$ and k, where dollar volume is denominated in \$1,000's of dollars.
SDOL_k	Same as BDOL_k, but for *sell* transactions.
NDOL_k	Same as BDOL_k, but for *neutral* (indeterminate) transactions.
DOL_{k-1}	Logarithm of total dollar volume between quote revisions $k-2$ and $k-1$, where dollar volume is again denominated in units of \$1,000.
DEPTHB_{k-1}	Logarithm of the lag of the quoted depth at the bid side, denominated in 100's of shares.
DEPTHA_{k-1}	Logarithm of the lag of the quoted depth at the ask side, denominated in 100's of shares.
NTRANS_k	Number of transactions between quote revisions $k-1$ and k.
DTIME_k	Interarrival time in seconds between quote revisions $k-1$ and k.
RISK_l	Absolute change in midquote between quote revisions $l-1$ and l, denominated in ticks.
HOUR1_k	Indicator variable for the first hour of the trading day.
HOURN_k	Indicator variable for the last hour of the trading day.
OFFEXCH_k	Indicator variable for trading off the NYSE between quote revisions $k-1$ and k.

4.4 Results

In this section I summarize the empirical results of estimating the parameters of the ordered probit model as specified in equation (4.4).

Due to space limitations I discuss the data used in this study at length in a separate data appendix (see Appendix B). The sample consists of the 30 stocks included in the Dow Jones Industrial Average in 1988, see Table 4.1. It is clearly not practical to report the complete results for all 30 stocks in my sample. I therefore present summary information concerning the signs of the estimated β-coefficients (positive or negative) in Table 4.2 and the complete maximum-likelihood results for a subset of 4 stocks in Table 4.3.

The columns below the ticker symbols in Table 4.3 report the parameter estimates and the columns labeled z report the ratios of the parameter estimate and the asymptotic standard error. Under the null hypothesis that the coefficient is zero, these z-statistics have an asymptotic standard normal distribution.

The standard errors associated with the partition boundaries α_i are quite small. Indeed, reporting these standard errors instead of the z-statistics would perhaps be more appropriate, as there is no reason to consider the hypothesis that $\alpha_i = 0$.

It seems hardly surprising that the sign of $\hat{\beta}_1$ is positive for all stocks: higher lagged spreads S_{k-1} are associated with higher current spreads S_k.

The results for the dollar-volume variables are quite interesting. In general, we see that larger volumes are associated with larger bid-ask spreads. However, "signed" dollar volume (BDOL_k for buys and SDOL_k for sells) apparently has a much stronger effect than "unsigned" dollar volume (NDOL_k), as exemplified by the consistently large values of $\hat{\beta}_2$ and $\hat{\beta}_3$ in comparison to $\hat{\beta}_4$. Similarly, lagged dollar volume (DOL_{k-1}) has a negligible effect compared to "current" volume. Subsection 4.4.1 contains a more detailed analysis of the relation between spreads and dollar volumes.

As expected, the quoted-depth variables (DEPTHB and DEPTHA) enter negatively. As far as the timing of trades is concerned, both the number of transactions (NTRANS) and the time between quote revisions (DTIME) generally enter negatively. I explore the role of the time between trades more closely in subsection 4.4.2. Note that $\beta_8 < 0$ is consistent with Demsetz (1968).

The sign pattern of the volatility measure (RISK) is largely as expected. The sign corresponding to this measure is almost always positive, meaning that higher volatility is associated with higher bid-ask spreads. A negative value for the measure RISK_{k-1} is found just once (UK) and twice for RISK_{k-2} (NAV and T), but these parameter estimates are all quite small (in absolute value) and insignificant. The economic importance of volatility for the behavior of the bid-ask spread is explored in more detail in subsection 4.4.3.

A slightly surprising result is that the time-of-day indicators do not consistently have

Name	Ticker
ALLIED SIGNAL INC	ALD
ALUMINUM CO AMER	AA
AMERICAN EXPRESS CO	AXP
AMERICAN TEL & TELEG CO	T
BETHLEHEM STL CORP	BS
BOEING CO	BA
CHEVRON CORP	CHV
COCA COLA CO	KO
DU PONT E I DE NEMOURS & CO	DD
EASTMAN KODAK CO	EK
EXXON CORP	XON
GENERAL ELEC CO	GE
GENERAL MTRS CORP	GM
GOODYEAR TIRE & RUBR CO	GT
INTERNATIONAL BUSINESS MACHS	IBM
INTERNATIONAL PAPER CO	IP
MCDONALDS CORP	MCD
MERCK & CO INC	MRK
MINNESOTA MNG & MFG CO	MMM
NAVISTAR INTL CORP	NAV
PHILIP MORRIS COS INS	MO
PRIMERICA CORP	PA
PROCTER & GAMBLE CO	PG
SEARS ROEBUCK & CO	S
TEXACO INC	TX
UNION CARBIDE CORP	UK
UNITED TECHNOLOGIES CORP	UTX
USX CORP	X
WESTINGHOUSE ELEC CORP	WX
WOOLWORTH F W CO	Z

Table 4.1: Dow Jones Industrials

Names and ticker symbols of the 30 stocks that were part of the Dow Jones Industrial Average (DJIA) in 1988.

Parameter	$\beta > 0$	$\beta < 0$
β_1: S_{k-1}	30	0
β_2: BDOL_k	30	0
β_3: SDOL_k	30	0
β_4: NDOL_k	23	7
β_5: DOL_{k-1}	19	11
β_6: DEPTHB_{k-1}	0	30
β_7: DEPTHA_{k-1}	0	30
β_8: NTRANS_k	2	28
β_9: DTIME_k	0	30
β_{10}: RISK_{k-1}	29	1
β_{11}: RISK_{k-2}	28	2
β_{12}: HOUR1_k	17	13
β_{13}: HOURN_k	26	4
β_{14}: OFFEXCH_k	1	29

Table 4.2: Summary of Estimation Results

Summary of the signs of the estimated coefficients corresponding to the specification of equation (4.4) for the 30 stocks in the sample.

the expected signs. The parameter estimate of the indicator variable HOUR1 for the first hour of trading is positive for roughly one half of the sample (17 out of 30). Moreover, the parameter estimates as well as z-values are typically quite small. The largest z-value (in absolute value) across all stocks is -3.45 for IBM (149,747 observations). The parameter estimates corresponding to the end-of-day indicator HOURN is more in line with the expected effect that spreads tend to be higher at the end of the trading day. The z-values, however, are not impressive given the sample sizes. Overall the estimates suggest that higher spreads at the beginning of the day should perhaps be attributed to other factors, such as high volatility or high volume at the beginning of the trading day (Admati and Pfleiderer, 1988).

The results for the competition variable OFFEXCH are reasonably consistent across stocks: for 29 out of 30 stocks the estimated parameter is negative, implying that off-NYSE trades tend to foreshadow spread decreases on the NYSE.[10] The magnitude of the

[10]The only exception is Du Pont (DD), for which $\hat{\beta}_{14}$ is positive but statistically insignificant.

Parameter	BS	z	GE	z	IBM	z	PG	z
α_1	1.141	20.11	1.464	42.49	1.342	64.72	0.485	17.15
α_2	3.592	51.03	3.998	100.81	3.234	143.16	2.021	70.07
α_3	4.733	42.39	5.272	91.29	4.531	152.71	3.269	105.97
α_4	5.120	39.85	6.385	63.09	5.525	102.25	4.883	109.46
β_1: S_{k-1}	1.143	58.14	1.607	165.19	1.154	205.89	0.911	128.77
β_2: BDOL_k	0.225	41.58	0.173	62.36	0.120	68.83	0.178	50.00
β_3: SDOL_k	0.240	42.70	0.186	67.45	0.115	65.21	0.147	39.38
β_4: NDOL_k	0.096	13.28	0.018	5.14	0.010	4.35	0.103	19.66
β_5: DOL_{k-1}	0.025	6.32	0.014	7.36	0.011	8.28	-0.007	-2.67
β_6: DEPTHB_{k-1}	-0.156	-17.51	-0.178	-35.95	-0.141	-40.55	-0.167	-33.16
β_7: DEPTHA_{k-1}	-0.140	-16.17	-0.167	-34.70	-0.124	-36.56	-0.184	-36.31
β_8: NTRANS_k	-0.141	-20.71	-0.081	-18.35	-0.014	-4.63	-0.088	-12.19
β_9: $\text{DTIME}_k/100$	-0.013	-5.11	-0.051	-9.41	-0.129	-15.11	-0.036	-10.90
β_{10}: RISK_{k-1}	0.101	4.02	0.107	6.01	0.061	6.06	0.028	2.24
β_{11}: RISK_{k-2}	0.063	2.24	0.075	4.20	0.101	10.26	0.081	7.16
β_{12}: HOUR1_k	0.031	1.48	-0.036	-3.00	-0.031	-3.45	0.033	2.29
β_{13}: HOURN_k	-0.033	-1.43	0.034	2.79	0.066	7.60	0.061	4.18
β_{14}: OFFEXCH_k	-0.050	-2.71	-0.056	-5.39	-0.008	-1.04	-0.000	-0.03

Table 4.3: Maximum Likelihood Results

Maximum likelihood estimates of the ordered probit model for bid-ask spreads for Bethlehem Steel Corporation (BS – 26,503 spreads), General Electric Company (GE – 93,460 spreads), International Business Machines Corporation (IBM – 149,747 spreads), and Procter & Gamble Company (PG – 41,162 spreads), for the sample period from 4 January 1988 to 30 December 1988. Each z-statistic is asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The ordered probit specification contains 5 states (1 tick, 2 ticks, 3 ticks, 4 ticks, and 5 ticks or more) for these stocks.

effect appears to be tiny, however.

4.4.1 Spreads and Trading Volumes

By examining the signs of the estimated parameters corresponding to the dollar-volume variables, I reached the qualitative conclusion that volumes matter for explaining the dynamic behavior of the bid-ask spread. Following large buys or sells, the bid-ask spread was shown to have a tendency to increase. The purpose of this section is to determine the magnitude or economic significance of these effects and to make comparisons across stocks.

Before doing so, however, I briefly digress to discuss the connection between the analysis here and the implications of the adverse-selection model of Easley and O'Hara (1987). In their model volume is correlated with private information in the sense that informed traders tend to trade large quantities. Market makers react by setting a larger bid-ask spread for large trades.

To examine this hypothesis empirically, we would need a schedule of bid-ask spreads with a separate bid-ask spread for each trading volume. Unfortunately, this information is not available. The only spread information available is the spread for transactions up to the quoted volumes.

Easley and O'Hara (1987) do not explicitly derive conditions under which the spread for the smaller quantity will increase subsequent to a block trade (trades are either small or large in their model). However, this appears to be a plausible implication of their basic model and the remainder of this section is an empirical examination of the presence of this effect.

Thus instead of examining whether large transactions are made at larger spreads, I examine to what extent the quoted spread responds to trading activity in the time interval between successive quotes. We saw that "signed" (dollar) volumes have a greater effect than both "unsigned" and lagged (dollar) volumes. I therefore focus on "signed" dollar volumes in this section and to avoid duplication I restrict myself to dollar volumes associated with buys. At the end of this subsection, however, I do consider the question whether buys and sell have different effects by examining whether $\beta_2 = \beta_3$.

To determine the effect of buy transactions of various sizes on the bid-ask spread, I follow the same method used by HLM (1992) to determine the price impact of trades of various sizes. Briefly, by choosing values for the regressors X_k we establish a "scenario" under which we calculate the probability distribution of the spread implied by the ordered probit maximum-likelihood estimates. Then we calculate the incremental effect on the bid-ask spread of successively larger buy transactions. By considering the effects in percentage terms it is then possible to make comparisons across stocks and to determine the quantitative magnitudes of these effects.

Since the focus of the analysis here is not on time-of-day effects or competition effects, I set the variables HOUR1_k, HOURN_k, and OFFEXCH_k equal to 0 throughout. Further, I assume a single buy transaction (i.e., $\text{NTRANS}_k = 1$ so that $\text{SDOL}_k = \text{NDOL}_k = 0$). For each stock, the remaining variables (except BDOL_k of course) are set equal to their sample medians.

To see the size of the effect of various buys, I then vary the buy dollar volume from a base case of \$10,000 to \$500,000 and explicitly calculate the implied spread probabilities using equation (4.3). The results are summarized in Table 4.4 and Figure 4.2 for a subset of 7 stocks spanning the range of observed sensitivities.

Table 4.4 gives the mean $E[S_k]$ of the spread distribution for the base case of a \$10,000 buy and the change in mean $\Delta E[S_k]$ associated with buy volumes exceeding the base amount of \$10,000. The upper panel gives the *absolute* magnitude of the effects in units of a tick (1 tick is \$0.125) and the lower panel gives the *relative* magnitude of the effects as a percentage of the stock price. For simplicity, the stock price is taken as a simple average of the highest and the lowest price observed in the whole year.

To illustrate the results in Table 4.4, consider Procter & Gamble (PG). For PG, the expected spread for a base case of a \$10,000 buy is 1.921 ticks, which corresponds to 0.302% of PG's average stock price in 1988. A buy of \$500,000 instead of \$10,000 gives a spread increase of 0.464 ticks or 0.073% of PG's price. For the case of a \$500,000 buy, a comparison of the last row of Table 4.4 shows that the effects vary from 0.017% for IBM

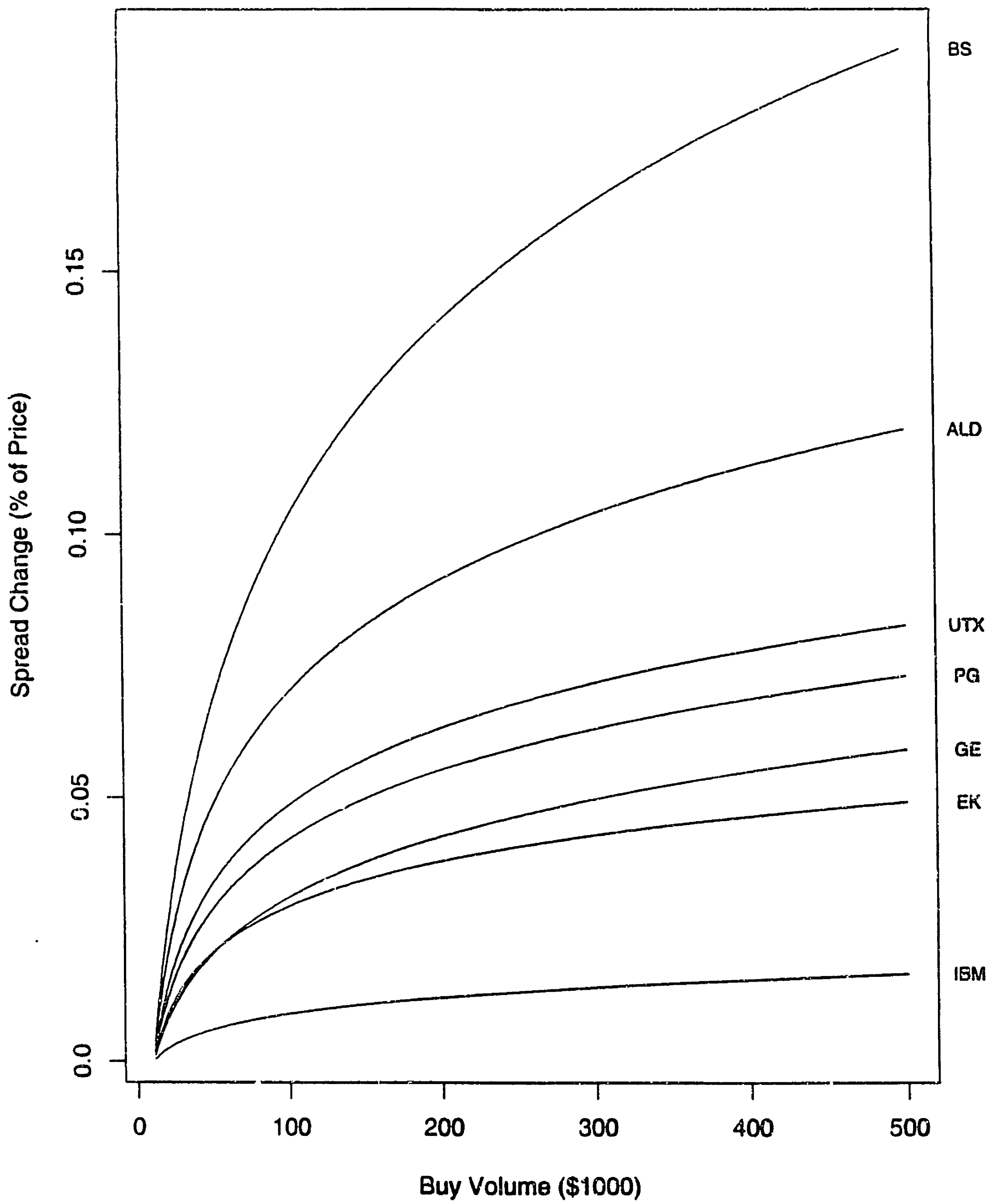


Figure 4.2: Spread Changes and Volumes

Buy ($)	Measure	ALD	BS	EK	GE	IBM	PG	UTX
					Ticks			
10,000	$E[S_k]$	1.966	1.180	1.735	1.138	1.182	1.921	1.856
25,000	$\Delta E[S_k]$	0.072	0.060	0.045	0.038	0.031	0.104	0.059
50,000	$\Delta E[S_k]$	0.127	0.111	0.077	0.071	0.057	0.185	0.103
100,000	$\Delta E[S_k]$	0.182	0.168	0.110	0.107	0.085	0.267	0.147
250,000	$\Delta E[S_k]$	0.255	0.248	0.152	0.160	0.124	0.378	0.206
500,000	$\Delta E[S_k]$	0.311	0.313	0.183	0.203	0.155	0.464	0.250
					% of Price			
10,000	$E[S_k]$	0.758	0.726	0.467	0.331	0.126	0.302	0.614
25,000	$\Delta E[S_k]$	0.028	0.037	0.012	0.011	0.003	0.016	0.019
50,000	$\Delta E[S_k]$	0.049	0.068	0.021	0.021	0.006	0.029	0.034
100,000	$\Delta E[S_k]$	0.070	0.103	0.029	0.031	0.009	0.042	0.049
250,000	$\Delta E[S_k]$	0.098	0.153	0.041	0.047	0.013	0.060	0.068
500,000	$\Delta E[S_k]$	0.120	0.192	0.049	0.059	0.017	0.073	0.083

Table 4.4: Effect of Buy Volume on Bid-Ask Spreads

to 0.192% for BS. These effects are substantial relative to the magnitude of the spread itself.

Finally, I consider the issue raised by Allen and Gorton (1991) regarding the asymmetry between purchases and sales, see section 4.2. The null hypothesis based on Allen and Gorton (1991) is that $\beta_2 = \beta_3$.

For my sample of 30 stocks, the maximum-likelihood estimates are such that $\hat{\beta}_2 > \hat{\beta}_3$ for 23 stocks and $\hat{\beta}_2 < \hat{\beta}_3$ for 7 stocks. Under the null hypothesis that $\beta_2 = \beta_3$, the quantity $(\hat{\beta}_2 - \hat{\beta}_3)^2/(\hat{\sigma}^2_{\beta_2} + \hat{\sigma}^2_{\beta_3} - 2\hat{\sigma}_{\beta_2,\beta_3})$ has an asymptotic χ^2 distribution with 1 degree of freedom. Calculation of these quantities for the sample of 30 stocks yields that for 14 (10) stocks $\hat{\beta}_2 > \hat{\beta}_3$ and for 4 (3) stocks $\hat{\beta}_2 > \hat{\beta}_3$ at the 5% (1%) significance level. The conclusion is that there appears to be some indication that buys and sells affect the spread differently, in accordance with the reasoning of Allen and Gorton (1991). The economic significance of the difference seems to be quite minor, however.

4.4.2 Does the Time Between Trades Matter?

In a model with event uncertainty, Easley and O'Hara (1992) show that the time between trades affects spreads. Specifically, spreads are shown to decrease as the time between trades increases. The reason is that in their model the absence of trades in some time interval is more likely to occur if the information event does not occur. Thus the adverse-selection problem is less severe and the spread is adjusted accordingly.

My empirical evidence is consistent with the prediction of Easley and O'Hara (1992), but the evidence is not very strong.

The specific prediction to be examined here is that the spread will decrease the longer the time between transactions. Note that my observations are the times at which quotes are revised. The time between successive quote revisions is DTIME and the number of transactions between successive transactions is NTRANS. Thus the time between transactions could be measured as NTRANS/DTIME, a quantity that is not included in my specification. It is perhaps not surprising that the variables NTRANS and DTIME are highly correlated, since there will be more transactions in a longer time interval, ceteris paribus. Indeed, for my sample of 30 stocks these correlations range from 40.0% (PG) to 89.4% (T) with an average of 64.7% and a standard deviation of 12.0%.

Note that when the number of transactions between quote revisions is fixed, having $\beta_9 < 0$ (β_9 is the parameter corresponding to DTIME) would be consistent with Easley and O'Hara (1992). Conversely, when the time between quote revisions is fixed, having $\beta_8 > 0$ (β_8 is the parameter corresponding to NTRANS) would be consistent with this theory.

From Table 4.2 we see that the estimated values $\hat{\beta}_9$ are indeed negative for all 30 stocks, consistent with the theoretical prediction. However, the estimated values $\hat{\beta}_8$ are almost all negative (28 out of 30, with NAV and XON the only cases for which $\hat{\beta}_8$ is insignificantly positive). Thus the net effect on bid-ask spreads of longer times between trades is indeterminate for the vast majority of the stocks considered here.

To circumvent this difficulty, I reestimate the ordered probit model with the variable NTRANS/DTIME included instead of the two variables NTRANS and DTIME. Note that the quantity DTIME/NTRANS (the average time between trades) cannot be included directly, since there are many observations with NTRANS = 0. If the effect predicted by Easley and O'Hara (1992) is indeed present, we would expect to see a positive coefficient for the parameter associated with the new variable NTRANS/DTIME. The empirical results give a positive value for 23 stocks and a negative value for the remaining 7 stocks. Since the z-values are mostly in the range from 1 to 6, this is seen as weak support for the prediction of Easley and O'Hara (1992) that the spread will decrease the longer the time between transactions.

4.4.3 Spreads and Stock-Price Volatility

Using the same method as in subsection 4.4.1, this subsection investigates the economic importance of volatility for the behavior of the bid-ask spread.

To evaluate the importance of volatility for the bid-ask spread, I specify market conditions with increasing price volatility by setting the values of the regressors δ_{k-1} and δ_{k-2} as follows: $\delta_{k-1} = \delta$ and $\delta_{k-2} = \delta$ for various values of δ.

As a base case I consider a "flat" market: $\delta = 0$. I then compare the incremental effect on the spread by considering successively more volatile markets: $\delta = 0.5$, $\delta = 1.0$, $\delta = 1.5$, and $\delta = 2.0$. A single transaction is assumed and all the other regressors are set equal to their medians.[11]

The results are summarized in Table 4.5, where the spread changes are given in percentage terms compared to the base case of a "flat" market ($\delta = 0$) for the full sample of 30 stocks. Note from Table 4.5 that the impacts of volatility on the spread are remarkably similar across stocks. It is seen from Table 4.5 that the volatility measure is economically non-trivial, although the effect is generally smaller than the effect associated with a large buy as demonstrated in subsection 4.4.1.

[11]Actually, it does not make sense to set the dollar-volume variables BDOL, SDOL, and NDOL equal to their sample medians (conditional on NTRANS = 1), because these medians are all 0 and BDOL = SDOL = NDOL = 0 is clearly inconsistent with NTRANS = 1. This issue is resolved by assuming that buy volume, sell volume, and neutral volume represent 40%, 40%, and 20%, respectively, of total volume and total volume is set equal to its sample median (conditional on NTRANS = 1).

Ticker	$E[S_k]$ $\delta = 0.0$	$\Delta E[S_k]$ $\delta = 0.5$	$\Delta E[S_k]$ $\delta = 1.0$	$\Delta E[S_k]$ $\delta = 1.5$	$\Delta E[S_k]$ $\delta = 2.0$
AA	0.534	0.018	0.036	0.055	0.075
ALD	0.836	0.022	0.045	0.067	0.090
AXP	0.666	0.021	0.043	0.065	0.087
BA	0.488	0.011	0.021	0.032	0.043
BS	0.851	0.020	0.040	0.061	0.082
CHV	0.434	0.009	0.018	0.028	0.037
DD	0.356	0.011	0.022	0.033	0.045
EK	0.514	0.017	0.035	0.054	0.074
GE	0.396	0.010	0.021	0.032	0.042
GM	0.317	0.007	0.014	0.021	0.028
GT	0.529	0.014	0.027	0.041	0.054
IBM	0.150	0.004	0.008	0.011	0.016
IP	0.447	0.015	0.031	0.047	0.063
KO	0.481	0.015	0.029	0.044	0.059
MCD	0.562	0.014	0.029	0.043	0.058
MMM	0.487	0.009	0.019	0.028	0.037
MO	0.314	0.009	0.018	0.028	0.037
MRK	0.292	0.007	0.015	0.022	0.030
NAV	2.354	0.011	0.024	0.037	0.051
PA	0.589	0.006	0.013	0.020	0.026
PG	0.382	0.006	0.012	0.018	0.024
S	0.463	0.018	0.037	0.056	0.075
T	0.577	0.011	0.023	0.036	0.049
TX	0.413	0.008	0.016	0.024	0.032
UK	0.692	0.009	0.018	0.027	0.037
UTX	0.703	0.013	0.027	0.040	0.054
WX	0.381	0.018	0.036	0.054	0.073
X	0.586	0.014	0.028	0.042	0.057
XON	0.462	0.019	0.038	0.056	0.074
Z	0.436	0.010	0.021	0.031	0.041

Table 4.5: Effect of Price Volatility on Percentage Bid-Ask Spreads

Illustration of the effect of price volatility on percentage bid-ask spreads. Taking a "flat" market as a base case, the ordered probit model is used to estimate the incremental effect of more volatile prices on the bid-ask spread. This is accomplished by letting $\delta_{k-1} = \delta_{k-2} = \delta$ for $\delta = 0.5$, 1.0, 1.5, and 2.0, while keeping the remaining regressors fixed.

4.5 Forecasting

One interesting question is how ordered probit can be used for forecasting and how the performance of ordered probit compares to alternative estimation methods.

The most obvious competitor to ordered probit is OLS (Ordinary Least Squares). I therefore carried out the OLS estimation corresponding to my specification of X_k. In other words, I regressed S_k on a constant and on the 14 regressors specified in equation (4.4).

Since it is most interesting to consider out-of-sample forecasting, I estimated the parameters for both ordered probit and OLS using data for the first 6 months of the year for all of the 30 sample stocks. These estimates were then used to forecast the actual values of the spread in the remaining 6 months of the year.

To compare the performance between ordered probit and OLS, we need a performance measure. One widely-used measure is the root mean square prediction error RMSPE, defined by

$$\text{RMSPE} = \sqrt{(\text{Mean of PE})^2 + (\text{Standard Deviation of PE})^2},$$

where PE is the prediction error, i.e., the actual spread S_k minus the fitted value $\hat{S}_k$. Calculating fitted values for OLS is trivial once the parameters have been estimated, since $\hat{S}_k = X_k'\hat{\beta}$. For ordered probit, we need to calculate the estimated probabilities of observing the various bid-ask spread values using equation (4.3) and then calculate the mean of this distribution to get the desired spread forecast.

The values of the RMSPE for each of the 30 stocks are reported in Table 4.6 for both OLS and ordered probit. Whereas OLS has smaller RMSPE than ordered probit in 19 out of 30 cases, the differences are quite small and insignificant: the z-value of the average difference between the RMSPE for ordered probit and OLS is 0.93.

One might question whether the mean squared error is the best yardstick for comparison. Indeed, OLS is designed to minimize mean squared error. An alternative measure that does recognize the discrete nature of spreads is the *hit rate*, i.e., the fraction of spreads forecasted correctly. For ordered probit, a correct forecast is achieved if the mode of the estimated probabilities is equal to the actual spread S_k. For OLS, a correct forecast is achieved if the rounded value of the conditional mean $X_k'\hat{\beta}$ is equal to S_k.

The results for the hit rate comparison are summarized in Table 4.7. A comparison of hit rates between OLS and ordered probit turns out to be more favorable for ordered probit than RMSPE: ordered probit beats OLS with 17 to 12 (with 1 tie). The differences are again quite small, however. The z-value of the average difference between the hit rate for ordered probit and OLS is 1.73, which is insignificant at the 5 percent level.

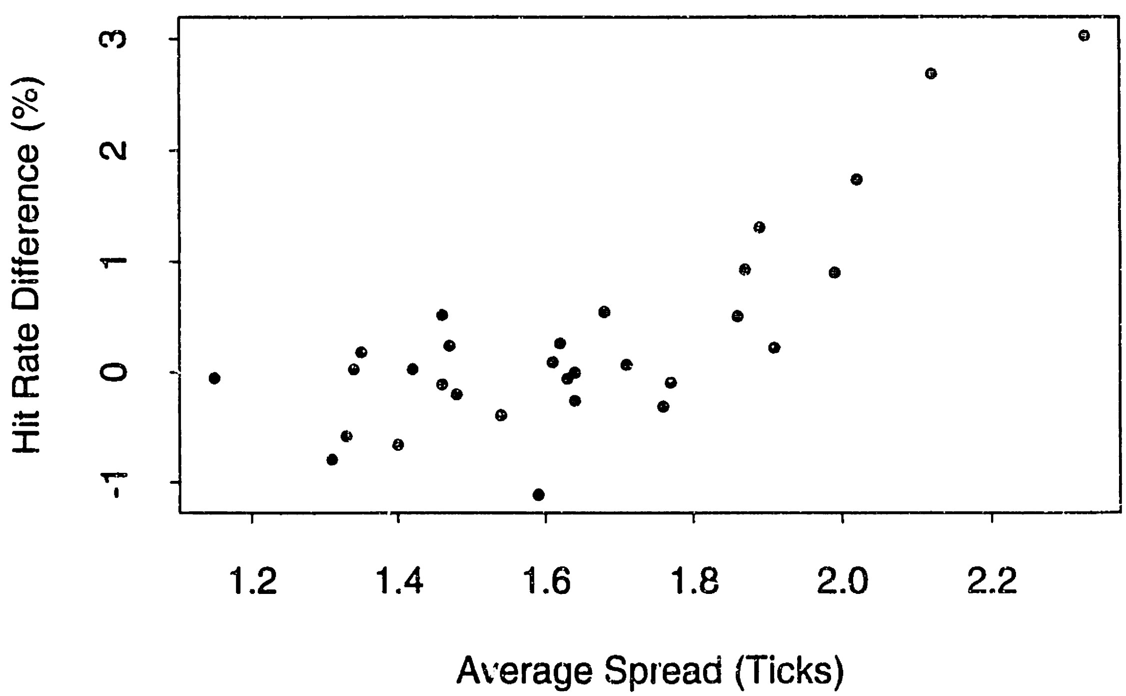


Figure 4.3: Hit Rate Difference Between Probit and OLS versus Average Spread

Analyzing the results in Table 4.7, I hypothesized that the difference in hit-rate performance between ordered probit and OLS might be related to the average spread of the stocks. See Figure 4.3 above for a scatter diagram of the hit rate difference (ordered probit hit rate minus OLS hit rate) against the average bid-ask spread (in ticks, see Table B.6) for my sample of 30 stocks. There appears to be a strong positive relation between the hit rate difference and the average spread. Indeed, if y denotes the percentage hit rate difference and x denotes the average spread in ticks, then an OLS regression of y on x for

Ticker	RMSPE OLS	RMSPE Probit	N_1	N_2
AA	0.5363	0.5375	9,976	7,033
ALD	0.5596	0.5609	8,126	6,039
AXP	0.4612	0.4613	15,622	14,144
BA	0.4997	0.5001	12,187	14,859
BS	0.4418	0.4432	14,181	12,322
CHV	0.5327	0.5337	20,223	14,344
DD	0.6287	0.6332	26,853	22,874
EK	0.4778	0.4788	13,482	13,958
GE	0.3913	0.3923	49,947	43,513
GM	0.4906	0.4903	15,913	13,894
GT	0.6391	0.6386	7,412	7,558
IBM	0.4628	0.4663	80,025	69,722
IP	0.5120	0.5132	23,946	17,822
KO	0.4684	0.4692	15,855	12,648
MCD	0.5230	0.5218	13,901	11,738
MMM	0.5708	0.5760	30,161	23,824
MO	0.5957	0.5958	21,132	17,695
MRK	0.6055	0.5985	24,024	16,898
NAV	0.3471	0.3458	2,941	1,993
PA	0.4679	0.4639	10,711	10,176
PG	0.6323	0.6364	22,716	18,446
S	0.4897	0.4876	8,991	11,670
T	0.4343	0.4319	12,657	9,793
TX	0.4977	0.4976	14,816	7,576
UK	0.4300	0.4288	18,492	20,144
UTX	0.5092	0.5108	16,591	11,609
WX	0.5802	0.5834	22,477	16,104
X	0.4572	0.4573	16,309	13,224
XON	0.4794	0.4778	11,620	9,452
Z	0.5487	0.5515	21,814	12,705
mean	0.5090	0.5095		

Table 4.6: Forecasting Comparison OLS and Ordered Probit Using RMSPE

Values of the root mean square prediction error (RMSPE) for OLS and ordered probit. The first six months were used to estimate the parameters and the last six months were used to calculate the prediction errors. N_1 and N_2 are the numbers of observations in the first half of the year and the second half of the year, respectively.

Ticker	Hit Rate OLS	Hit Rate Probit	N_1	N_2
AA	0.6545	0.6536	9,976	7,033
ALD	0.6181	0.6274	8,126	6,039
AXP	0.6637	0.6640	15,622	14,144
BA	0.6605	0.6599	12,187	14,859
BS	0.6981	0.6923	14,181	12,322
CHV	0.6550	0.6557	20,223	14,344
DD	0.5873	0.5963	26,853	22,874
EK	0.6738	0.6747	13,482	13,958
GE	0.7832	0.7753	49,947	43,513
GM	0.6209	0.6235	15,913	13,894
GT	0.5266	0.5535	7,412	7,558
IBM	0.7064	0.6998	80,025	69,722
IP	0.6707	0.6681	23,946	17,822
KO	0.6941	0.6830	15,855	12,648
MCD	0.6332	0.6387	13,901	11,738
MMM	0.6403	0.6454	30,161	23,824
MO	0.5751	0.5882	21,132	17,695
MRK	0.5760	0.6063	24,024	16,898
NAV	0.8580	0.8575	2,941	1,993
PA	0.7215	0.7176	10,711	10,176
PG	0.5934	0.6108	22,716	18,446
S	0.6376	0.6356	8,991	11,670
T	0.7197	0.7200	12,657	9,793
TX	0.6007	0.6059	14,816	7,576
UK	0.7319	0.7337	18,492	20,144
UTX	0.6911	0.6880	16,591	11,609
WX	0.6394	0.6416	22,477	16,104
X	0.7091	0.7080	16,309	13,224
XON	0.6454	0.6478	11,620	9,452
Z	0.6223	0.6223	21,814	12,705
mean	0.6603	0.6631		

Table 4.7: Forecasting Comparison OLS and Ordered Probit Using Hit Rate

Values of the hit rate for OLS and ordered probit. The hit rate is defined as the fraction of spreads forecasted correctly. The OLS forecast is the rounded value of the estimated spread. The ordered probit forecast is the mode of the estimated spread distribution. The first six months were used to estimate the parameters and the last six months were used to calculate the hit rates. N_1 and N_2 are the numbers of observations in the first half of the year and the second half of the year, respectively.

the sample of 30 stocks gives the following (t-statistics in parentheses):

$$y = -4.06 + 2.64x, \qquad R^2 = 0.60$$
$$(-6.01) \quad (6.53)$$

The explanation is that for stocks with low bid-ask spreads, discreteness is not an important issue. Ordinary Least Squares is therefore an adequate estimation method. However, for stocks with high bid-ask spreads, the discreteness of the spread is more important and ordered probit is the preferred estimation method. To illustrate this in an extreme case, consider a forecasting experiment for Crown Cork & Seal (CCK), one of the stocks analyzed in Chapter 5 of this thesis. This stock had a high bid-ask spread during its pre-split period, about 4 ticks ($0.50), and it was estimated using $m = 7$ states. The results of out-of-sample forecasting of 1,984 observations with CCK were as follows. Using OLS, the RMSPE was 0.9639 and the hit rate was 0.5136 (1,019 hits). Using ordered probit, the RMSPE was 0.9647 and the hit rate was 0.5938 (1,178 hits). Thus the hit-rate difference in favor of ordered probit was 8.0%, even larger than a naive application of the regression result discussed above suggests.[12]

For CCK there is significant clustering of bid-ask spreads on even eighths—a phenomenon that OLS cannot capture. The difference in estimated probabilities for the spread between ordered probit and OLS is illustrated in Figure 4.4, which shows the first 4 predictions of the spread for the forecasting experiment described above. The four predictions on the left are made using ordered probit and the four predictions on the right are made using OLS. The forecasting experiment tells us that ordered probit is better able to predict the value of the next spread.

Finally, I give an example of the forecasting distribution for IBM, a stock that does not have a nonlinear or nonmonotonic spread distribution. The first four predictions for IBM are given in Figure 4.5. The predicted distributions look fairly similar, although ordered probit seems to produce distributions that are less spread out than OLS.

[12]Substituting $x = 4$ ticks yields an estimate of the hit-rate difference y of 6.5% in favor of ordered probit. Of course, application of this regression equation for x-values outside the range of the original sample (see Figure 4.3) is questionable.

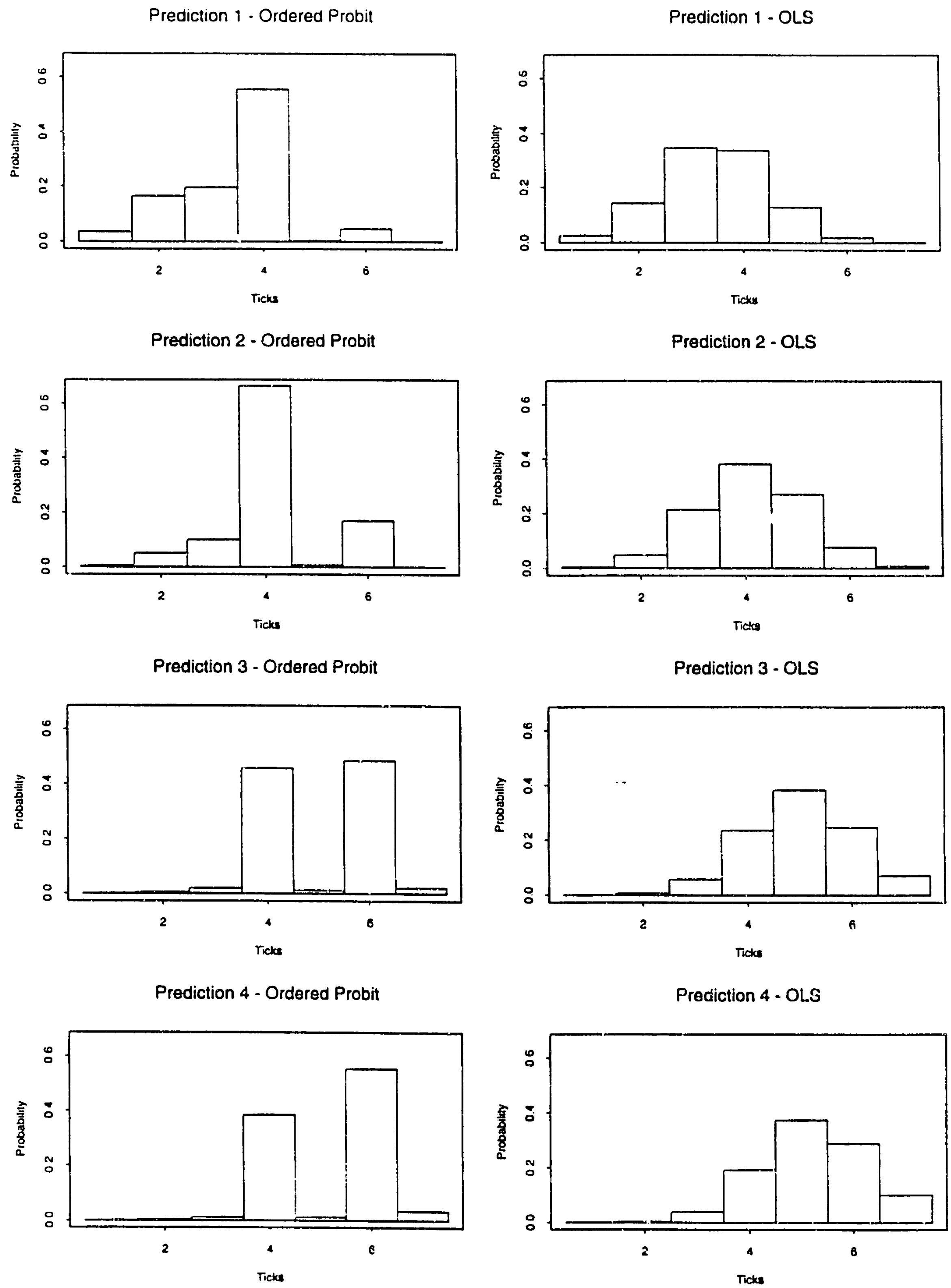


Figure 4.4: Comparison of Forecasted Spread Distributions for CCK

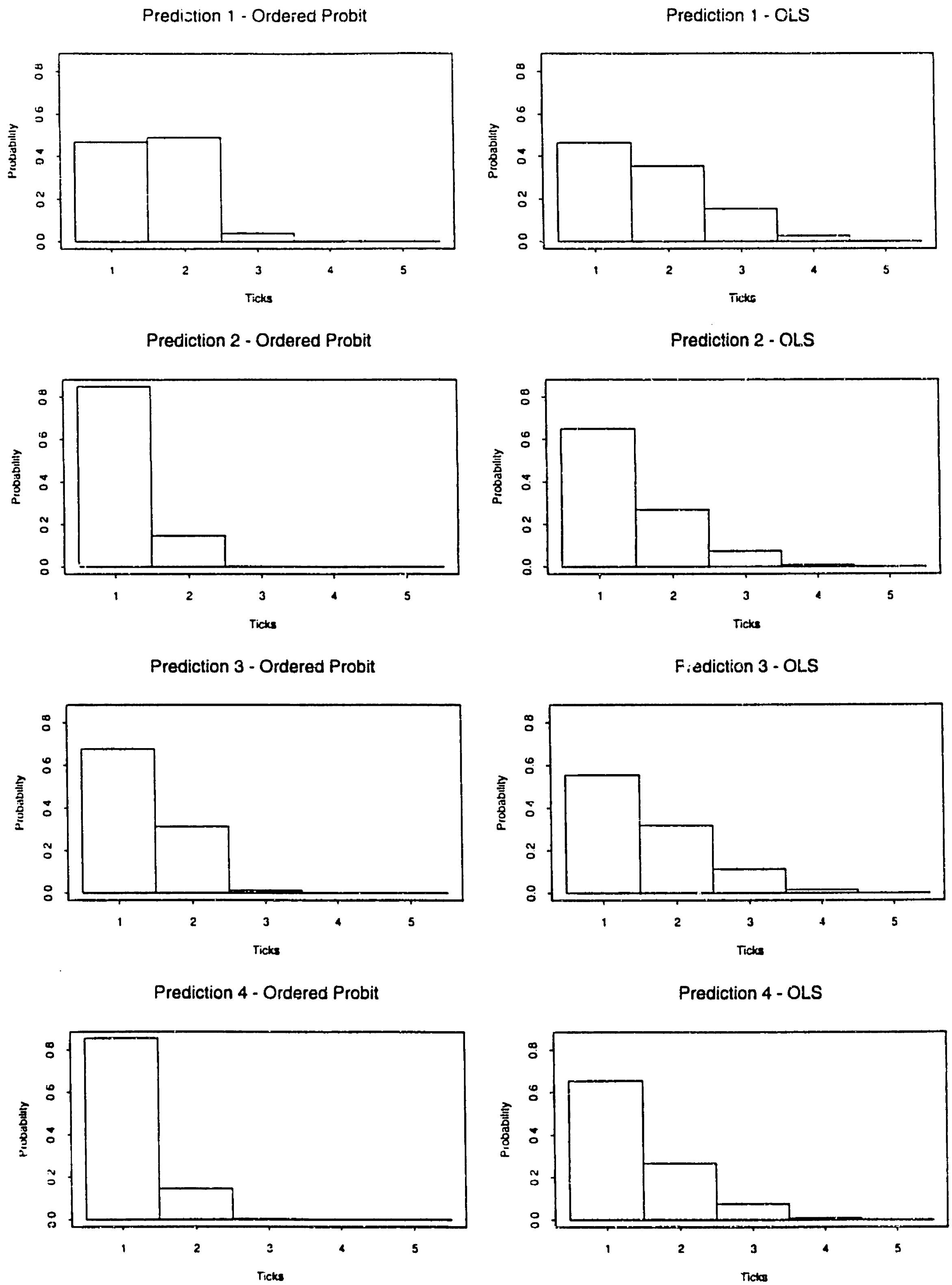


Figure 4.5: Comparison of Forecasted Spread Distributions for IBM

4.6 Conclusion

In this chapter I have studied the dynamic behavior of the bid-ask spread for the collection of NYSE stocks included in the Dow Jones Industrial Average. Since the spread is intimately related to market liquidity, this study is an attempt to explore how liquidity changes during the trading process of these stocks on a transaction-by-transaction and a quote-by-quote basis.

Several economic theories have been developed to derive properties of optimal quote-setting behavior of market makers and their implications for the bid-ask spread under various market conditions. For example, adverse-selection models focus on the importance of trading volumes as a signal of private information. My empirical results are consistent with the prediction that high trading volumes are associated with a subsequent widening of the bid-ask spread. The effect of trades occurring at the midpoint of the bid-ask spread is negligible in comparison to trades occurring at (or close) to the quotes. This is interpreted as evidence in support of the notion that trades initiated by market participants willing to pay (part of) the bid-ask spread have a different information content from trades that are most likely liquidity trades crossing at the midquote, a finding consistent with the ideas of the adverse-selection theory. The effect of lagged volume on spreads appears to be negligible. In other words, the information contained in the trading process is processed very quickly.

Another economic variable whose theoretical importance is supported in my study is stock-price volatility. Since specialists cannot always have a zero net position in the stocks for which they make a market, they bear the risk associated with unfavorable movements of the stock price. It is therefore plausible that the spread, their reward for bearing this risk, is related to price volatility. My empirical results suggest that this is indeed the case. I find strong evidence that spreads tend to increase when prices are volatile.

Finally, I show that ordered probit produces slightly better forecasts than OLS if performance is measured by the hit rate, the fraction of spreads forecasted correctly. This is particularly true for stocks whose bid-ask spread is relatively large.

Chapter 5

Stock Splits and Market Liquidity

5.1 Introduction

Stock splits have posed a number of challenging questions for researchers interested in the price behavior and the trading process of stocks. At first glance, stock splits appear to be purely cosmetic, since they do not change the cash flows of the firm (ignoring the costs associated with the split), nor do they change the ownership structure of the firm.[1] Nevertheless, several real effects have been observed to accompany splits. The most prominent example is the positive market reaction to splits. Numerous papers have addressed the question what the various effects are that are associated with stock splits, and how we can explain those effects.

In addition to direct survey of management's rationale for the stock-split decision, various hypotheses have been brought forward to explain the phenomenon of stock splits in light of the empirical evidence. According to the *signalling* hypothesis, managers use the stock-split decision to convey favorable information to investors. Those in favor of the *desirable trading range* hypothesis argue that the stock price should be kept in a certain range in order to achieve a compromise between the preferences of small investors

[1]The major components of the costs associated with a stock split are Exchange listing fees, mailing and distribution costs, and printing costs. Mr. Thomas E. Shea III, Assistant Treasurer of Merck, informed me that the total costs associated with Merck's 1992 three-for-one stock split were $560,522. This amount includes Exchange listing fees for New York ($250,000) and for Philadelphia ($1,250), mailing and distribution costs ($203,038), printing costs ($76,615), legal costs ($1,890), and other costs ($27,729). The costs of this split correspond roughly to $0.00145 per pre-split share of Merck.

on the one hand and wealthy investors and institutions on the other hand.[2] A more specific version of this hypothesis states that stock splits create "wider" markets (more shareholders), resulting in higher volume, a better price-discovery process, and lower bid-ask spreads. Stock splits enhance liquidity in this view.

Liquidity is a vague concept that is difficult to make precise. Most observers will agree that the stock market is more liquid than the real-estate market, but how do we compare the liquidity across stocks or across different time periods for the same stock?

This chapter is an empirical study of the liquidity effects associated with stock splits, focusing on measures of trading activity (e.g., the number of transactions and trading volumes), the bid-ask spread, and market depth for a sample of NYSE companies that split their stocks by 2-for-1 or greater in the two years 1988 and 1991. Basically, the question is whether stock splits have an effect on liquidity and if so, whether liquidity increases or decreases after the split. Although not all measures considered are in agreement, the broad picture that emerges is that liquidity actually *decreases* after a split. In light of the nontrivial costs of effecting a split, this is not a result to be ignored.

I use a broad spectrum of liquidity measures—some of which have not been considered before—to investigate the relation between stock splits and liquidity. The first measure considered is the (percentage) bid-ask spread. I use the *complete* record of all bid-ask spreads quoted for a sample of NYSE stocks, not just closing bid-ask spreads. There are two reasons why the use of all spreads is preferable to the use of closing bid-ask prices. First, since no trades will occur at these quotes, they are not necessarily reliable. Indeed, I show in an appendix that closing spreads are on average narrower than spreads quoted at the very end of the trading day. Second, even if closing quotes are representative for spreads at the end of the trading session, it has been shown by McInish and Wood (1992) that bid-ask spreads exhibit a U-shape pattern over the course of the trading day. In other words, the spread at the end of the day only gives partial information about the "typical" bid-ask spread of a stock. By aligning stocks by their split date and by looking at averages

[2] A curious "counterexample" to the desirable or optimal trading range hypothesis is the NYSE stock Berkshire Hathaway (ticker symbol BRK), whose chairman Warren Buffett opposes splitting the stock. BRK's closing price on December 20, 1993 was $16,810.

across stocks in event time, I show that the percentage bid-ask spread increases following the split.

I also consider the *depth* in addition to the bid-ask spread. A complete quote issued by a NYSE specialist consists not just of a bid price and an ask price, but also of the number of shares available at each of these prices. Ignoring the depth can be misleading. Clearly, spreads and depths are separate dimensions of market liquidity that should not be considered in isolation. The results for the depth are inconclusive. It appears that the post-split depth decreased for the 1988 sample and increased for the 1991 sample. In light of this conflicting evidence, the conclusion is that the depth does not change after the split. Thus if liquidity is measured by spread and depth, liquidity decreases after the split.

In addition to examining the behavior of the depth around stock splits, I also explore the depth in relation to trading volumes and the behavior of trading volumes around splits. I show that for the stocks in my sample the daily number of transactions increases after the split, but total daily dollar volume and split-adjusted share volume remain unchanged. The dollar volume and split-adjusted share volume of a typical transaction decline on average by about 20% and 25%, respectively. Both the mean and the median are used to define a "typical" transaction and the conclusions are remarkably similar for both measures. Transaction sizes for individual trades have not been examined in earlier papers.

As far as the relation between volumes and depths is concerned, I show that the fraction of trades with volumes not exceeding the quoted depth increases after the split for transactions classified as buys, whereas it remains unchanged for transactions classified as sells. Thus from this perspective it appears that liquidity actually increased after the split.

Finally, I consider the intraday properties of the bid-ask spread, an illustration of the liquidity characteristics of splits that is not purely descriptive. By using the ordered-probit model for the discrete-valued bid-ask spread, it is shown that stock splits decrease liquidity in the sense that large trades tend to widen the percentage spread more after the split than before the split.

There are many other issues related to stock splits that are not touched on in this chapter. For example, another empirical regularity that has been documented regarding stock splits is that stock price volatility increases subsequent to stock splits. The relation between splits and stock return behavior or firm performance is not discussed here either. Another limitation is that this study does not deal with stock dividends nor reverse splits.

The chapter is organized as follows. Section 5.2 presents a brief review of the existing literature on splits and liquidity. The data used in this study are briefly discussed in section 5.3. In sections 5.4 and 5.5, I present a pre-split versus post-split comparison of percentage bid-ask spreads and quoted depths—two distinct, but interrelated liquidity measures. An analysis of various measures of trading activity around splits is given in section 5.6. An intraday analysis of spreads using the ordered probit model is given in section 5.7 and section 5.8 concludes. There are two appendixes to this chapter with information on the peculiarities of closing quotes (appendix C) and on the issue of missing data for the 1991 sample (appendix D).

5.2 Literature Review

In this section I present a brief review of the literature regarding stock splits and liquidity. Harris (1990a) discusses spreads and depths as separate dimensions of market liquidity. See also Black (1971), Bernstein (1987), and Grossman and Miller (1988) for some general discussion of liquidity and Lee, Mucklow, and Ready (1993) for an empirical study that does consider both spreads and depths in the context of the impact of earnings announcements.

Although this study does not focus on the relation between splits and stock return behavior or firm performance, any discussion on stock splits that does not mention the classic article by Fama, Fisher, Jensen, and Roll (1969) would seem incomplete.

In a survey of chief financial officers, Baker and Gallagher (1980) find that one major reason managers gave was that lowering the stock price keeps it in a range attractive to small investors, as well as broadening the ownership base. In an early study of the liquidity effects of stock splits, Copeland (1979) finds that trading volume increases less than proportionately after the split and that post-split bid-ask spreads increase significantly as a percentage of the value of the stock. In other words, the surprising conclusion of his study is that liquidity actually *decreases* following the split. By considering both the short- and long-term liquidity effects of stock splits, Murray (1985) finds no statistically significant change in proportional trading volume. In addition, Murray finds that stock splits do not have an effect on the percentage bid-ask spread.

The thorough study by Lakonishok and Lev (1987) does not support a decrease in marketability as far as the volume of trade—measured by turnover—is concerned. They document that splitting firms' stocks experience large trading volumes during the year before the announcement and that this can be attributed to the unusual operational performance of these firms. However, they add the following: "Yet it might be that splits affect other aspects of marketability, such as the composition of stockholders, the number of stockholders, or transaction costs. The effect of splits on marketability is, therefore, still an open issue."

Lamoureux and Poon (1987) confirm Copeland's finding that the *value* of shares traded falls subsequent to the ex-split day. They attempt to explain the announcement effect with a theory based on an increase of *raw* volume, which is used as a proxy for the number of transactions.

In a recent study of the effects of stock splits on the bid-ask spread, Conroy, Harris, and Benet (1990) find that the absolute (dollar) bid-ask spread decreases after the split and that the percentage spread increases for a sample of 147 splits of NYSE-listed stocks during the period from January 1981 to April 1983. Unlike Copeland (1979) and Murray (1985), who use representative bid-ask quotes, Conroy et al. use closing bid-ask spreads—a total of 20 spread observations for each security. Closing spreads are derived from closing quotes, specially-marked quotes that are disseminated to signal the end of the trading day for a given stock.

See Brennan and Copeland (1988) for a model of the signalling hypothesis of stock splits. For a discussion of stock splits and stock price volatility, see Ohlson and Penman

(1985) and the papers by Dravid (1987), Dubofsky (1991), and Lamoureux and Poon (1987) for extensions.

5.3 Data

The data for this study were obtained from the ISSM transactions database for the years 1988 and 1991. Using information provided by *NYSE Fact Books* for these two years, those companies were selected that split their stocks with a split factor of at least 2-for-1. Stocks that do not satisfy the requirement of having at least 5 pre-split trading days and 5 post-split trading days in their respective calendar years are excluded from the analysis. See Tables 5.1 and 5.2 for a list of the companies that are included for the years 1988 and 1991, respectively. There are 26 companies in the 1988 sample, whereas 46 companies satisfy the criteria for inclusion in the 1991 sample. The split factors for all companies are also included in Tables 5.1 and 5.2.

5.4 Splits and Bid-Ask Spreads

Previous studies have examined the relation between stock splits and bid-ask spreads. Basically, the question is whether percentage spreads increase subsequent to stock splits. I claim to have an innovation by considering the complete record of all spreads, not just closing ones. In Appendix C, I demonstrate that closing quotes do not accurately reflect the quotes prevailing at the end of the trading day. In particular, spreads calculated from closing quotes tend to be narrower than spreads prevailing at the end of the trading session. For example, the final closing spread is on average 8.9% smaller than the last intraday spread for the 1988 sample, whereas the corresponding decline for the 1991 sample is 4.8%.

Using the complete record of all quotes, I use the following procedure to get typical spreads for a given stock. For each trading day, I construct a second-by-second time series of its percentage bid-ask spread. The average of this series is calculated to get an observation of a representative percentage spread on this day for this particular stock. This procedure is repeated for the 100-day period centered around the split.

Index	Ticker	Name	Split Factor
1	ASH	ASHLAND OIL INC	2 for 1
2	BCR	BARD CR INC	2 for 1
3	BMS	BEMIS INC	2 for 1
4	CCK	CROWN CORK & SEAL INC	3 for 1
5	CTB	COOPER TIRE & RUBR CO	2 for 1
6	DCI	DONALDSON INC	2 for 1
7	DOV	DOVER CORP	2 for 1
8	F	FORD MTR CO DEL	2 for 1
9	FBO	FEDERAL PAPER BRD INC	2 for 1
10	FCB	FOOTE CONE & BELDING	2 for 1
11	GW	GULF & WESTN INC	2 for 1
12	HLT	HILTON HOTELS CORP	2 for 1
13	IMD	IMO DELAVAL INC	2 for 1
14	MRK	MERCK & CO INC	3 for 1
15	NHY	NORSK HYDRO A S	2 for 1
16	NWL	NEWELL CO	2 for 1
17	PC	PENN CENT CORP	2 for 1
18	SBL	SYMBOL TECHNOLOGIES INC	2 for 1
19	SC	SHELL TRANS & TRADING LTD	2 for 1
20	SDW	SOUTHDOWN INC	2 for 1
21	SPP	SCOTT PAPER CO	2 for 1
22	STH	STANHOME INC	2 for 1
23	TKR	TIMKEN CO	2 for 1
24	UB	UNITED BRANDS CO	3 for 1
25	VDC	VAN DORN CO	3 for 1
26	WWY	WRIGLEY WM JR CO	2 for 1

Table 5.1: NYSE Stocks Splitting At Least 2-for-1 (1988)

Index	Ticker	Name	Split Factor
1	ASC	AMERICAN STORES CO NEW	2 for 1
2	AUD	AUTOMATIC DATA PROCESSING	2 for 1
3	BOL	BAUSCH LOMB INC	2 for 1
4	BV	BLOCKBUSTER ENTERTAINMENT CORP	2 for 1
5	CIR	CIRCUS CIRCUS ENTERPRISES INC	2 for 1
6	CL	COLGATE-PALMOLIVE CO	2 for 1
7	CML	CML GROUP INC	2 for 1
8	CNC	CONSECO INC	2 for 1
9	CPB	CAMPBELL SOUP CO CAPITAL	2 for 1
10	GLX	GLAXO HOLDINGS P.L.C. ADR 2ORD	2 for 1
11	GPS	THE GAP INC	2 for 1
12	GPU	GENERAL PUB UTILITIES CORP	2 for 1
13	GS	GILLETTE CO	2 for 1
14	GWW	GRAINGER W.W. INC	2 for 1
15	HF	HOUSE OF FABRICS INC	2 for 1
16	HKF	HANCOCK FABRIC INC	2 for 1
17	HRB	BLOCK H R INC	2 for 1
18	IGT	INTL GAME TECHNOLOGY	2 for 1
19	IMA	IMCERA GROUP INC	3 for 1
20	JEC	JACOBS ENGINEERING GROUP INC	2 for 1
21	K	KELLOGG CO	2 for 1
22	LC	LIBERTY CORP	2 for 1
23	LDL	LYDALL INC	2 for 1
24	LGN	LOGICON INC	2 for 1
25	LOC	LOCTITE CORPORATION	2 for 1
26	MDT	MEDTRONIC INC	2 for 1
27	MS	MORGAN STANLEY GROUP INC	2 for 1
28	NLC	NALCO CHEMICAL CO	2 for 1
29	NME	NATL MED ENTERPRISES INC	2 for 1
30	PFE	PFIZER INC	2 for 1
31	PVH	PHILLIPS-VAN HEUSEN CORP	2 for 1
32	RAD	RITE AID CORP	2 for 1
33	RAL	RALSTON PURINA CO	2 for 1
34	RBD	RUBBERMAID INC	2 for 1
35	RPR	RHONE POULENC RORER INC	2 for 1
36	SCR	SEA CONTAINER LTD UNPAIRED	2 for 1
37	SHW	SHERWIN-WILLIAMS CO	2 for 1
38	SRR	STRIDE-RITE CORP	2 for 1
39	STR	QUESTAR CORP	2 for 1
40	SYN	SYNTEX CORPORATION	2 for 1
41	UNP	UNION PACIFIC CORPORATION	2 for 1
42	USS	UNITED STATES SURGICAL CORP	2 for 1
43	VCD	VALUE CITY DEPT STORES INC	2 for 1
44	VNO	VORNADO INCORPORATED	5 for 1
45	WAG	WALGREEN COMPANY	2 for 1
46	WTI	WHEELABRATOR TECHNOLOGIES NEW	2 for 1

Table 5.2: NYSE Stocks Splitting At Least 2-for-1 (1991)

See Figure 5.1 for the resulting series for Merck, which split 3-for-1 in 1988. The average percentage spread is shown versus the number of days relative to the split date. Days −50 to −1 refer to the pre-split period and days 0 to 49 refer to the post-split period. As can be seen from this plot, Merck's average percentage spread increased after the split. The average percentage spread for the 50-day pre-split period is 0.208% and the average percentage spread for the 50-day post-split period is 0.385%, an increase of about 85%. A standard two-sample t test of the difference in mean confirms that the shift in mean is highly statistically significant. Assuming normal samples and equal variances for the pre-split and post-split periods, the resulting test statistic has a value of 27.33, much larger than any reasonable cutoff value for a t distribution with 98 degrees of freedom (or a standard normal distribution).[3]

I carried out the same procedure for all stocks in both of the two sample periods. The results are summarized in Tables 5.3 and 5.4. Comparing the average pre-split percentage spreads to the average post-split percentage spreads, the results indicate that spreads increased substantially for almost all stocks if we consider the 100 trading days around the stock split. For all 26 companies comprising the 1988 sample, the percentage bid-ask spread increased after the split (see Table 5.3). The equally-weighted average of these 26 stocks increases from 0.67% before the split to 1.06% after the split, an increase of about 60%. For the 1991 sample we see from Table 5.4 that the average bid-ask spread increases for 45 stocks and decreases for 1 stock, VNO.[4] Overall, the equally-weighted average spread of the 46 stocks in the 1991 sample increases from 0.56% to 0.82%, an increase of 47.5%. Thus these results confirm that if the bid-ask spread is taken as the measure of market liquidity, stock splits are associated with a decrease of liquidity.

There is some indication that a higher split ratio is associated with a greater decrease in liquidity for the 1988 sample of 26 stocks.[5] From Table 5.3, the average increase of the

[3]Although the sample standard deviation increases from 0.0286% for the pre-split period to 0.0357% for the post-split period, this increase is not statistically significant at the 5 percent level.

[4]Note that VNO is the only company that split 5-for-1. It appears that VNO is an extremely thinly traded stock with on average 4 transactions per day (see Table 5.19 and also Appendix D).

[5]Since virtually all split factors for the 1991 sample are equal to 2 (see Table 5.2), this sample is unsuitable for an examination of the relation between split factors and changes in liquidity.

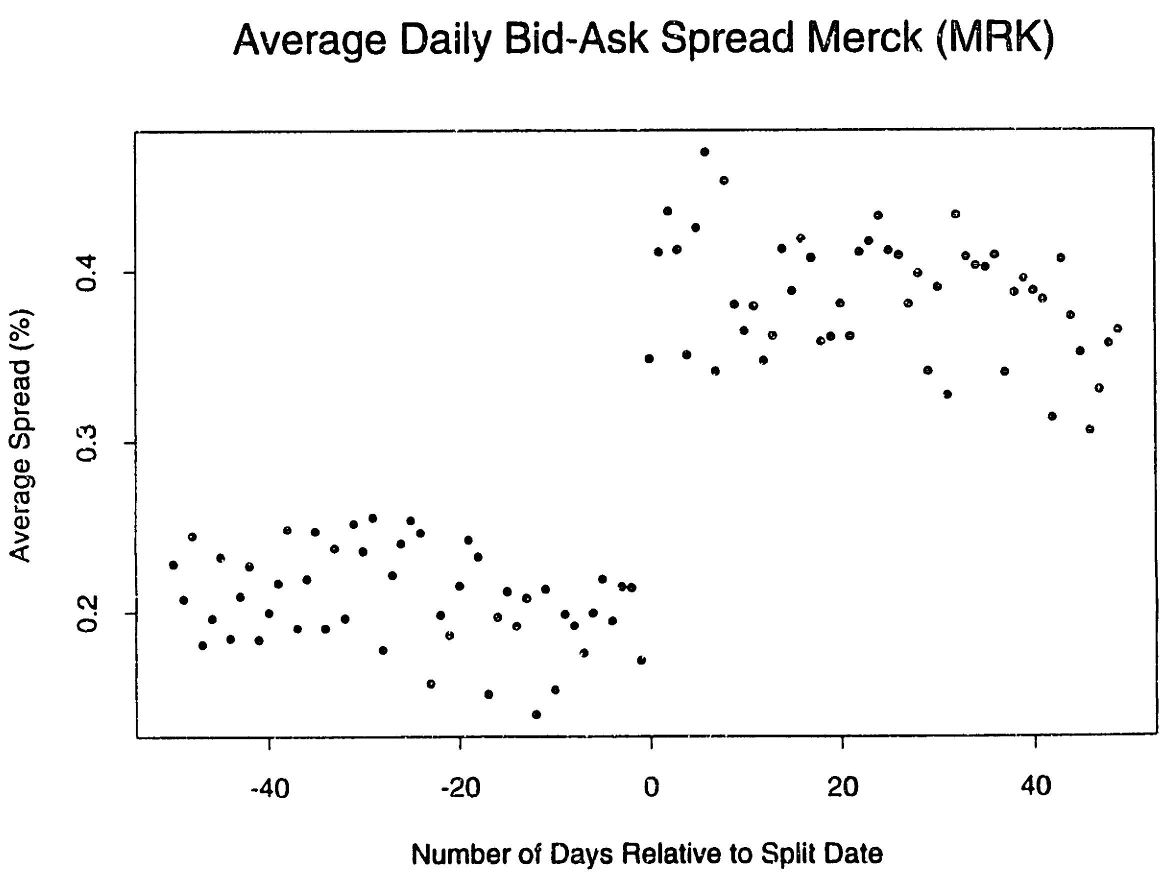


Figure 5.1: Average Daily Bid-Ask Spread Merck (MRK) Around 1988 Split

Ticker	Average Spread Before Split (%)	Average Spread After Split (%)	N_1	N_2	ζ
ASH	0.54	1.01	50	50	19.80
BCR	0.62	1.01	50	50	14.59
BMS	0.88	1.30	50	50	9.29
CCK	0.44	0.71	50	50	14.44
CTB	0.86	1.18	50	50	7.02
DCI	0.89	1.63	50	50	14.53
DOV	0.64	1.09	50	50	13.72
F	0.29	0.41	7	50	6.65
FBO	0.79	1.36	50	50	15.62
FCB	1.27	1.43	50	50	2.25
GW	0.31	0.48	50	50	13.16
HLT	0.37	0.63	50	50	17.31
IMD	0.87	1.29	50	50	9.84
MRK	0.21	0.38	50	50	27.33
NHY	0.67	1.09	50	50	11.27
NWL	0.73	1.11	50	50	9.20
PC	0.71	1.14	49	50	11.50
SBL	0.75	1.01	50	50	8.43
SC	0.37	0.76	50	14	15.02
SDW	0.93	1.36	50	50	9.47
SPP	0.56	0.76	50	50	9.62
STH	0.91	1.42	50	50	12.10
TKR	0.52	0.92	50	50	14.78
UB	0.78	1.58	50	50	12.32
VDC	0.88	1.71	50	50	14.28
WWY	0.52	0.86	50	50	14.35
Average	0.67	1.06			

Table 5.3: Percentage Spread Comparison for Individual Stocks (1988)

The average percentage spread statistics are calculated for the 50 trading days prior to the split (1st column) and for the 50 trading days subsequent to the split (2nd column). For each stock, N_1 is the number of trading days for which data are available during the 50-day pre-split period and N_2 is the number of trading days for which data are available during the 50-day post-split period. The following procedure is used to calculate the average percentage spread for a given stock. For each of the N_i ($i = 1, 2$) trading days for which data are available, a second-by-second time series of the stock's percentage bid-ask spread is calculated. The average of this series is taken as the typical spread for that day and the average over the N_i days is given in the table.

The column labeled ζ gives the value of the test statistic whether the mean of the N_1 pre-split spreads is equal to the mean of the N_2 post-split spreads. Define $X_1, \ldots, X_{N_1}$ as the daily pre-split percentage spreads and $Y_1, \ldots, Y_{N_2}$ as the daily post-split percentage spreads. If the respective mean estimators are denoted by $\overline{X}$ and $\overline{Y}$ and the respective variance estimators are denoted by S_x^2 and S_y^2, the test statistic ζ is defined by $\zeta = (\overline{y} - \overline{x})/[s_p\sqrt{(1/N_1) + (1/N_2)}]$, where S_p^2 is the pooled variance estimator $S_p^2 \equiv [(N_1 - 1)S_x^2 + (N_2 - 1)S_y^2]/(N_1 + N_2 - 2)$ and lower case letters are used to differentiate between estimators and estimates. If the X's and Y's are independent and normally distributed, the test statistic has a t distribution with $N_1 + N_2 - 2$ degrees of freedom (approximately standard normal when N_1 and N_2 are large).

Ticker	Average Spread Before Split (%)	Average Spread After Split (%)	N_1	N_2	ζ
ASC	0.43	0.68	50	50	12.47
AUD	0.50	0.70	50	50	9.69
BOL	0.38	0.67	50	50	12.97
BV	0.66	1.24	47	50	29.91
CIR	0.43	0.67	50	50	13.76
CL	0.28	0.47	50	50	23.61
CML	0.86	1.36	50	50	12.98
CNC	0.60	0.75	50	50	7.71
CPB	0.31	0.52	50	5	10.25
GLX	0.32	0.50	50	34	29.17
GPS	0.41	0.51	50	50	8.86
GPU	0.43	0.72	50	50	13.47
GS	0.32	0.50	50	50	21.93
GWW	0.36	0.54	50	50	13.47
HF	0.64	1.00	50	50	10.50
HKF	0.70	1.05	50	50	9.84
HRB	0.44	0.68	50	50	13.31
IGT	0.56	0.89	50	50	13.42
IMA	0.33	0.65	50	33	20.12
JEC	1.00	1.28	50	50	6.44
K	0.31	0.42	50	10	5.34
LC	0.89	1.35	50	20	7.23
LDL	0.96	1.54	50	50	8.54
LGN	0.97	1.57	50	19	9.69
LOC	0.50	0.76	50	50	13.63
MDT	0.29	0.37	50	50	7.06
MS	0.39	0.52	50	50	8.07
NLC	0.52	0.85	50	50	12.96
NME	0.48	0.93	50	50	20.67
PFE	0.26	0.35	50	50	11.42
PVH	0.87	1.63	50	50	13.70
RAD	0.51	0.79	50	50	14.31
RAL	0.28	0.40	46	50	12.83
RBD	0.38	0.56	50	21	10.10
RPR	0.33	0.49	50	50	9.46
SCR	1.03	1.04	50	50	0.16
SHW	0.64	0.92	50	50	13.07
SRR	0.58	0.90	50	8	9.26
STR	0.73	1.25	50	50	15.02
SYN	0.31	0.49	50	50	20.81
UNP	0.29	0.44	50	50	19.52
USS	0.49	0.51	50	50	0.77
VCD	0.97	1.26	50	20	5.09
VNO	1.46	1.28	50	50	−2.11
WAG	0.43	0.64	22	50	16.40
WTI	0.74	1.10	50	50	13.38
Average	0.56	0.82			

Table 5.4: Percentage Spread Comparison for Individual Stocks (1991)

percentage bid-ask spread for those stocks that split by 2-for-1 (22 stocks, see Table 5.1) is 58.3% and the average increase of the percentage bid-ask spread for those stocks that split by 3-for-1 (4 stocks, see Table 5.1) is 86.0%. Define y_i as the change in percentage bid-ask spread following the split, f_i as the split ratio, and D_i as an indicator variable with $D_i = 0$ if $f_i = 2$ and $D_i = 1$ if $f_i = 3$. For the sample of 26 stocks, the OLS regression equation is given by $y_i = 0.583 + 0.277D_i$. The t-statistic of the parameter estimate for the split-factor indicator is 2.49, which is significant. Thus it appears that there is a positive relation between split ratio and decrease in liquidity for the stocks included in the 1988 sample.[6]

One comment about the way Table 5.4 is constructed is in order. For each stock, the 100-day period around the split is taken as the 50 trading days before and after the split for which data are available in the ISSM database. As explained in Appendix D, there are several trading days in 1991 for which no data are available for many of the stocks. These days are excluded so that the summary statistics are calculated for the first 50 days before and after the split for which data are available.

Tables 5.3 and 5.4 also report statistics for a test whether the differences between pre-split spreads and post-split spreads for individual stocks are significant. The statistic ζ is defined as the ratio of the difference between post-split and pre-split average spread and its standard deviation. If the two means are equal, the quantity ζ has an asymptotic standard normal distribution under the appropriate assumptions (see Table 5.3 for details). Inspection of Tables 5.3 and 5.4 reveals that the change in spread is highly statistically significant for almost all of the stocks.[7]

Rather than considering each stock's spread individually, we can also combine them in a portfolio and consider the equally-weighted average spread before and after the split. By aligning the various stocks by their split date, I calculate the cross-sectional average spread for the K trading days preceding the split and also for the K trading days following

[6] An alternative illustration of the same phenomenon is the observation that the correlation between the change in percentage bid-ask spread and the split factor is about 45%.

[7] One of the assumptions underlying the two-sample t test is that the variances of the two samples are equal. A standard F test for the equality of these variances rejects this equality for many of the stocks. Therefore, the statistical results reported in Tables 5.3 and 5.4 should be interpreted cautiously.

the split. To make the comparisons valid, a stock is only included in the portfolio if its split date is at least K days away from the beginning and from the end of the calendar year so that there are complete data for all of the days in the $2K$-day period. Table 5.5 reports results for the 1988 sample. Complete results are given in panel A for the 14-day period around the split ($K = 7$) for which data are available for all of the 26 stocks ($N = 26$). Summarized results are given in panel B for $K = 14$ and in panel C for $K = 49$. In panel B, the portfolio consists of 25 stocks—F is excluded since it has only 7 pre-split trading days ($N_1 = 7$ in Table 5.3). In panel C, the portfolio consists of 24 stocks—additionally, SC is excluded, since it has only 14 post-split trading days ($N_2 = 14$ in Table 5.3). See Figure 5.2 for the complete results for all three portfolios. It is clear from this graph of cross-sectional means that the percentage spreads increase subsequent to the split for the 1988 sample.

From Table 5.5, a two-sample t test reveals that the differences in average percentage spread between the pre-split period and the post-split period are all highly statistically significant. However, there is some indication that the variance of the observations in the post-split period is larger than the variance of the observations in the pre-split period (the variance ratio exceeds unity in all three cases). In the last panel of Table 5.5, the p-value of the F test for the equality of variances is equal to 0.0169, which is significant at the 5 percent level but not at the 1 percent level. Figure 5.3 shows the cross-sectional standard deviation of the percentage spreads for each stock for each day of the $2K$-day period around the split. There appears to be an increased variability of relative spreads following the split for the 1988 sample.

The results for the 1991 sample are reported in Table 5.6. A portfolio of the complete set of stocks ($N = 46$) can only be constructed for $K = 5$, i.e., for the 10-day period around the split (see Panel A). For longer time periods, some stocks that have split dates close to the beginning or end of the year have to be dropped to keep the sample constant for all days of the period under consideration. Thus there are $N = 42$ stocks for the 40-day period around the split ($K = 20$, see Panel B) and there are $N = 34$ stocks in

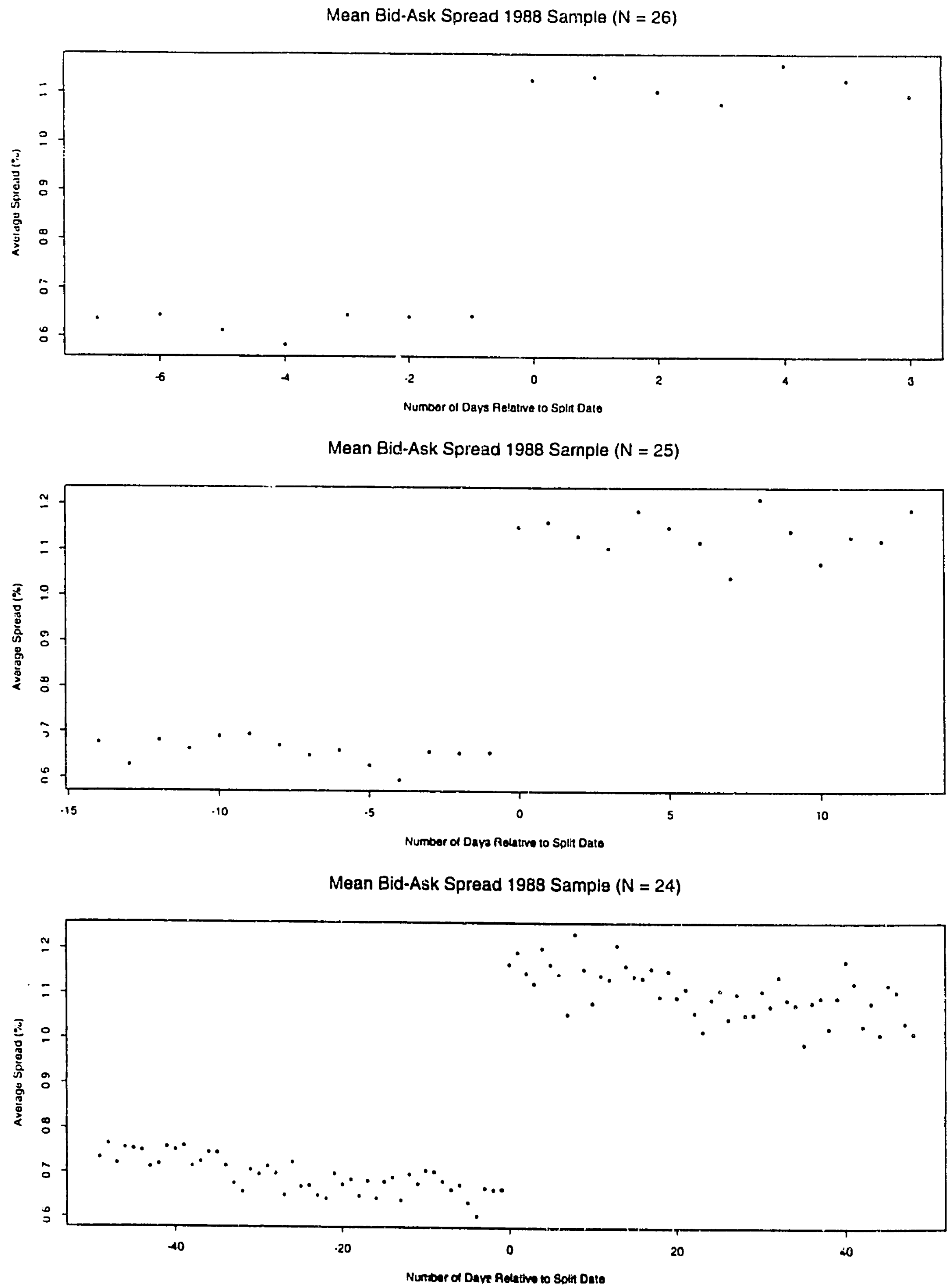


Figure 5.2: Average Percentage Bid-Ask Spread for Portfolio (1988)

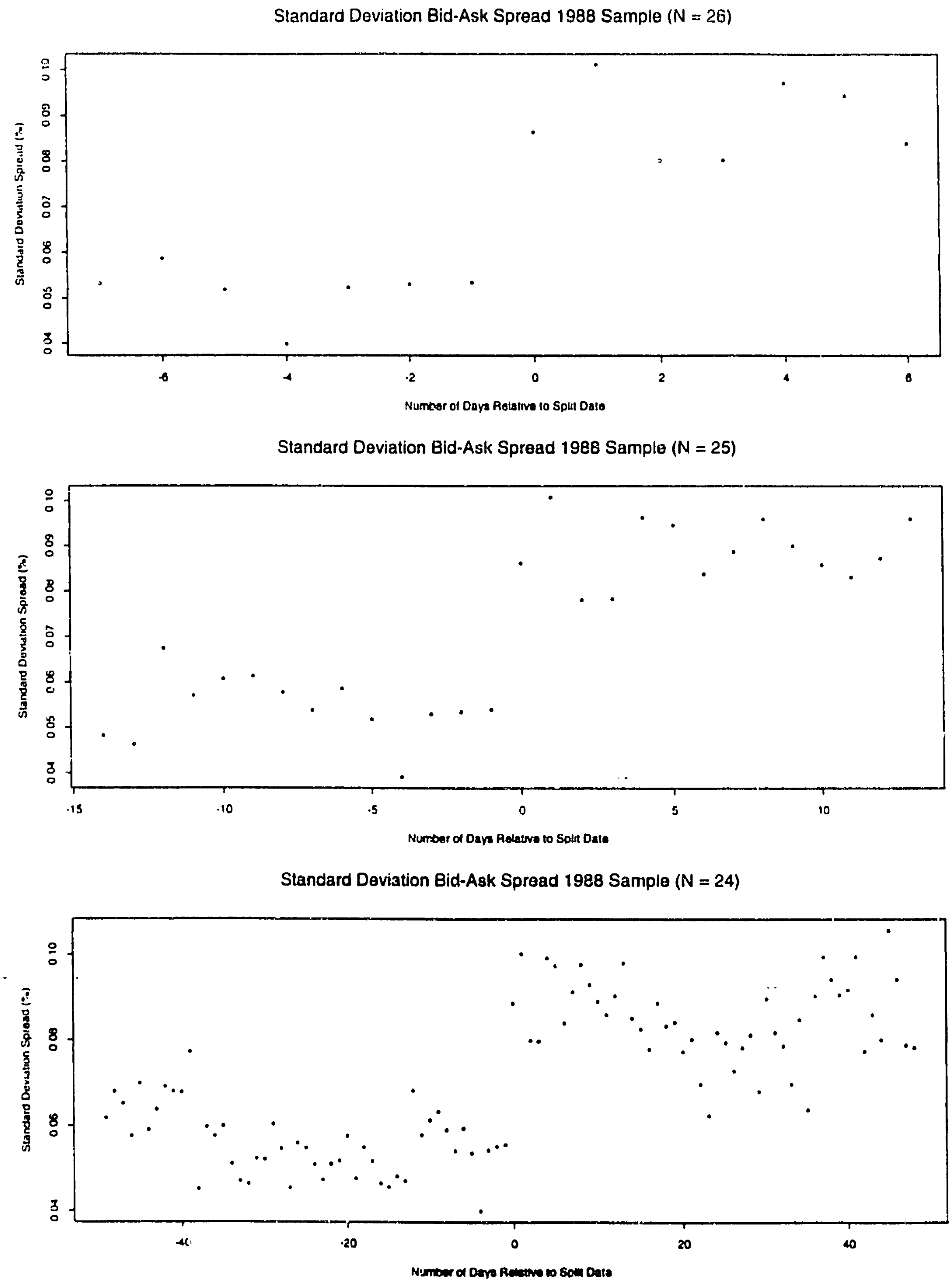


Figure 5.3: Standard Deviation of Percentage Bid-Ask Spread for Portfolio (1988)

A. $K = 7$, $N = 26$ (full sample)

Days	Before Split	After Split
1	0.6419	1.1252
2	0.6405	1.1317
3	0.6440	1.1023
4	0.5819	1.0767
5	0.6122	1.1563
6	0.6428	1.1251
7	0.6354	1.0945
Mean	0.6284	1.1160
St. dev.	0.0233	0.0266

The ζ-statistic for equality of means is 36.52, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 1.30, which is insignificant at the 5 percent level (2.5% quantile is 0.17 and 97.5% quantile is 5.82 for F distribution with 6 and 6 degrees of freedom).

B. $K = 14$, $N = 25$ (1 stock excluded)

	Before Split	After Split
Mean	0.6566	1.1362
St. dev.	0.0267	0.0455

The ζ-statistic for equality of means is 33.99, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 2.91, which is insignificant at the 5 percent level (2.5% quantile is 0.32 and 97.5% quantile is 3.12 for F distribution with 13 and 13 degrees of freedom).

C. $K = 49$, $N = 24$ (2 stocks excluded)

	Before Split	After Split
Mean	0.6939	1.1044
St. dev.	0.0393	0.0558

The ζ-statistic for equality of means is 42.12, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 2.01, which is significant at the 5 percent level with a p-value of 0.0169 (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

Table 5.5: Percentage Spread Comparison for Portfolios (1988)

N is the number of stocks in an equally-weighted portfolio whose percentage spread is calculated for the K-day pre-split period and also for the K-day post-split period. For this purpose, all stocks are aligned by their respective split dates.

A. $K = 5$, $N = 46$ (full sample)

Days	Before Split	After Split
1	0.5012	0.8386
2	0.5183	0.7786
3	0.4992	0.7962
4	0.5145	0.8338
5	0.5440	0.7977
Mean	0.5154	0.8090
St. dev.	0.0180	0.0260

The ζ-statistic for equality of means is 20.76, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 2.10, which is insignificant at the 5 percent level (2.5% quantile is 0.10 and 97.5% quantile is 9.60 for F distribution with 4 and 4 degrees of freedom).

B. $K = 20$, $N = 42$ (4 stocks excluded)

	Before Split	After Split
Mean	0.5324	0.8212
St. dev.	0.0234	0.0275

The ζ-statistic for equality of means is 35.77, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 1.38, which is insignificant at the 5 percent level (2.5% quantile is 0.40 and 97.5% quantile is 2.53 for F distribution with 19 and 19 degrees of freedom).

C. $K = 50$, $N = 34$ (12 stocks excluded)

	Before Split	After Split
Mean	0.5635	0.8157
St. dev.	0.0321	0.0317

The ζ-statistic for equality of means is 39.55, which is highly statistically significant. The ratio of post-split sample variance to pre-split sample variance is 0.97, which is insignificant at the 5 percent level (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

Table 5.6: Percentage Spread Comparison for Portfolios (1991)

N is the number of stocks in an equally-weighted portfolio whose percentage spread is calculated for the K-day pre-split period and also for the K-day post-split period. For this purpose, all stocks are aligned by their respective split dates.

the full 100-day period around the split ($K = 50$, see Panel C).[8] The main conclusion is the same as for the 1988 sample: percentage spreads increased substantially after the split. Note from Table 5.6 that the variance in the pre-split and after-split periods did not seem to change: the ratios of post-split sample variance to pre-split sample variance are all insignificantly different from 1. See Figures 5.4 and 5.5 for a graphical presentation of the cross-sectional means and standard deviations of the daily percentage spreads for the three 1991 portfolios.

5.5 Splits and Depths

In this section I explore the behavior of another liquidity measure, the *depth*, around stock splits. Recall that a complete specialist quote on the NYSE has a price and a volume dimension on both sides of the market: bid and ask prices and bid and ask depths, the number of shares available on each side of the market. Greater liquidity is associated with a "deeper" market, a market with more depth.

For the same reason that I considered the relative (percentage) spread as opposed to the absolute (dollar) spread in the previous section, I consider the depth in units of *dollars* rather than in units of *shares* for analyzing how stock splits affect the depth: the change in stock price associated with the split would make a pre-split versus post-split comparison meaningless if depth were measured in units of shares. For example, it is hard to argue that the depth remained unchanged if after a 2-for-1 stock split the number of shares available at the bid or at the ask remained the same.

I therefore focus on *dollar bid depth* and *dollar ask depth*, defined as the product of the bid price and the number of shares available at the bid price, and the product of the ask price and the number of shares available at the ask price, respectively. I also define the *total dollar depth* as the sum of the dollar bid depth and the dollar ask depth. For brevity, I will henceforth simply use "depth" for "dollar depth." The next two subsections

[8]The stocks excluded from the portfolio of Panel B are those stocks with $N_1 < 20$ or $N_2 < 20$ in Table 5.4 (CPB, K, LGN, and SRR). The additional stocks excluded from the portfolio of Panel C are those stocks with $20 \leq N_1 < 40$ or $20 \leq N_2 < 40$ in Table 5.4 (BV, GLX, IMA, LC, RAL, RBD, VCD, and WAG).

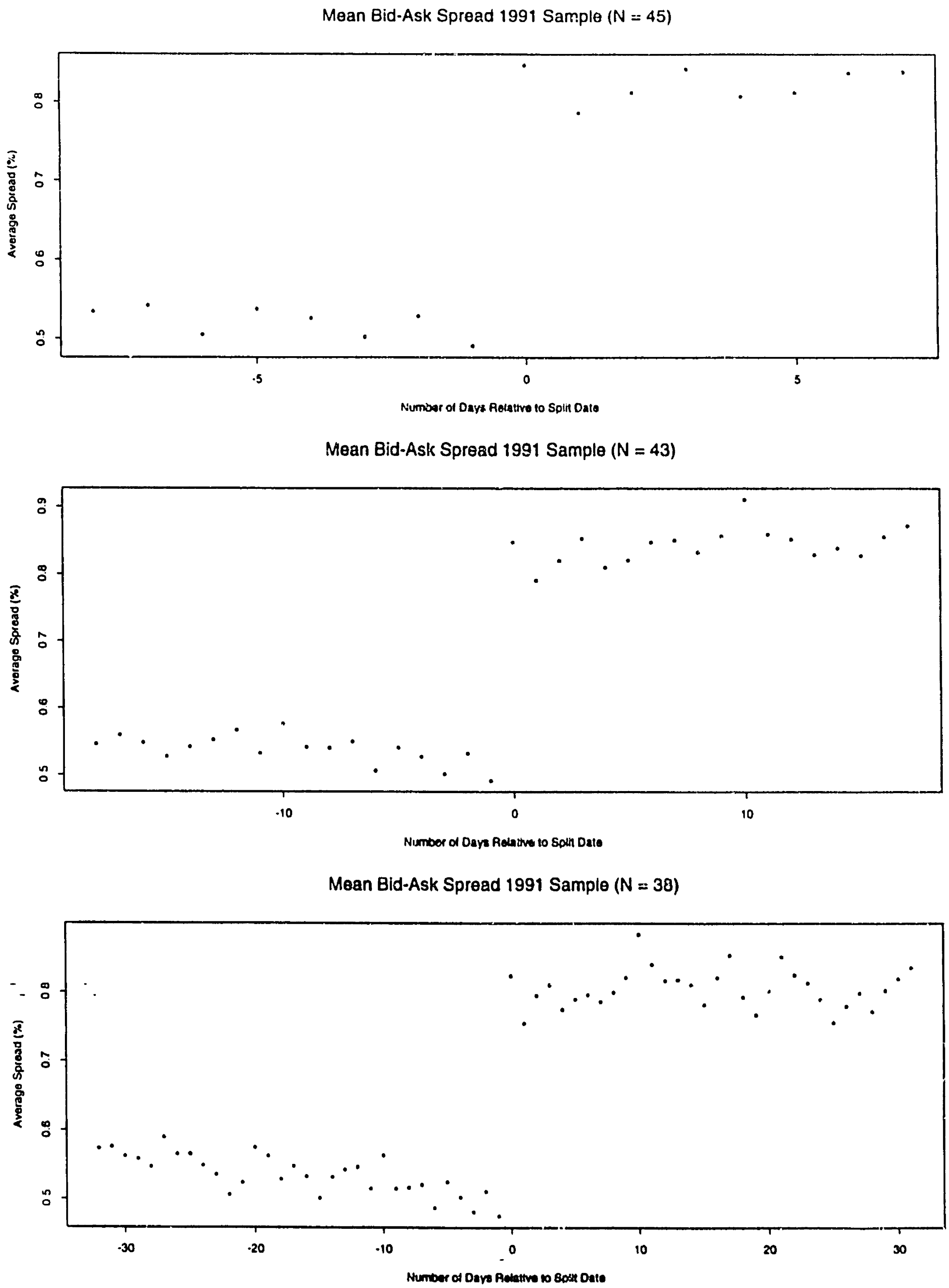


Figure 5.4: Average Percentage Bid-Ask Spread for Portfolio (1991)

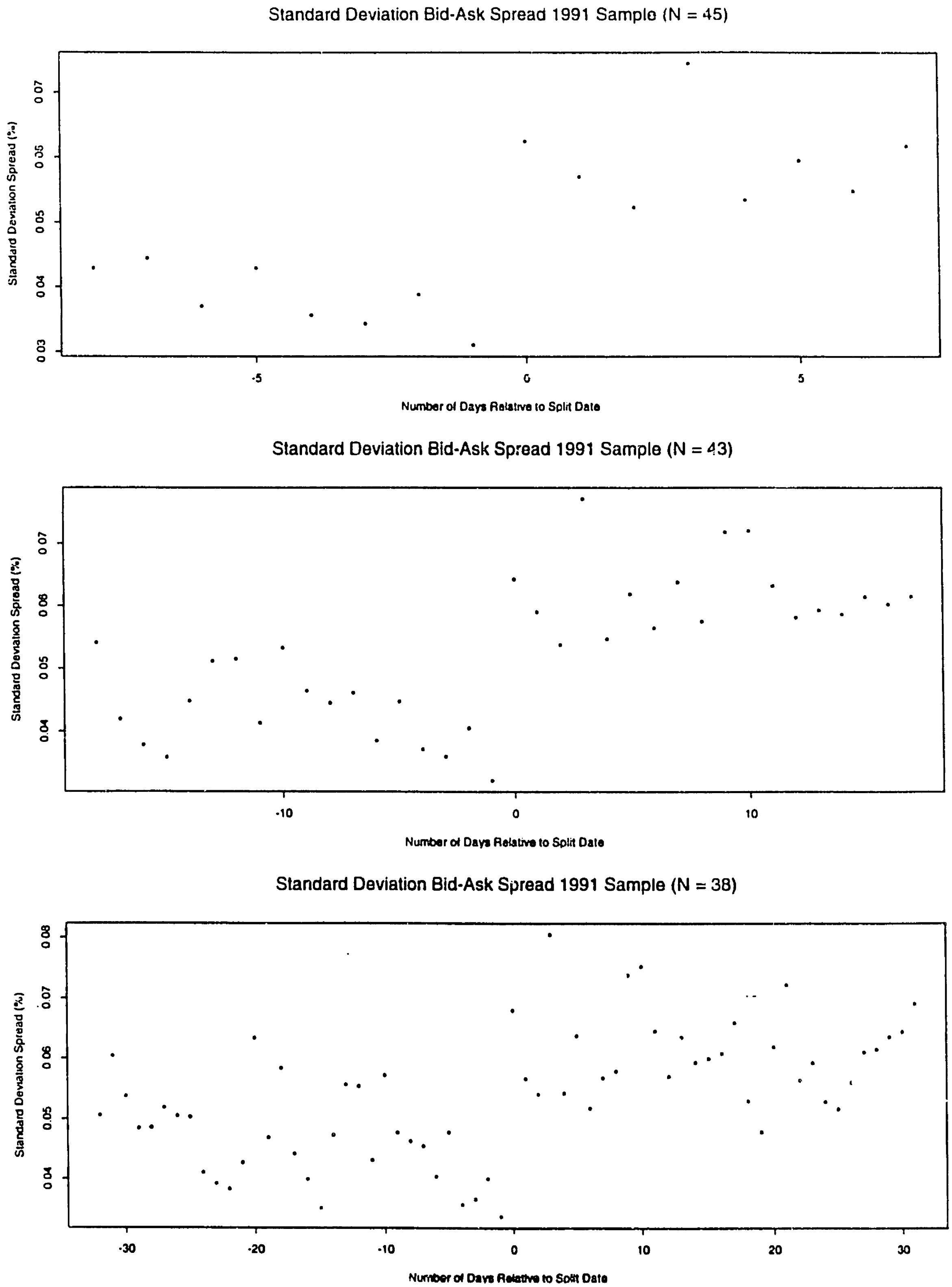


Figure 5.5: Standard Deviation of Percentage Bid-Ask Spread for Portfolio (1991)

investigate the changes in depth associated with stock splits for individual stocks and for portfolios.

5.5.1 Splits and Depths for Individual Stocks

Using the same procedure as in the previous section, I first calculate the average (dollar) depth for each stock on each day of the 100-day period around the split. Stock-by-stock averages are then calculated for the pre-split period and for the post-split period. For the 1988 sample, the results are summarized in Tables 5.7–5.9. The results for this sample period are decidedly mixed and no clear pattern emerges. The average bid depth increases for 15 stocks and decreases for 11 stocks (Table 5.7), the average ask depth increases for 13 stocks and decreases for 13 stocks (Table 5.8), and the total depth increases for 14 stocks and decreases for 12 stocks (Table 5.9). Moreover, the magnitude of changes is often relatively small and statistically insignificant. As before, ζ is the value of the test statistic whether the mean of the N_1 pre-split depths is equal to the mean of the N_2 post-split depths in Tables 5.7–5.9. As an aside, the reason that the values for N_1 and N_2 are not equal to 50 for all stocks is that for some stocks there are fewer than 50 trading days prior to the split (so that $N_1 < 50$) or there are fewer than 50 trading days subsequent to the split (so that $N_2 < 50$).

One might question whether it is necessary to consider the bid depth and the ask depth separately. Why not focus on the total dollar depth instead? If the behavior of the bid depth mirrored the behavior of the ask depth around the split exactly, we would have a correlation between the percentage bid depth change and the percentage ask depth change of 1. It turns out that this correlation is in fact 0.56, making it worthwhile to consider bid depth and ask depth separately.

The corresponding results for the 1991 sample are given in Tables 5.10–5.12. For the 1991 sample, the average bid depth increases for 23 stocks and decreases for 23 stocks (Table 5.10), the average ask depth increases for 24 stocks and decreases for 22 stocks (Table 5.11), and the total depth increases for 25 stocks and decreases for 21 stocks (Table 5.12). The correlation between the percentage bid depth change and the percentage

Ticker	Bid Depth Before Split ($)	Bid Depth After Split ($)	N_1	N_2	ζ
ASH	93,938	58,919	50	50	−4.11
BCR	90,260	85,663	50	50	−0.46
BMS	26,622	26,630	50	50	0.00
CCK	53,729	84,213	50	50	1.68
CTB	73,289	60,655	50	50	−1.37
DCI	39,479	28,507	50	50	−1.44
DOV	119,126	129,283	50	50	0.53
F	305,328	375,409	7	50	1.26
FBO	63,694	74,704	50	50	1.36
FCB	41,420	47,322	50	50	0.85
GW	251,437	189,354	50	50	−2.69
HLT	152,180	90,277	50	50	−4.38
IMD	49,716	47,472	50	50	−0.32
MRK	356,354	326,081	50	50	−1.18
NHY	289,561	380,787	50	50	2.10
NWL	74,694	62,471	50	50	−1.10
PC	188,645	127,902	49	50	−2.23
SBL	82,378	95,165	50	50	0.81
SC	582,586	278,963	50	14	−3.56
SDW	45,892	56,886	50	50	1.49
SPP	104,708	152,824	50	50	3.60
STH	42,085	54,779	50	50	1.42
TKR	41,042	56,171	50	50	2.16
UB	40,770	44,039	50	50	0.33
VDC	34,770	36,329	50	50	0.26
WWY	24,546	31,808	50	50	1.83

Table 5.7: Dollar Bid Depth Comparison for Individual Stocks (1988)

Ticker	Ask Depth Before Split ($)	Ask Depth After Split ($)	N_1	N_2	ζ
ASH	66,589	60,177	50	50	−0.82
BCR	100,797	108,218	50	50	0.66
BMS	42,204	27,611	50	50	−1.60
CCK	54,573	81,217	50	50	1.75
CTB	90,586	93,144	50	50	0.24
DCI	43,090	19,433	50	50	−3.34
DOV	88,873	124,378	50	50	2.71
F	331,171	476,177	7	50	1.68
FBO	84,781	81,796	50	50	−0.32
FCB	45,503	26,325	50	50	−2.82
GW	234,156	213,519	50	50	−0.63
HLT	206,959	101,370	50	50	−5.23
IMD	32,066	45,634	50	50	1.88
MRK	423,369	417,554	50	50	−0.18
NHY	296,760	326,799	50	50	0.84
NWL	85,486	86,193	50	50	0.06
PC	184,090	123,973	49	50	−2.10
SBL	80,590	143,340	50	50	4.25
SC	543,385	255,534	50	14	−4.34
SDW	50,528	73,214	50	50	2.10
SPP	123,862	154,881	50	50	2.52
STH	65,293	74,474	50	50	0.75
TKR	68,155	66,254	50	50	−0.22
UB	27,023	28,457	50	50	0.31
VDC	121,363	31,509	50	50	−4.48
WWY	49,510	44,979	50	50	−0.62

Table 5.8: Dollar Ask Depth Comparison for Individual Stocks (1988)

Ticker	Total Depth Before Split ($)	Total Depth After Split ($)	N_1	N_2	ζ
ASH	160,527	119,096	50	50	−2.93
BCR	191,057	193,881	50	50	0.15
BMS	68,826	54,241	50	50	−1.25
CCK	108,302	165,430	50	50	2.30
CTB	163,875	153,798	50	50	−0.62
DCI	82,569	47,941	50	50	−3.20
DOV	208,000	253,661	50	50	1.64
F	636,499	851,586	7	50	1.62
FBO	148,475	156,500	50	50	0.51
FCB	86,923	73,647	50	50	−1.31
GW	485,593	402,873	50	50	−1.81
HLT	359,139	191,647	50	50	−6.61
IMD	81,782	93,106	50	50	0.93
MRK	779,723	743,635	50	50	−0.70
NHY	586,321	707,585	50	50	1.79
NWL	160,181	148,665	50	50	−0.73
PC	372,735	251,875	49	50	−2.52
SBL	162,968	238,505	50	50	2.93
SC	1,125,972	534,496	50	14	−5.02
SDW	96,420	130,099	50	50	2.28
SPP	228,570	307,704	50	50	3.70
STH	107,378	129,253	50	50	1.29
TKR	109,197	122,426	50	50	1.21
UB	67,793	72,496	50	50	0.42
VDC	156,133	67,838	50	50	−4.03
WWY	74,056	76,787	50	50	0.29

Table 5.9: Total Dollar Depth Comparison for Individual Stocks (1988)

Ticker	Bid Depth Before Split ($)	Bid Depth After Split ($)	N_1	N_2	ζ
ASC	105,979	99,094	50	50	−0.57
AUD	127,380	128,505	50	50	0.10
BOL	165,883	155,566	50	50	−0.42
BV	261,237	623,527	47	50	11.02
CIR	181,790	133,790	50	50	−2.77
CL	210,840	245,088	50	50	1.79
CML	94,836	80,239	50	50	−1.23
CNC	64,724	65,253	50	50	0.07
CPB	177,970	213,405	50	5	0.89
GLX	1,002,805	691,286	50	34	−4.35
GPS	165,095	162,862	50	50	−0.19
GPU	189,524	123,891	50	50	−2.47
GS	254,329	283,741	50	50	0.93
GWW	121,639	141,893	50	50	1.22
HF	128,141	41,754	50	50	−3.52
HKF	62,803	43,419	50	50	−2.08
HRB	151,060	206,861	50	50	2.02
IGT	110,655	125,078	50	50	0.63
IMA	114,978	123,316	50	33	0.74
JEC	39,648	48,312	50	50	1.67
K	209,888	150,758	50	10	−1.93
LC	23,714	11,171	50	20	−2.23
LDL	25,231	18,328	50	50	−1.70
LGN	36,272	26,225	50	19	−1.82
LOC	64,524	80,417	50	50	1.35
MDT	95,693	150,676	50	50	2.03
MS	144,807	149,312	50	50	0.31
NLC	108,358	171,794	50	50	3.44
NME	271,007	482,540	50	50	5.22
PFE	308,966	335,857	50	50	1.07
PVH	69,318	33,698	50	50	−4.19
RAD	165,206	192,952	50	50	0.83
RAL	210,095	162,598	46	50	−2.94
RBD	109,496	83,178	50	21	−2.04
RPR	226,732	192,010	50	50	−1.57
SCR	48,910	49,951	50	50	0.10
SHW	116,579	148,159	50	50	2.16
SRR	73,795	73,073	50	8	−0.05
STR	76,609	57,374	50	50	−2.16
SYN	219,936	170,772	50	50	−3.77
UNP	240,399	269,768	50	50	1.20
USS	150,962	171,751	50	50	1.40
VCD	37,069	47,139	50	20	1.19
VNO	18,276	14,875	50	50	−1.43
WAG	142,488	117,392	22	50	−1.67
WTI	62,727	53,075	50	50	−1.69

Table 5.10: Dollar Bid Depth Comparison for Individual Stocks (1991)

Ticker	Ask Depth Before Split ($)	Ask Depth After Split ($)	N_1	N_2	ζ
ASC	153,934	165,424	50	50	0.61
AUD	164,853	193,814	50	50	1.40
BOL	250,262	210,790	50	50	−1.02
BV	445,685	705,338	47	50	5.24
CIR	290,671	218,917	50	50	−2.96
CL	318,252	304,474	50	50	−0.41
CML	92,837	92,592	50	50	−0.02
CNC	86,536	95,482	50	50	0.77
CPB	206,680	237,078	50	5	0.72
GLX	1,419,618	1,031,961	50	34	−4.26
GPS	264,522	247,776	50	50	−0.69
GPU	493,275	340,548	50	50	−2.79
GS	348,121	410,425	50	50	1.96
GWW	142,593	173,661	50	50	1.50
HF	132,138	76,517	50	50	−1.75
HKF	88,320	54,026	50	50	−3.78
HRB	182,811	265,483	50	50	3.24
IGT	153,271	148,085	50	50	−0.20
IMA	171,949	159,305	50	33	−0.93
JEC	62,191	70,872	50	50	0.79
K	282,470	305,480	50	10	0.55
LC	19,891	14,061	50	20	−1.14
LDL	16,545	36,002	50	50	3.32
LGN	50,419	31,863	50	19	−1.76
LOC	44,438	145,124	50	50	6.07
MDT	156,886	265,935	50	50	2.64
MS	145,558	201,633	50	50	2.54
NLC	152,121	224,700	50	50	2.89
NME	285,723	457,473	50	50	4.59
PFE	391,254	463,497	50	50	1.90
PVH	64,247	38,944	50	50	−2.53
RAD	198,956	255,645	50	50	1.62
RAL	309,560	218,468	46	50	−3.67
RBD	192,796	179,196	50	21	−0.44
RPR	236,547	275,229	50	50	1.62
SCR	66,287	120,272	50	50	2.52
SHW	120,033	143,251	50	50	1.93
SRR	134,072	114,128	50	8	−0.61
STR	83,144	60,033	50	50	−2.16
SYN	253,170	225,129	50	50	−1.87
UNP	238,209	270,194	50	50	1.51
USS	219,420	245,919	50	50	1.25
VCD	58,521	88,985	50	20	1.71
VNO	26,023	9,347	50	50	−3.17
WAG	166,026	158,293	22	50	−0.40
WTI	79,477	70,421	50	50	−0.92

Table 5.11: Dollar Ask Depth Comparison for Individual Stocks (1991)

Ticker	Total Depth Before Split ($)	Total Depth After Split ($)	N_1	N_2	ζ
ASC	259,912	264,518	50	50	0.17
AUD	292,233	322,318	50	50	1.14
BOL	416,144	366,356	50	50	−0.98
BV	706,921	1328,865	47	50	9.19
CIR	472,461	352,706	50	50	−3.68
CL	529,092	549,562	50	50	0.45
CML	187,673	172,831	50	50	−0.72
CNC	151,260	160,735	50	50	0.59
CPB	384,650	450,483	50	5	0.98
GLX	2,422,423	1,723,247	50	34	−5.11
GPS	429,617	410,638	50	50	−0.67
GPU	682,799	464,438	50	50	−3.08
GS	602,450	694,166	50	50	1.62
GWW	264,232	315,554	50	50	1.70
HF	260,280	118,271	50	50	−2.76
HKF	151,123	97,445	50	50	−3.65
HRB	333,871	472,343	50	50	3.41
IGT	263,927	273,163	50	50	0.22
IMA	286,926	282,622	50	33	−0.22
JEC	101,839	119,184	50	50	1.31
K	492,358	456,238	50	10	−0.57
LC	43,605	25,232	50	20	−2.39
LDL	41,777	54,330	50	50	1.55
LGN	86,691	58,088	50	19	−2.15
LOC	108,962	225,541	50	50	4.88
MDT	252,579	416,611	50	50	3.16
MS	290,365	350,945	50	50	2.07
NLC	260,479	396,494	50	50	3.45
NME	556,729	940,013	50	50	5.94
PFE	700,220	799,354	50	50	1.79
PVH	133,565	72,642	50	50	−4.79
RAD	364,162	448,596	50	50	1.33
RAL	519,655	381,066	46	50	−3.84
RBD	302,291	262,374	50	21	−1.11
RPR	463,279	467,239	50	50	0.11
SCR	115,197	170,223	50	50	2.06
SHW	236,612	291,409	50	50	2.53
SRR	207,868	187,201	50	8	−0.49
STR	159,753	117,406	50	50	−2.54
SYN	473,105	395,901	50	50	−3.14
UNP	478,608	539,962	50	50	1.54
USS	370,382	417,670	50	50	1.45
VCD	95,590	136,123	50	20	1.82
VNO	44,298	24,222	50	50	−3.28
WAG	308,514	275,684	22	50	−1.05
WTI	142,204	123,496	50	50	−1.48

Table 5.12: Total Dollar Depth Comparison for Individual Stocks (1991)

ask depth change for the 46 stocks in the 1991 sample is 0.53.

5.5.2 Splits and Depths for Portfolios

To get a more aggregate measure of what happens to the depth following a stock split, I combine stocks into portfolios and align them by their respective split dates.

One factor that may obscure a cross-sectional analysis of depths is that the level of dollar depths varies quite substantially across stocks. This is also true for percentage bid-ask spreads, but the differences are not nearly as pronounced. For example, the ratio of the largest pre-split percentage spread to the smallest pre-split percentage spread is around 6 for both the 1988 sample (FCB versus MRK, see Table 5.3) and for the 1991 sample (VNO versus PFE, see Table 5.4). But for the total dollar depth the ratios are 17 (SC versus UB, see Table 5.9) and 58 (GLX versus VNO, see Table 5.12), almost an order of magnitude greater than for percentage spreads. Thus it is conceivable that the results for portfolios are mostly driven by those stocks whose dollar depth is relatively large.

I therefore decided to construct both *unweighted* and *weighted* portfolios. Specifically, suppose we are considering a portfolio of N stocks over a $2K$-day period around the split date with K pre-split days and K post-split days. Define D_{ik} as the time-average absolute (dollar) depth of stock i on day k, where $i = 1, \ldots, N$ and $-K \leq k < K$. The convention is that the pre-split period consists of days k with $k < 0$ and the post-split period consists of days k with $k \geq 0$. The designation "day k" is thus always interpreted relative to the split date. The generic term "depth" denotes either bid depth, ask depth, or total depth (all in dollar terms), as defined earlier in this section. The quantities D_{ik} are simply the values whose pre-split means $(\sum_{k<0} D_{ik}/N_1)$ and whose post-split means $(\sum_{k\geq 0} D_{ik}/N_2)$ are reported in Tables 5.7–5.12. The unweighted portfolio depth M_k on day k is defined by

$$M_k \equiv \frac{1}{N} \sum_{i=1}^{N} D_{ik}.$$

Define the overall average dollar depth of stock i by $\overline{D}_i \equiv \sum_k D_{ik}/(2K)$, the average over both of the pre-split and the post-split periods. The relative depth D'_{ik} of stock i on day

k is defined by $D'_{ik} \equiv D_{ik}/\overline{D}_i$ and the weighted portfolio depth M'_k on day k is defined by

$$M'_k \equiv \frac{1}{N}\sum_{i=1}^{N} D'_{ik}.$$

The next two subsections summarize the results for both types of portfolios. To see what effect stock splits have on the depth, I will compare the pre-split mean of $\{M_k\}_{k<0}$ versus the post-split mean of $\{M_k\}_{k\geq 0}$ for unweighted portfolios and the pre-split mean of $\{M'_k\}_{k<0}$ versus the post-split mean of $\{M'_k\}_{k\geq 0}$ for weighted portfolios. I also compare the conclusions derived from the weighted and unweighted portfolios.

Unweighted Portfolios

Tables 5.13 and 5.14 summarize the results for the values of the unweighted portfolio depth M_k (bid depth, ask depth, and total depth are considered separately) for the years 1988 and 1991, respectively. The results for the equally-weighted portfolio for 1988 indicate that depth decreased mildly after the split, but the decrease is insignificant. On the other hand, the results for 1991 indicate that depth actually increased after the split. The bid depth increased insignificantly but the ask depth increased quite significantly: the value of the test statistic ζ whether the mean of the pre-split series is equal to the mean of the post-split series is 3.79, clearly rejecting the equality of the means. The behavior of the total depth mirrors the behavior of the ask depth. Unfortunately, there are indications that the variance of the post-split series is smaller than the variance of the pre-split series (see Table 5.14).

Weighted Portfolios

Finally, the results for the weighted portfolio depth M'_k are summarized in Tables 5.15 (1988) and 5.16 (1991). For the weighted portfolio M'_k, the bid depth changed insignificantly for both of the two sample periods. However, the ask depth decreased significantly for the 1988 sample ($\zeta = -2.02$) and it increased significantly for the 1991 sample ($\zeta = 2.28$). Again, the ratio of post-split sample variance to pre-split sample variance is significantly below 1. Note that if we had just examined the total depth, we would con-

A. Bid Depth ($)

	Before Split	After Split
Mean	99,364	97,821
St. dev.	17,306	19,358

The ζ-statistic for equality of means is -0.42, which is insignificant.
The ratio of post-split sample variance to pre-split sample variance is 1.25, which is insignificant at the 5 percent level (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

B. Ask Depth ($)

	Before Split	After Split
Mean	111,177	106,167
St. dev.	19,697	16,494

The ζ-statistic for equality of means is -1.37, which is insignificant.
The ratio of post-split sample variance to pre-split sample variance is 0.70, which is insignificant at the 5 percent level (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

C. Total Depth ($)

	Before Split	After Split
Mean	210,540	203,988
St. dev.	28,245	29,130

The ζ-statistic for equality of means is -1.13, which is insignificant.
The ratio of post-split sample variance to pre-split sample variance is 1.06, which is insignificant at the 5 percent level (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

Table 5.13: Dollar Depth Comparison for Unweighted Portfolios (1988)

A. Bid Depth ($)

	Before Split	After Split
Mean	134,958	142,019
St. dev.	19,558	19,113

The ζ-statistic for equality of means is 1.83, which is insignificant.
The ratio of post-split sample variance to pre-split sample variance is 0.95, which is insignificant at the 5 percent level (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

B. Ask Depth ($)

	Before Split	After Split
Mean	176,548	193,460
St. dev.	25,981	17,907

The ζ-statistic for equality of means is 3.79, which is highly statistically significant.
The ratio of post-split sample variance to pre-split sample variance is 0.48, which is significant with a p-value of 0.01 (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

C. Total Depth ($)

	Before Split	After Split
Mean	311,506	335,479
St. dev.	38,116	31,021

The ζ-statistic for equality of means is 3.45, which is highly statistically significant.
The ratio of post-split sample variance to pre-split sample variance is 0.66, which is insignificant at the 5 percent level (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

Table 5.14: Dollar Depth Comparison for Unweighted Portfolios (1991)

clude that its change was insignificant for both sample periods. Moreover, the variance ratios are insignificantly different from 1 in both cases.

Thus the results for the depth are inconclusive. The most accurate overall assessment appears to be that the depth remains unchanged after the split. Subsection 5.6.3 considers the depth within the context of transaction volumes.

5.6 Splits and Transaction Volumes

In this section I explore the relation between splits and transaction volumes. Using the complete record of all NYSE transactions for the 50-day pre-split period and for the 50-day post-split period for the stocks in my two samples, subsection 5.6.1 contains a pre-split versus post-split comparison of several daily volume measures and subsection 5.6.2 contains a similar comparison for individual transaction sizes. Finally, subsection 5.6.3 explores the relation between transaction volumes and depth and how this relation is affected by the split.

To put the analysis here in perspective, I first summarize the major findings regarding splits and volumes in earlier empirical work. In an early study of the liquidity effects of stock splits, Copeland (1979) finds that trading volume increases less than proportionately after the split. By considering both the short- and long-term liquidity effects of stock splits, Murray (1985) finds no statistically significant change in proportional trading volume. Lakonishok and Lev (1987) do not find a decrease in marketability as far as the volume of trade is concerned. They document that splitting firms' stocks experience large trading volumes during the year before the announcement and that this can be attributed to the unusual operational performance of these firms.

Lamoureux and Poon (1987) confirm Copeland's finding that the *value* of shares traded falls subsequent to the ex-split day. They attempt to explain the announcement effect with a theory based on an increase of *raw* volume, which is used as a proxy for the number of transactions.

For the two samples of stocks that split by a factor of 2-for-1 or greater, I find that

A. Bid Depth

	Before Split	After Split
Mean	0.9954	1.0020
St. dev.	0.2006	0.2010

The ζ-statistic for equality of means is 0.16, which is insignificant at the 5% level. The ratio of post-split sample variance to pre-split sample variance is 1.00, which is insignificant at the 5 percent level (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

B. Ask Depth

	Before Split	After Split
Mean	1.0350	0.9605
St. dev.	0.2128	0.1454

The ζ-statistic for equality of means is -2.02, which is significant at the 5% level with a p-value of 0.046.
The ratio of post-split sample variance to pre-split sample variance is 0.47, which is significant at the 5 percent level with a p-value of 0.0048 (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

C. Total Depth

	Before Split	After Split
Mean	1.0211	0.9757
St. dev.	0.1494	0.1476

The ζ-statistic for equality of means is -1.52, which is insignificant at the 5% level. The ratio of post-split sample variance to pre-split sample variance is 0.98, which is insignificant at the 5 percent level (2.5% quantile is 0.56 and 97.5% quantile is 1.77 for F distribution with 48 and 48 degrees of freedom).

Table 5.15: Dollar Depth Comparison for Weighted Portfolios (1988)

A. Bid Depth

	Before Split	After Split
Mean	1.0145	0.9855
St. dev.	0.1430	0.1456

The ζ-statistic for equality of means is -1.01, which is insignificant at the 5% level. The ratio of post-split sample variance to pre-split sample variance is 1.04, which is insignificant at the 5 percent level (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

B. Ask Depth

	Before Split	After Split
Mean	0.9666	1.0334
St. dev.	0.1710	0.1163

The ζ-statistic for equality of means is 2.28, which is significant at the 5% level with a p-value of 0.0226.
The ratio of post-split sample variance to pre-split sample variance is 0.46, which is significant at the 5 percent level with a p-value of 0.0079 (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

C. Total Depth

	Before Split	After Split
Mean	0.9867	1.0133
St. dev.	0.1237	0.1087

The ζ-statistic for equality of means is 1.14, which is insignificant at the 5% level. The ratio of post-split sample variance to pre-split sample variance is 0.77, which is insignificant at the 5 percent level (2.5% quantile is 0.57 and 97.5% quantile is 1.76 for F distribution with 49 and 49 degrees of freedom).

Table 5.16: Dollar Depth Comparison for Weighted Portfolios (1991)

the number of daily transactions increases on average by about 36%. Total daily share volume increases roughly proportional to the split factor and total daily dollar volume remains roughly unchanged following the split.

The typical transaction size drops significantly after the split. Both the mean and the median number of shares per trade drop on average by about 25%, whereas both the mean and median dollar volume per trade drop on average by about 20%. Table 5.17 contains an aggregated summary of the results for both of the two samples (1988 and 1991) and for the combined (merged) sample. This table was compiled using the information from Tables 5.18 to 5.23, which are discussed in more detail below.

5.6.1 Splits and Daily Trading Activity

Since my data focus on the microstructure of the trading process, they are not too suitable for a study of daily trading volumes. Given the large variability of daily trading volumes, it is desirable to have data over longer periods of time. Nevertheless, I did calculate the following measures of daily trading activity for the 50-day pre-split period and for the 50-day post-split period for all stocks: the average daily number of trades (NTRADES), the average daily number of shares traded (NSHARES), and the average daily dollar volume of shares traded (NDOLLARS). Only transactions that were reported from the NYSE are included here. The results are summarized in Tables 5.18 and 5.19 for the years 1988 and 1991, respectively. For example, in Table 5.18 we see that Ashland Oil (ticker symbol ASH) had an average of 32 trades per day for the 50-day pre-split period and an average of 33 trades per day for the post-split period.

From Tables 5.18 and 5.19, we see that the number of trades per day increases on average by 37.0% for the 1988 sample and by 36.2% for the 1991 sample. Both of these increases are statistically significant at the 5 percent level (see Table 5.17 for z-values and more details).

Similarly, we conclude from Tables 5.18 and 5.19 that the change in split-adjusted daily share volume and the change in daily dollar volume are both statistically insignificant. What I mean by "split-adjusted" is best illustrated by an example. For Ashland Oil, we

(A) Daily Volume Measures

	1988		1991		1988 and 1991	
Measure	Change	z-value	Change	z-value	Change	z-value
Daily Number of Trades	+37.0%	+2.26	+36.2%	+6.34	+36.5%	+5.30
Daily Share Volume	+8.5%	+0.72	−0.5%	−0.09	+2.7%	+0.50
Daily Dollar Volume	+15.4%	+1.07	+5.7%	+1.10	+9.2%	+1.50

(B) Trade-by-Trade Volume Measures

	1988		1991		1988 and 1991	
Measure	Change	z-value	Change	z-value	Change	z-value
Average Share Volume	−20.0%	−4.42	−26.3%	−8.47	−24.0%	−9.33
Average Dollar Volume	−16.2%	−3.53	−21.7%	−7.53	−19.7%	−7.96
Median Share Volume	−22.5%	−4.19	−26.5%	−8.30	−25.1%	−8.95
Median Dollar Volume	−20.3%	−4.28	−23.2%	−8.46	−22.1%	−9.09

Panel A of this table is compiled from Tables 5.18 and 5.19. Panel B is compiled from Tables 5.20 to 5.23.
In Panel A, the change of the daily number of trades is obtained by averaging across stocks the percentage change of the post-split value of NTRADES (from Tables 5.18 or 5.19) relative to the pre-split value of this variable. The z-value is the value of the statistic obtained by testing the null hypothesis that the daily number of trades remained unchanged. Under this null hypothesis, the test statistic has an approximate standard normal distribution.
Similarly, the daily share volume statistics are obtained from the values of NSHARES and the daily dollar volume statistics are obtained from the values of NDOLLARS. Daily share volumes are split-adjusted, i.e., the ratio of the post-split number of shares to the pre-split number of shares is normalized by the split factor f_i for stock i. The split factors f_i are given in Tables 5.1 and 5.2.
In Panel B, the rows corresponding to averages are obtained from Tables 5.20 and 5.21, whereas the rows corresponding to medians are obtained from Tables 5.22 and 5.23. Again, the statistics for share volumes are split-adjusted.
The pre-split period is taken as the 50-day period prior to the split and the post-split day is taken as the 50-day period subsequent to the split. However, for some stocks the split date is within 50 days of the beginning or end of the calendar year. As long as at least 5 pre-split days and 5 post-split days are available, these stocks are still included in the sample. Thus the actual number of days in the pre-split period and the post-split period is 50 with the following exceptions:

- 1988 sample
  For the pre-split period: F (7 days) and PC (49 days)
  For the post-split period: SC (14 days)
- 1991 sample
  For the pre-split period: BV (47 days), RAL (45 days), VNO (32 days), and WAG (22 days)
  For the post-split period: CPB (5 days), GLX (34 days), IMA (33 days), K (10 days), LC (18 days), LGN (19 days), RBD (21 days), SRR (8 days), and VCD (20 days)

Table 5.17: Summary of Changes in Daily and Individual Transaction Volumes

	NTRADES		NSHARES		NDOLLARS	
Ticker	pre	post	pre	post	pre	post
ASH	32	33	65,900	78,300	$4,689,000	$2,703,000
BCR	40	73	66,400	149,200	2,900,000	3,227,000
BMS	21	29	8,700	18,600	332,000	394,000
CCK	21	34	6,600	42,100	764,000	1,732,000
CTB	13	66	17,400	115,300	690,000	2,837,000
DCI	13	13	9,000	11,900	374,000	258,000
DOV	39	57	53,800	140,100	3,319,000	4,092,000
F	458	446	769,100	988,700	61,098,000	42,722,000
FBO	37	47	41,500	109,600	1,686,000	2,067,000
FCB	7	6	3,600	7,100	176,000	181,000
GW	85	112	139,200	236,100	10,910,000	9,888,000
HLT	37	52	40,000	64,000	3,608,000	3,017,000
IMD	20	28	17,300	47,500	647,000	978,000
MRK	271	273	344,000	661,500	54,300,000	36,466,000
NHY	12	11	47,700	87,700	1,474,000	1,366,000
NWL	14	17	14,500	31,100	638,000	796,000
PC	30	38	58,900	78,700	2,558,000	1,804,000
SBL	55	85	49,600	127,500	2,260,000	2,755,000
SC	30	23	76,500	40,100	5,334,000	1,484,000
SDW	8	7	10,500	15,500	426,000	345,000
SPP	58	64	66,700	144,100	4,713,000	5,389,000
STH	11	26	12,800	46,300	471,000	941,000
TKR	42	42	42,100	66,700	2,903,000	2,063,000
UB	10	8	3,500	10,800	179,000	183,000
VDC	7	6	5,100	9,100	207,000	160,000
WWY	41	53	13,900	32,200	1,049,000	1,177,000

Table 5.18: Daily Trading Activity Measures (1988)

	NTRADES		NSHARES		NDOLLARS	
Ticker	pre	post	pre	post	pre	post
ASC	66	68	68,100	113,700	$5,606,000	$4,355,000
AUD	101	123	91,500	229,600	5,944,000	7,457,000
BOL	66	77	66,000	117,400	5,580,000	5,205,000
BV	361	651	723,900	1,472,200	19,338,000	16,720,000
CIR	76	93	159,000	178,900	11,851,000	6,829,000
CL	153	154	170,000	266,000	13,307,000	10,217,000
CML	34	52	40,600	105,800	1,801,000	2,602,000
CNC	128	145	81,300	120,500	5,489,000	4,698,000
CPB	119	206	129,600	323,900	10,000,000	13,461,000
GLX	798	1,590	1,155,300	2,371,700	58,259,000	69,984,000
GPS	187	308	244,300	414,400	15,129,000	16,886,000
GPU	118	111	88,100	137,500	4,235,000	3,241,000
GS	281	317	404,100	570,600	29,619,000	20,927,000
GWW	50	59	43,000	90,600	3,637,000	4,261,000
HF	18	23	25,500	39,000	1,021,000	937,000
HKF	23	24	31,900	54,800	1,433,000	1,360,000
HRB	85	133	75,400	192,700	4,553,000	6,564,000
IGT	91	125	88,300	171,100	4,349,000	4,869,000
IMA	91	205	74,500	365,000	8,538,000	13,467,000
JEC	24	53	18,900	52,100	708,000	1,166,000
K	148	284	151,400	334,400	16,032,000	20,290,000
LC	6	4	2,800	2,700	119,000	60,000
LDL	6	6	2,700	4,900	108,000	105,000
LGN	22	19	15,600	20,100	559,000	384,000
LOC	31	43	27,600	80,400	1,784,000	2,546,000
MDT	99	155	65,500	138,600	7,984,000	10,009,000
MS	37	69	53,400	156,500	4,711,000	7,890,000
NLC	84	87	95,400	188,200	5,727,000	5,381,000
NME	119	274	264,500	1,169,400	12,072,000	20,155,000
PFE	438	477	525,100	743,000	50,777,000	42,248,000
PVH	20	24	22,200	41,800	743,000	722,000
RAD	87	87	136,200	223,200	5,747,000	4,858,000
RAL	143	149	148,700	192,000	14,865,000	10,684,000
RBD	122	237	81,300	194,300	4,787,000	6,428,000
RPR	39	50	71,400	108,200	5,813,000	4,724,000
SCR	15	19	9,500	40,100	471,000	940,000
SHW	88	99	108,200	212,400	4,523,000	5,077,000
SRR	107	148	119,400	192,400	5,770,000	5,405,000
STR	23	28	24,400	34,200	930,000	671,000
SYN	357	428	378,000	619,400	28,465,000	25,091,000
UNP	141	200	153,700	362,500	13,727,000	17,400,000
USS	170	219	185,400	272,000	21,759,000	18,750,000
VCD	42	58	36,000	79,000	1,241,000	1,357,000
VNO	4	4	3,100	2,400	349,000	65,000
WAG	102	174	120,300	252,700	6,368,000	8,162,000
WTI	59	62	61,800	101,300	2,986,000	2,734,000

Table 5.19: Daily Trading Activity Measures (1991)

see from Table 5.18 that the daily average number of shares traded increases from 65,900 for the pre-split period to 78,300 for the post-split period. The "raw" change in NSHARES is therefore an increase of 18.8%, but since ASH split 2-for-1, the split-adjusted change in the number of shares traded is $[(78{,}300/65{,}900)/2] - 1 = -0.406$, a decline of 40.6%. If we define f_i as the split factor for stock i (the values of f_i are reported in Tables 5.1 and 5.2), the change in NSHARES is thus obtained by normalizing the ratio of the post-split quantity to the pre-split quantity by f_i. Clearly, no such normalization is necessary for the ratios associated with NDOLLARS, the daily dollar volumes.

5.6.2 Splits and Transaction Size

In this subsection I examine what happens to the size of individual transactions following the split. Transaction size is measured using both *share* volume and *dollar* volume.

The average values of all share volumes and dollar volumes for the 50-day pre-split period and for the 50-day post-split period are reported in Tables 5.20 (1988 sample) and 5.21 (1991 sample). We see from these tables that the average post-split share volumes tend to increase after the split. However, after adjusting the share volumes by the split factors f_i (see previous subsection), the average share volume actually *decreased* substantially following the split: a decrease of 20.0% for the 1988 sample and a decrease of 26.3% for the 1991 sample. Both of these declines are highly statistically significant (see Table 5.17 for the z-values and more details).

In light of this evidence, it is not surprising that the average dollar volume per trade also decreased substantially after the split. Indeed, the average dollar volume per trade decreased by 16.2% for the 1988 sample and by 21.7% for the 1991 sample.

It is well known that the empirical distribution of transaction volumes is right-skewed, so that the *median* might be a better summary measure than the mean. I therefore calculated the median share volume and median dollar volume for the 50-day pre- and post-split periods for 1988 (Table 5.22) and for 1991 (Table 5.23). The conclusions closely mirror those obtained for the means. The typical pattern is that the median share volume increases after the split, but the increase is for the most part less than proportional as

	Share Volume		Dollar Volume	
Ticker	Pre Split	Post Split	Pre Split	Post Split
ASH	2,057	2,353	$146,336	$81,184
BCR	1,646	2,042	71,809	44,148
BMS	408	638	15,590	13,517
CCK	317	1,236	36,613	50,793
CTB	1,304	1,757	51,727	43,253
DCI	713	948	29,554	20,680
DOV	1,389	2,469	85,622	72,126
F	1,679	2,217	133,401	95,815
FBO	1,116	2,310	45,303	43,561
FCB	496	1,104	24,296	28,282
GW	1,642	2,109	128,661	88,303
HLT	1,092	1,238	98,573	58,305
IMD	882	1,697	33,040	34,941
MRK	1,267	2,419	200,074	133,362
NHY	3,974	7,682	122,873	119,656
NWL	1,056	1,823	46,521	46,692
PC	1,947	2,088	84,517	47,866
SBL	903	1,492	41,128	32,240
SC	2,521	1,761	175,697	65,144
SDW	1,382	2,138	55,894	47,719
SPP	1,154	2,239	81,575	83,736
STH	1,113	1,760	41,131	35,769
TKR	1,005	1,581	69,321	48,872
UB	353	1,397	18,153	23,783
VDC	787	1,488	31,595	26,073
WWY	342	605	25,870	22,138

Table 5.20: Average Transaction-by-Transaction Trading Activity Measures (1988)

	Share Volume		Dollar Volume	
Ticker	Pre Split	Post Split	Pre Split	Post Split
ASC	1,029	1,663	84,704	63,669
AUD	909	1,873	59,065	60,833
BOL	995	1,524	84,163	67,556
BV	2,003	2,263	53,514	25,703
CIR	2,083	1,914	155,237	73,072
CL	1,109	1,723	86,791	66,192
CML	1,206	2,023	53,472	49,762
CNC	637	831	42,974	32,407
CPB	1,089	1,575	84,017	65,472
GLX	1,447	1,492	72,989	44,020
GPS	1,306	1,345	80,897	54,793
GPU	744	1,243	35,736	29,289
GS	1,439	1,799	105,482	65,978
GWW	853	1,543	72,193	72,565
HF	1,408	1,726	56,368	41,406
HKF	1,360	2,302	61,049	57,194
HRB	887	1,450	53,560	49,375
IGT	971	1,364	47,828	38,813
IMA	814	1,776	93,329	65,537
JEC	773	992	29,033	22,198
K	1,024	1,178	108,384	71,467
LC	501	631	21,409	14,126
LDL	425	844	17,128	17,913
LGN	724	1,050	26,016	20,099
LOC	882	1,882	57,060	59,565
MDT	661	895	80,488	64,630
MS	1,436	2,263	126,648	114,049
NLC	1,142	2,153	68,550	61,565
NME	2,226	4,264	101,584	73,493
PFE	1,198	1,557	115,898	88,526
PVH	1,103	1,773	36,933	30,605
RAD	1,575	2,573	66,445	55,993
RAL	1,038	1,293	103,722	71,907
RBD	665	819	39,168	27,095
RPR	1,848	2,183	150,367	95,310
SCR	647	2,079	31,927	48,735
SHW	1,227	2,138	51,294	51,116
SRR	1,117	1,302	53,962	36,584
STR	1,064	1,227	40,526	24,097
SYN	1,060	1,446	79,838	58,567
UNP	1,089	1,814	97,312	87,054
USS	1,087	1,242	127,616	85,638
VCD	851	1,355	29,299	23,256
VNO	781	625	88,049	17,045
WAG	1,178	1,452	62,374	46,897
WTI	1,052	1,623	50,815	43,807

Table 5.21: Average Transaction-by-Transaction Trading Activity Measures (1991)

	Share Volume		Dollar Volume	
Ticker	Pre Split	Post Split	Pre Split	Post Split
ASH	500	700	$34,500	$24,500
BCR	400	500	17,600	11,813
BMS	100	200	4,113	4,375
CCK	200	300	22,350	12,938
CTB	500	1,000	20,125	23,700
DCI	200	400	8,800	8,200
DOV	400	500	23,550	14,438
F	900	600	69,300	25,800
FBO	300	800	13,013	15,500
FCB	200	500	10,300	12,250
GW	500	800	37,813	33,750
HLT	400	500	36,550	23,375
IMD	300	400	11,100	8,600
MRK	700	1,000	112,525	55,250
NHY	1,500	2,500	46,500	38,750
NWL	300	500	13,125	11,500
PC	500	600	21,500	13,200
SBL	300	400	13,125	9,125
SC	400	400	27,800	14,400
SDW	500	600	20,938	13,800
SPP	500	700	33,375	27,125
STH	300	500	11,250	10,563
TKR	200	300	14,325	9,638
UB	200	500	10,350	7,875
VDC	300	500	11,663	8,125
WWY	100	300	8,100	10,950

Table 5.22: Medians of Per Trade Share Volumes and Dollar Volumes (1988)

	Share Volume		Dollar Volume	
Ticker	Pre Split	Post Split	Pre Split	Post Split
ASC	300	500	$24,413	$19,813
AUD	400	600	25,800	21,450
BOL	200	400	17,675	17,450
BV	400	500	11,850	5,875
CIR	600	600	42,900	22,200
CL	400	600	31,900	22,425
CML	500	600	20,250	15,525
CNC	500	500	30,625	19,125
CPB	500	700	39,188	30,013
GLX	300	400	14,100	11,750
GPS	500	500	31,625	21,813
GPU	100	300	4,963	6,788
GS	500	500	36,438	18,938
GWW	200	400	16,575	18,800
HF	500	600	19,650	14,550
HKF	300	400	14,925	10,000
HRB	200	400	13,050	13,850
IGT	300	500	16,725	13,000
IMA	200	500	24,375	18,438
JEC	300	400	11,438	8,800
K	500	500	51,625	28,000
LC	300	400	12,600	9,000
LDL	200	400	8,550	8,600
LGN	300	500	11,138	9,500
LOC	200	500	13,900	15,750
MDT	200	300	25,650	22,950
MS	700	1,000	63,700	49,275
NLC	300	500	18,338	14,563
NME	600	1,000	25,800	15,188
PFE	500	500	45,313	28,188
PVH	500	600	15,750	10,800
RAD	400	600	16,900	12,300
RAL	400	500	40,650	27,000
RBD	300	300	17,513	10,350
RPR	600	700	48,525	29,225
SCR	300	700	14,363	16,100
SHW	300	500	12,825	11,625
SRR	400	400	18,200	11,950
STR	400	400	15,200	8,200
SYN	400	500	30,400	19,938
UNP	500	800	41,438	37,400
USS	500	500	60,000	35,875
VCD	300	500	11,325	8,313
VNO	100	400	13,600	10,500
WAG	400	500	21,650	16,000
WTI	300	400	15,263	10,800

Table 5.23: Medians of Per Trade Share Volumes and Dollar Volumes (1991)

measured by the split factors f_i. If we calculate the ratios r_i of post-split medians to pre-split medians for each stock i and compare these ratios to the split factors f_i from Tables 5.1 and 5.2, we find that the average value of the quantities r_i/f_i declined by 22.5% for the 1988 sample and by 26.5% for the 1991 sample. Median dollar volumes give a similar conclusion. The average ratio of the post-split median dollar volume to the pre-split median dollar volume declined by 20.3% for the 1988 sample and by 23.2% for the 1991 sample. All of these declines are highly statistically significant (see Table 5.17).

5.6.3 Splits and Transaction Size Vis-A-Vis Depth

In addition to analyzing the behavior of transaction volumes by themselves, I also carried out an exploratory analysis comparing transaction volumes to prevailing depths. Recall that the bid (ask) depth is the number of shares available at the bid (ask). It is thus interesting to know what the fraction of trades is for which this "constraint" is binding and to see whether this fraction changes after the split. One could argue that the higher the fraction of trades at or below the quoted depth, the higher the liquidity of the market for that stock.

One difficulty with this analysis is that we need to know whether a given trade is a buy or a sell, because we need to compare the volume of the trade with the quoted depth. I used the convention to classify all trades as buy, sell, or neutral, depending on the transaction price relative to the midquote.[9] Thus the comparison is only carried out for trades classified as either buys or sells—neutral transactions are discarded. Only those trades and quotes that originated from the NYSE are considered here.

The results for the 1988 sample are reported in Tables 5.24 (for buys) and 5.25 (for sells). For example, we see from Table 5.24 that Ashland Oil (ticker symbol ASH) had 712 buys in the 50-day pre-split period, corresponding to 42.92% of all transactions in this period. The percentage of these buys that were classified as "small," meaning that the

[9] Specifically, let P denote the transaction price of a given trade and let B and A denote the outstanding bid and ask quote at the time of this trade. If $P = (B+A)/2$, the transaction is classified as being neutral. If $P > (B + A)/2$, the transaction is classified as a buy (buyer-initiated) and if $P < (B + A)/2$, the transaction is classified as a sell (seller-initiated).

volume was smaller than or equal to the quoted ask depth, was 61.24%. For the post-split period, the percentage of small buys was 61.09%, a marginal decline. The definition of "small" sells is analogous: sell transactions whose volume is smaller than or equal to the quoted bid depth. From Table 5.25, we see that the percentage of small sells for Ashland Oil increases from 66.72% in the pre-split period to 69.28% in the post-split period. The corresponding results for the 1991 sample are reported in Tables 5.26 (for buys) and 5.27 (for sells).

Ticker	# buys pre	% buys pre	% small pre	# buys post	% buys post	% small post
ASH	712	42.92	61.24	640	37.23	61.09
BCR	963	46.41	80.48	1,569	42.28	83.37
BMS	561	49.73	86.10	775	51.05	83.48
CCK	439	39.41	83.83	865	48.93	83.24
CTB	373	50.61	84.72	1,716	51.45	83.86
DCI	280	40.11	69.29	292	42.44	61.64
DOV	759	38.14	68.38	1,156	40.01	84.86
F	1,438	44.76	82.27	10,036	44.92	90.14
FBO	881	46.10	80.48	1,066	43.94	84.33
FCB	181	43.10	88.40	128	33.60	71.88
GW	1,863	43.43	76.22	2,489	44.07	80.63
HLT	964	51.11	86.83	1,183	44.83	83.94
IMD	398	38.16	71.11	561	38.56	82.89
MRK	6,394	46.94	79.75	5,838	42.54	84.52
NHY	253	37.93	85.38	263	42.08	83.27
NWL	314	42.03	80.89	439	47.98	83.14
PC	697	45.38	80.20	833	43.00	85.11
SBL	1,165	41.64	84.46	2,190	50.67	92.74
SC	874	54.73	91.53	170	50.00	93.53
SDW	193	42.42	70.47	155	36.13	79.35
SPP	1,237	42.09	75.75	1,292	39.53	79.26
STH	309	48.36	77.35	799	57.98	84.98
TKR	1,048	48.74	76.34	1,094	50.55	81.08
UB	230	44.15	84.35	204	45.43	87.25
VDC	164	46.86	93.29	134	35.45	82.09
WWY	1,044	50.12	76.63	1,505	55.51	79.07

Table 5.24: Buy Transactions Relative to Depth (1988)

Ticker	# sells pre	% sells pre	% small pre	# sells post	% sells post	% small post
ASH	679	40.93	66.72	765	44.50	69.28
BCR	707	34.07	74.82	1,286	34.65	83.13
BMS	436	38.65	79.59	597	39.33	82.75
CCK	397	35.64	85.64	723	40.89	80.36
CTB	264	35.82	78.05	1,095	32.83	74.43
DCI	355	50.86	77.75	303	44.04	70.63
DOV	767	38.54	70.93	909	31.46	78.55
F	1,139	35.45	82.18	7,915	35.43	88.29
FBO	639	33.44	75.27	687	28.32	77.44
FCB	203	48.33	88.67	212	55.64	83.49
GW	1,944	45.31	79.94	2,344	41.50	82.89
HLT	644	34.15	81.21	1,109	42.02	78.36
IMD	493	47.27	76.27	612	42.06	81.86
MRK	4,537	33.31	77.03	4,096	29.85	82.42
NHY	259	38.83	82.24	192	30.72	80.21
NWL	296	39.63	81.76	310	33.88	80.32
PC	579	37.70	77.03	744	38.41	82.39
SBL	966	34.52	78.36	1,121	25.94	78.95
SC	432	27.05	84.03	104	30.59	85.58
SDW	174	38.24	56.32	225	52.45	68.44
SPP	1,103	37.53	73.89	1,269	38.83	81.09
STH	228	35.68	69.74	357	25.91	72.83
TKR	876	40.74	69.75	726	33.55	69.70
UB	195	37.43	76.92	189	42.09	84.66
VDC	137	39.14	75.18	193	51.06	77.72
WWY	738	35.43	65.31	904	33.35	60.95

Table 5.25: Sell Transactions Relative to Depth (1988)

Ticker	# buys pre	% buys pre	% small pre	# buys post	% buys post	% small post
ASC	1,360	40.43	74.56	1,625	46.83	82.22
AUD	1,964	38.59	83.30	2,530	40.81	84.94
BOL	1,487	44.14	85.68	1,777	45.49	84.41
BV	7,354	43.15	95.39	19,556	59.99	99.05
CIR	1,797	46.35	82.97	1,972	41.52	88.03
CL	3,534	45.74	90.46	3,924	50.50	92.28
CML	650	37.27	78.31	1,006	37.51	83.50
CNC	2,579	40.03	84.92	3,508	48.05	87.46
CPB	2,616	43.58	84.56	504	48.79	91.07
GLX	24,075	60.24	98.22	31,675	58.56	97.99
GPS	4,060	43.16	89.83	6,309	40.79	92.47
GPU	4,281	71.64	97.99	3,631	65.03	97.41
GS	6,119	43.41	88.74	7,752	48.72	93.49
GWW	1,072	41.52	84.79	1,223	40.92	86.10
HF	382	39.59	66.75	585	49.04	77.61
HKF	675	54.35	78.07	617	49.24	80.71
HRB	2,022	47.02	87.29	3,367	50.28	89.66
IGT	2,245	48.81	83.30	3,218	50.90	86.76
IMA	2,285	49.39	83.06	3,374	49.51	86.43
JEC	538	41.80	86.99	1,472	54.82	91.51
K	3,801	51.04	90.16	1,715	60.20	91.31
LC	125	38.58	78.40	41	35.96	63.41
LDL	67	22.19	83.58	180	52.79	88.33
LGN	526	46.43	84.41	159	41.19	79.87
LOC	667	40.62	71.81	1,053	47.78	85.19
MDT	2,218	44.21	81.47	3,777	48.43	85.62
MS	774	40.10	73.51	1,458	41.42	76.82
NLC	1,715	40.42	84.96	2,046	46.03	88.95
NME	2,417	40.26	87.55	6,798	49.38	94.41
PFE	9,842	44.81	82.25	11,495	48.03	89.46
PVH	439	40.80	80.18	572	46.47	76.40
RAD	1,925	43.86	90.39	2,112	47.96	91.86
RAL	3,204	49.31	87.64	3,376	45.05	89.54
RBD	3,281	53.25	93.84	3,072	61.40	94.86
RPR	955	47.32	75.39	1,418	55.52	85.47
SCR	425	54.84	87.76	552	53.28	79.89
SHW	1,736	38.85	89.63	1,700	33.78	88.94
SRR	2,391	44.19	83.02	600	50.17	87.67
STR	545	44.78	81.65	584	39.46	86.30
SYN	7,020	39.27	86.62	10,257	47.77	92.03
UNP	2,822	39.58	85.05	3,732	37.04	87.06
USS	3,609	42.05	81.08	5,193	47.18	87.23
VCD	1,059	48.25	80.83	638	53.39	92.32
VNO	58	30.69	70.69	58	28.29	70.69
WAG	860	37.87	88.14	3,643	41.55	90.83
WTI	1,187	39.51	82.90	1,619	50.93	85.55

Table 5.26: Buy Transactions Relative to Depth (1991)

Ticker	# sells pre	% sells pre	% small pre	# sells post	% sells post	% small post
ASC	1,282	38.11	69.97	1,211	34.90	79.11
AUD	1,614	31.71	81.78	1,877	30.28	76.08
BOL	1,157	34.34	82.28	1,328	34.00	80.35
BV	4,321	25.35	91.92	9,133	28.02	96.55
CIR	1,237	31.91	73.65	1,255	26.42	76.57
CL	2,493	32.27	86.96	2,319	29.85	86.63
CML	576	33.03	82.47	843	31.43	76.87
CNC	2,160	33.52	79.58	2,661	36.45	85.08
CPB	2,116	35.25	76.28	333	32.24	73.57
GLX	8,787	21.99	94.71	15,125	27.96	96.63
GPS	2,656	28.23	80.87	4,122	26.65	85.98
GPU	898	15.03	82.96	1,128	20.20	89.36
GS	3,777	26.79	82.00	3,579	22.49	85.22
GWW	981	37.99	82.06	1,128	37.74	82.36
HF	403	41.76	78.41	401	33.61	69.83
HKF	404	32.53	70.54	448	35.75	69.42
HRB	1,569	36.49	85.40	2,233	33.34	83.70
IGT	1,714	37.27	79.52	1,920	30.37	80.05
IMA	1,583	34.22	77.70	2,094	30.73	78.03
JEC	382	29.68	81.94	762	28.38	80.05
K	2,674	35.91	84.26	787	27.62	85.39
LC	122	37.65	65.57	56	49.12	55.36
LDL	122	40.40	89.34	96	28.15	68.75
LGN	388	34.25	82.73	185	47.93	83.78
LOC	670	40.80	72.54	556	25.23	69.24
MDT	1,502	29.94	72.30	2,572	32.98	77.10
MS	627	32.49	68.10	1,181	33.55	75.11
NLC	1,281	30.19	78.92	1,208	27.18	83.86
NME	1,882	31.35	82.52	4,170	30.29	88.35
PFE	8,718	39.69	81.31	7,503	31.35	83.47
PVH	444	41.26	80.18	494	40.13	68.02
RAD	1,453	33.11	83.14	1,295	29.41	89.03
RAL	2,243	34.52	83.82	2,446	32.64	83.20
RBD	1,703	27.64	85.85	1,180	23.59	85.17
RPR	663	32.85	73.00	552	21.61	72.28
SCR	275	35.48	76.36	343	33.11	74.34
SHW	1,633	36.55	88.18	1,610	32.00	92.36
SRR	1,720	31.79	79.24	276	23.08	82.97
STR	450	36.98	81.33	435	29.39	82.53
SYN	6,596	36.89	86.78	5,685	26.48	83.55
UNP	2,417	33.90	84.86	3,281	32.56	86.28
USS	2,471	28.79	74.18	2,993	27.19	77.61
VCD	687	31.30	75.55	327	27.36	80.73
VNO	68	35.98	76.47	81	39.51	69.14
WAG	656	28.89	82.16	2,227	25.40	83.43
WTI	1,100	36.62	83.18	1,000	31.46	79.50

Table 5.27: Sell Transactions Relative to Depth (1991)

From Tables 5.24 to 5.27, we see that for the combined sample of the two years 1988 and 1991, the average fraction of small buys increased from 82.6% before the split to 85.2% after the split and the average fraction of small sells increased from 78.9% before the split to 79.6% after the split. This change is statistically significant for small buys, but insignificant for small sells.[10] It is not obvious how to make a judgment whether the statistically significant change is also economically significant. If it were, then liquidity has increased using this measure.

As an aside, inspection of Tables 5.24 to 5.27 reveals a curious empirical regularity. Taking the average of the pre-split and post-split fractions of small trades, the fraction of small buys is greater than the fraction of small sells for both of the two samples. For the 1988 sample, the fraction of small buys is 80.9% and the fraction of small sells is 77.3%, giving a z-value of 3.05 for the test of equality of the two fractions. For the 1991 sample, the fraction of small buys is 85.6% and the fraction of small sells is 80.3%, giving a z-value of 10.36 for the standardized mean of the series of paired differences. Transaction volumes and depths are apparently such that buys are smaller relative to ask depths than sells relative to bid depths. Perhaps it is the case that large buy orders are broken into more and smaller pieces than large sell orders. Various theoretical models have been proposed to focus on some aspect of buy-sell asymmetries (e.g., Allen and Gorton, 1991), but I am not aware of any models that have predictions of this nature.

5.7 Application of Ordered-Probit Model

In this section I use the ordered probit methodology from the previous chapter to analyze the liquidity effects associated with stock splits. Using the 1988 and 1991 samples of stocks that split by 2-for-1 or greater, I show that the event of a split changes the bid-ask spread distribution after taking the level shift into account. Specifically, large trades widen the percentage spread more after the split than before the split. This result reinforces the conclusion that, contrary to market folklore, stock splits actually *decrease* market liquidity.

[10]The z-value for the test of a zero mean of the series of $26 + 46 = 72$ paired differences is 4.00 for the case of small buys and 1.12 for the case of small sells.

Using the maximum-likelihood parameter estimates for the ordered-probit parameters corresponding to equation (4.4) for the pre-split period and the post-split period, I calculate the distribution of the bid-ask spread for two scenarios: (1) a base case in which all variables are set equal to their sample medians; and (2) a situation where a large trade has just taken place (buys and sells are considered separately).

For concreteness, I illustrate the methodology for Crown Cork & Seal (ticker symbol CCK), a stock that split 3-for-1 in 1988. Figures 5.6.a and 5.6.b display two probability distributions for the pre-split bid-ask spread. The scenarios assumed for these two distributions are identical except for the value of BDOL_k, the dollar volume of a buy transaction at time k. Specifically, (i) all regressors except BDOL_k are set equal to their (pre-split) median values; (ii) BDOL_k is set at the median (transformed) buy dollar volume in figure 5.6.a and at the 90th percentile of the (transformed) buy dollar volume in figure 5.6.b. The probabilities are then calculated using the maximum-likelihood estimates obtained for the pre-split period. A comparison of figures 5.6.a and 5.6.b allows us to examine the effect on the spread of a "large" buy ($102,600) compared to a "medium-size" buy ($22,350). It can be seen that the spread distribution shifts to the right. Detailed calculations reveal that the average spread increases from 4.2859 ticks to 4.4127 ticks, roughly a 3 percent increase. See Table 5.28 for complete details.

Figures 5.6.c and 5.6.d illustrate the same basic comparison for the post-split period. In this case, a "large" buy ($113,750) increases the expected spread from 2.1671 ticks for a "medium-size" buy ($18,750) to 2.3212 ticks, an increase of about 7 percent. The conclusion is that for this particular stock, a similar event (a large buy vis-à-vis a medium-size buy) increases the spread more (on average) after the split than before the split. Thus the split effectively *reduced* the liquidity of the market for Crown Cork & Seal. See Table 5.28 for a summary of the basic logic applied to CCK.

Table 5.29 summarizes the results of the same calculations for all 26 NYSE stocks that experienced a split of 100% or greater in 1988. For 23 out of these 26 stocks, the same result applies: following a large buy, spreads increase more after the split than before the split. Although the sample is not large, a simple test whether the effects are the same easily

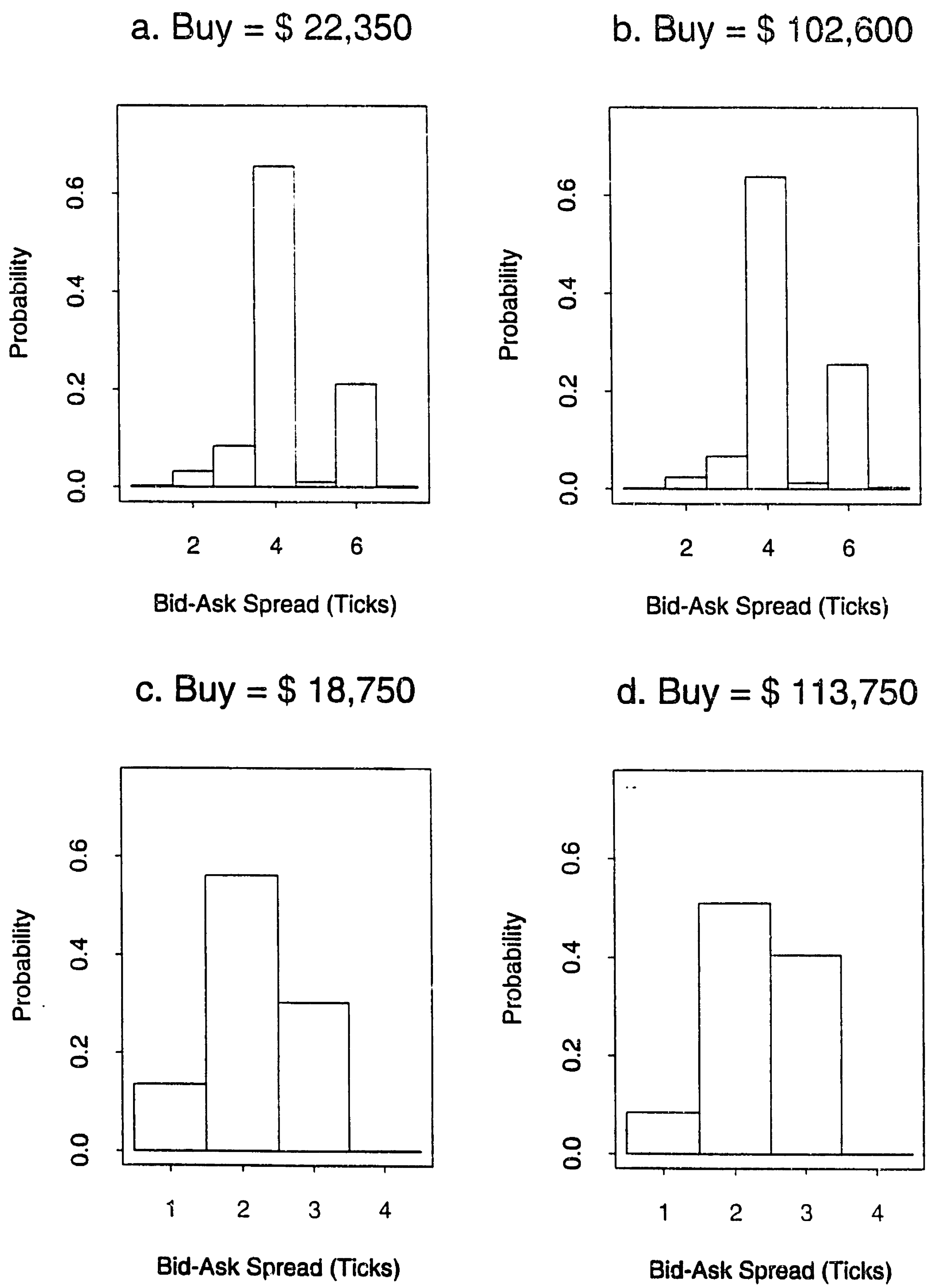


Figure 5.6: Spread Distributions for Crown Cork & Seal

CCK Before Split Figure a and b	Medium Buy	Large Buy	% Change
$P\{S_k = 1\}$	0.0023	0.0014	
$P\{S_k = 2\}$	0.0324	0.0236	
$P\{S_k = 3\}$	0.0836	0.0672	
$P\{S_k = 4\}$	0.6570	0.6378	
$P\{S_k = 5\}$	0.0106	0.0117	
$P\{S_k = 6\}$	0.2121	0.2549	
$P\{S_k = 7\}$	0.0021	0.0033	
$E[S_k]$ (ticks)	4.2859	4.4127	2.957
$E[S_k]$ (% price)	0.504	0.519	2.957

CCK After Split Figure c and d	Medium Buy	Large Buy	% Change
$P\{S_k = 1\}$	0.1360	0.0846	
$P\{S_k = 2\}$	0.5610	0.5100	
$P\{S_k = 3\}$	0.3028	0.4050	
$P\{S_k = 4\}$	0.0001	0.0004	
$E[S_k]$ (ticks)	2.1671	2.3212	7.112
$E[S_k]$ (% price)	0.637	0.683	7.112

Difference (After − Before)	4.155

Table 5.28: Split Effect Analysis on Crown Cork & Seal

rejects this null hypothesis if we assume independence across stocks. Indeed, the mean and standard deviation of the series of differences are 2.579 and 2.920, respectively, so that the value of the test statistic is given by $2.579/(2.920/\sqrt{26}) = 4.50$. This test statistic is approximately standard normal under the null hypothesis of no effect associated with the split, so that the p-value associated with this test is negligible.[11] The results for the 1991 sample closely mirror those obtained for the 1988 sample. The corresponding z-statistics for the 1991 sample are 3.31 for buys and 5.03 for sells, thus reinforcing the conclusions of this section.[12]

5.8 Conclusion

This chapter is an empirical study of the liquidity effects associated with stock splits, focusing on measures of trading activity, bid-ask spreads, and depths for a sample of NYSE companies that split their stocks by 2-for-1 or greater in the two years 1988 and 1991. The objective is to investigate whether stock splits have an effect on liquidity and if so, whether liquidity increases or decreases after the split.

Although the liquidity measures considered here are not all in agreement, the picture that emerges from this study is that liquidity tends to decline after the split. It is unambiguously clear that percentage bid-ask spreads increase after the split. The results for depths are mixed, however. Using weighted portfolios to account for the different magnitude of (dollar) depths across stocks, I show that the bid depth remains unchanged for both of the two sample periods, but the ask depth decreased significantly for the 1988 sample and it increased significantly for the 1991 sample. In light of this conflicting evidence, the conclusion is that the depth does not change after the split. Thus if liquidity

[11] The same result holds if we consider sells instead of buys. For sells, the mean and standard deviation of the series of differences are 2.429 and 2.369, respectively, so that the test statistic equals $2.429/(2.369/\sqrt{26}) = 5.23$, again a very strong rejection of the null hypothesis of no difference.

[12] The ordered probit model was estimated for all stocks of the 1991 sample with the exception of VNO, whose pre-split spread was never 1 tick, while it exceeded 4 ticks in about 95% of the observations. Since the number of observations for VNO is quite small (well below 1,000) for both the pre-split period and the post-split period, I decided to exclude this stock rather than redefining the values of s_j (see section 4.3.1) to make estimation possible. The number of stocks in the 1991 sample was therefore equal to 45 for the purposes of the experiment described here.

	% Change Before	% Change After	Difference
ASH	7.038	5.811	−1.227
BCR	7.115	7.532	0.417
BMS	5.216	10.661	5.445
CCK	2.957	7.112	4.155
CTB	5.540	10.392	4.852
DCI	7.255	11.357	4.102
DOV	4.810	9.855	5.045
F	7.499	9.996	2.497
FBO	7.617	9.208	1.591
FCB	6.811	9.649	2.838
GW	9.628	9.694	0.066
HLT	7.429	7.045	−0.384
IMD	6.863	11.357	4.494
MRK	5.391	6.501	1.110
NHY	7.722	10.265	2.543
NWL	7.245	9.680	2.435
PC	8.769	9.704	0.935
SBL	5.541	7.797	2.256
SC	15.588	8.722	−6.866
SDW	8.846	12.634	3.788
SPP	3.768	6.368	2.600
STH	7.569	10.913	3.344
TKR	6.004	9.022	3.018
UB	6.001	11.659	5.658
VDC	4.487	13.673	9.186
WWY	2.605	5.759	3.154
mean	6.743	9.321	2.579
SD	2.503	2.088	2.920

Table 5.29: Comparison of Spread Changes Using Ordered Probit (1988)

is measured by spread and depth, liquidity decreases after the split.

In addition to examining what happens to the depth around stock splits, I also explore the depth in relation to trading volumes and the behavior of trading volumes around splits. I show that for the stocks in my sample the daily number of transactions increases after the split, but total daily dollar volume and split-adjusted share volume remain unchanged. The dollar volume and split-adjusted share volume of a typical transaction decline on average by about 20% and 25%, respectively. Both the mean and the median are used for a "typical" transaction and the conclusions are remarkably similar for both measures.

As far as the relation between volumes and depths is concerned, I show that the fraction of trades with volumes not exceeding the quoted depth increases after the split for transactions classified as buys, whereas it remains unchanged for transactions classified as sells. Thus from this perspective it appears that liquidity actually increased after the split.

Finally, this chapter considers the intraday properties of the bid-ask spread. By using the ordered-probit model for the discrete-valued bid-ask spread, it is shown that stock splits decrease liquidity in the sense that large trades tend to widen the percentage spread more after the split than before the split.

Appendix A

Theory of Runs

Suppose we have a sequence consisting of B's and S's. We are interested in the distribution of the number of runs. Specifically, assume that we have b symbols marked "B" and s symbols marked "S." Assume that all distinguishable orderings are equally likely and denote by P_k the probability that the arrangement of B's and S's contains exactly k runs.

To find the probability P_k, first note that there are $\binom{b+s}{s}$ distinguishable orderings of the collection of b B's and s S's. Consider the case that k is even, i.e., $k = 2\nu$ for some integer ν. In this case there must be ν B-runs and also ν S-runs. Since there are $\binom{b-1}{\nu-1}$ ways to distribute b symbols over ν cells leaving none of the cells empty and $\binom{s-1}{\nu-1}$ ways to distribute s symbols over ν cells under the same requirement, it follows that

$$P_{2\nu} = 2\binom{b-1}{\nu-1}\binom{s-1}{\nu-1} \Big/ \binom{b+s}{s}$$

when $k = 2\nu$ is even, since there are two possible permutations of the B-runs and S-runs.

Now consider the case that k is odd, i.e., $k = 2\nu + 1$ for some integer ν. In this case there are either $\nu + 1$ B-runs and ν S-runs or there are ν B-runs and $\nu + 1$ S-runs. Using the same argument as before, we conclude that

$$P_{2\nu+1} = \left\{\binom{b-1}{\nu}\binom{s-1}{\nu-1} + \binom{b-1}{\nu-1}\binom{s-1}{\nu}\right\} \Big/ \binom{b+s}{s}$$

when $k = 2\nu + 1$ is odd. Note that the probability distribution derived here is conditional on b and s.[1]

[1] The probability distribution derived above is also given in exercise II.20 of Feller (1968, Vol. 1, p. 62).

There is a normal approximation to this exact distribution. If b and s are large enough (say at least 10), then the distribution for P_k tends to the distribution of a normal random variable Z whose first two moments are given by (e.g., see Mood, Graybill, and Boes, p. 521)

$$E[Z] = \frac{2bs}{b+s} + 1,$$

and

$$\text{var}(Z) = \frac{2bs(2bs - b - s)}{(b+s)^2(b+s-1)}.$$

Note that this approximation is most accurate when b and s are roughly equal to each other. In particular for tail probabilities, exact and approximate values can differ by orders of magnitude when b and s are substantially different (as they frequently are in our sample).

Appendix B

The Dynamics of the Bid-Ask Spread: Data

The data used in Chapter 4 of this thesis is taken from the database for 1988 provided by the Institute for the Study of Security Markets (ISSM). This database contains a time-stamped record of all trades and quotes for all securities listed on the NYSE and the AMEX for the entire year.

My sample consists of the 30 NYSE stocks that were included in the Dow Jones Industrial Average (DJIA) in 1988. See Table B.1 for a list of these stocks and their associated ticker symbols.[1] These stocks were chosen to get a representative sample of the most heavily-traded stocks from the New York Stock Exchange.

The remainder of this appendix is devoted to a detailed discussion of the data. As I will distinguish between quotes and trades originating from the New York Stock Exchange versus quotes and trades originating from elsewhere, I first give a summary of the originating exchanges for all quotes and trades in section B.1. A discussion of the data and of the various screens imposed on the data is presented in sections B.2 and B.3 for quotes and trades, respectively. This discussion includes a detailed summary of the various quote and trade condition codes encountered on the ISSM tapes. Since I will distinguish between buy transactions and sell transactions as far as their impact on the bid-ask spread

[1]Subsequent to 1988 the collection of stocks comprising the DJIA has changed. As of November 1993, NAV, PA, and X have left the index and their places are taken by CATERPILLAR INC DE (CAT), DISNEY WALT CO (DIS), and MORGAN J P & CO INC (JPM).

Name	Ticker
ALLIED SIGNAL INC	ALD
ALUMINUM CO AMER	AA
AMERICAN EXPRESS CO	AXP
AMERICAN TEL & TELEG CO	T
BETHLEHEM STL CORP	BS
BOEING CO	BA
CHEVRON CORP	CHV
COCA COLA CO	KO
DU PONT E I DE NEMOURS & CO	DD
EASTMAN KODAK CO	EK
EXXON CORP	XON
GENERAL ELEC CO	GE
GENERAL MTRS CORP	GM
GOODYEAR TIRE & RUBR CO	GT
INTERNATIONAL BUSINESS MACHS	IBM
INTERNATIONAL PAPER CO	IP
MCDONALDS CORP	MCD
MERCK & CO INC	MRK
MINNESOTA MNG & MFG CO	MMM
NAVISTAR INTL CORP	NAV
PHILIP MORRIS COS INS	MO
PRIMERICA CORP	PA
PROCTER & GAMBLE CO	PG
SEARS ROEBUCK & CO	S
TEXACO INC	TX
UNION CARBIDE CORP	UK
UNITED TECHNOLOGIES CORP	UTX
USX CORP	X
WESTINGHOUSE ELEC CORP	WX
WOOLWORTH F W CO	Z

Table B.1: Dow Jones Industrials

Names and ticker symbols of the 30 stocks that were part of the Dow Jones Industrial Average (DJIA) in 1988.

is concerned, I describe the buy-sell classification of trades in section B.4.

B.1 Originating Exchanges

The focus of my bid-ask spread study on DJIA stocks is on the activity taking place on the New York Stock Exchange. For this reason and also to see empirically how activity is distributed across the various exchanges and networks, I have summarized in Tables B.2 and B.3 the frequency distribution by exchange of all quotes and trades that were reported on the 1988 ISSM tapes for the stocks in my sample. As can be expected for stocks whose primary exchange is the New York Stock Exchange, the NYSE (identifier 'N') is the dominating exchange for all stocks, both for quotes and for trades. However, the percentages of NYSE quotes versus non-NYSE quotes and of NYSE trades versus non-NYSE trades vary widely. The fraction of NYSE quotes relative to all quotes ranges from 28.6 percent for Exxon (XON) to 67.0 percent for International Paper (IP) and the average for the 30 stocks is 50.1 percent (see Table B.2). Similarly, the fraction of NYSE trades relative to all trades ranges from 47.4 for Navistar (NAV) to 81.2 percent for International Paper (IP) and the average for the 30 stocks is 67.7 percent (see Table B.3).[2]

Perhaps the least-known exchange appearing in Table B.3 is Instinet (identifier 'O'), originally called Institutional Network. This automated proprietary network is used by subscribing institutional investors, who can submit anonymous bids and offers and execute trades with other subscribers. Trading volume on Instinet is relatively low. Finally, none of the 30 stocks had any trades reported on the Consolidated Tape System (identifier 'S').

B.2 Quotes

This section contains a discussion of the various screens applied to the raw quotation data extracted from the ISSM tapes and a summary of the screened NYSE quotes.

[2] Note that this comparison ignores the fact that trade sizes differ substantially across the various exchanges.

	B	C	M	N	P	T	X	NQ
AA	0.2	6.2	8.0	57.8	19.0	4.5	4.2	31,554
ALD	0.1	4.0	12.1	56.7	22.5	4.0	0.6	27,095
AXP	1.1	12.2	12.1	38.0	20.6	2.8	13.3	82,024
BA	1.7	3.9	9.9	65.1	12.7	3.5	3.2	43,379
BS	0.4	8.7	19.5	49.3	15.7	6.1	0.4	56,277
CHV	1.1	12.8	12.3	46.9	21.4	1.4	4.1	76,449
DD	6.9	26.4	8.3	48.2	7.6	1.1	1.5	106,376
EK	2.2	17.6	21.4	34.7	16.5	2.6	5.0	83,274
GE	1.1	11.4	5.8	65.1	13.1	1.9	1.6	146,362
GM	4.1	27.5	12.4	40.9	8.8	2.5	3.9	76,876
GT	1.1	7.6	3.5	58.1	23.0	6.0	0.7	27,966
IBM	0.3	13.4	25.1	54.4	5.8	0.8	0.1	280,101
IP	1.4	9.3	10.8	67.0	9.5	1.6	0.4	64,380
KO	1.0	8.2	18.5	41.0	26.4	1.9	3.0	72,698
MCD	2.8	6.0	15.5	60.7	10.3	4.4	0.2	44,525
MMM	1.1	22.8	10.3	57.0	6.4	2.2	0.2	97,426
MO	0.4	31.0	6.6	48.8	11.3	1.4	0.6	82,511
MRK	1.2	22.9	28.0	33.1	6.7	1.3	6.9	128,463
NAV	7.6	9.0	9.5	32.4	18.1	16.7	6.7	18,303
PA	0.8	2.5	11.4	64.2	16.6	2.5	1.9	34,372
PG	0.8	22.6	3.2	55.8	12.0	1.4	4.2	76,411
S	1.7	18.4	14.5	37.5	15.4	4.0	8.6	58,887
T	5.1	15.8	8.9	33.7	17.8	7.3	11.4	70,249
TX	0.6	2.2	14.3	48.0	26.1	2.4	6.4	50,699
UK	0.3	8.4	8.1	54.4	13.4	5.4	10.1	73,677
UTX	2.1	4.6	11.1	64.2	13.3	3.2	1.4	46,069
WX	1.4	1.7	12.0	64.8	16.1	1.6	2.3	61,736
X	1.6	14.6	12.8	42.0	17.6	4.0	7.3	73,471
XON	2.9	20.2	19.6	28.6	17.6	5.2	5.9	78,248
Z	0.6	3.8	2.4	53.6	17.6	2.6	19.5	66,997

Table B.2: Originating Exchanges for Quotes

Originating exchanges (percentages) for quotes for the sample of 30 stocks included in the Dow Jones Industrial Average in 1988. Abbreviations used: B – Boston Stock Exchange, C – Cincinnati Stock Exchange, M – Midwest Stock Exchange, N – New York Stock Exchange, P – Pacific Stock Exchange, T – National Association of Securities Dealers, and X – Philadelphia Stock Exchange. NQ is the total number of quotes reported on the ISSM tape.

	B	C	M	N	O	P	T	X	NT
AA	3.2	1.3	8.5	75.0	0.08	7.0	3.0	1.9	39,375
ALD	4.6	0.8	9.0	64.8	0.01	14.0	5.1	1.8	38,824
AXP	1.6	1.0	8.9	71.8	0.03	9.9	4.0	2.8	89,223
BA	2.2	1.0	8.4	72.0	0.08	6.5	6.1	3.8	67,761
BS	1.2	2.1	17.5	58.4	0.04	8.8	6.7	5.3	56,842
CHV	1.1	1.5	15.2	63.4	0.04	11.2	5.4	2.1	69,318
DD	6.2	3.0	8.8	71.7	0.05	4.4	2.7	3.1	78,484
EK	3.4	1.0	7.5	59.6	0.04	12.2	6.9	9.2	135,312
GE	1.9	0.9	10.0	59.4	0.04	13.1	8.9	5.6	181,587
GM	4.2	2.1	10.3	66.3	0.07	5.6	7.0	4.5	103,884
GT	1.4	1.2	8.7	66.9	0.06	17.3	2.8	1.6	29,800
IBM	4.2	1.3	9.3	69.0	0.08	7.4	6.1	2.5	239,282
IP	1.7	1.4	8.3	81.2	0.05	2.8	2.7	1.9	62,933
KO	3.8	1.3	10.3	63.8	0.05	5.4	5.6	9.8	70,445
MCD	5.2	1.4	13.6	69.0	0.03	4.3	4.4	2.1	58,460
MMM	1.7	3.1	12.9	72.9	0.06	4.3	3.8	1.2	70,475
MO	5.3	1.9	8.5	74.1	0.07	5.3	3.3	1.7	91,943
MRK	5.0	2.0	11.2	68.9	0.09	4.6	4.1	4.1	95,461
NAV	4.5	0.4	10.1	47.4	0.02	23.8	8.8	5.0	109,991
PA	1.2	0.7	7.6	78.1	0.02	4.2	5.6	2.6	39,618
PG	1.4	2.7	6.5	78.2	0.03	6.7	2.4	2.1	58,732
S	3.0	1.0	10.9	59.2	0.08	11.5	9.6	4.8	107,767
T	4.7	0.9	7.4	54.2	0.01	17.4	9.8	5.6	208,553
TX	0.9	0.4	11.9	57.8	0.02	19.8	5.2	3.9	150,168
UK	0.7	0.8	6.7	75.7	0.04	7.2	5.8	3.0	102,028
UTX	3.0	1.0	8.7	76.3	0.06	6.2	2.7	2.1	39,667
WX	1.6	0.9	8.7	72.9	0.03	7.9	4.0	3.9	53,955
X	2.0	1.5	7.5	69.0	0.02	6.5	5.5	8.1	79,187
XON	6.0	1.6	10.6	57.9	0.04	10.9	8.2	4.8	130,507
Z	1.1	1.3	6.0	75.2	0.02	8.2	3.5	4.7	49,568

Table B.3: Originating Exchanges for Trades

Originating exchanges (percentages) for trades for the sample of 30 stocks included in the Dow Jones Industrial Average in 1988. Abbreviations used: B – Boston Stock Exchange, C – Cincinnati Stock Exchange, M – Midwest Stock Exchange, N – New York Stock Exchange, O – Instinet, P – Pacific Stock Exchange, T – National Association of Securities Dealers, and X – Philadelphia Stock Exchange. NT is the total number of trades reported on the ISSM tape.

B.2.1 Quotes: Data Screening

The first screen is for data errors and missing data. An obvious requirement is that bid price, ask price, quoted depth at the bid, and quoted depth at the ask are all positive. Less-obvious data errors are corrected whenever discovered. For example, it appears that bid and ask prices in the ISSM database are not subjected to the error filters that are applied to transaction prices.[3]

The second screen is for certain unusual quotes. The ISSM database flags some trades and quotes with condition codes. Table B.4 gives a complete summary of the occurrence of all quote and trade condition codes for the stocks in my sample. A brief description of the various condition codes is contained in Table B.5—see the ISSM documentation for more details. As in several other studies dealing with ISSM transactions data, I screened out quotes flagged with quote condition codes C, D, F, G, I, L, N, P, S, V, X, and Z, i.e., all condition codes associated with "BBO-Ineligible" quotes.[4] Note from Table B.4 that quote condition codes referring to situations other than open and close, i.e., quote conditions 'C' (Closing), 'L' (Market Maker Quotes Closed, NASD), and 'O' (Opening), are relatively rare. Indeed, the only other BBO-ineligible quote conditions I encountered were quote conditions 'B' (Depth on the Bid Side: 1 case), 'D' (News Dissemination: 12 cases), 'I' (Order Imbalance: 67 cases), 'N' (Non-Firm: 1,632 cases), 'P' (News Pending: 15 cases), and 'Z' (No Open or No Resume: 1 case).

Since the focus in this part of the study is on the New York Stock Exchange, the third screen is for quotes that originated from exchanges other than the NYSE.

Finally, for the remaining NYSE quotes there are cases of identical time stamps among successive quotes. Typically, multiple quotes with the same time stamp are identical. There are also instances, however, that *different* successive quotes have identical time stamps. Since the NYSE specialist can only have one quote outstanding at any given

[3]As an illustration, the database reports NYSE bid and ask prices (in ticks) of 1188 and 1276 at time 13:36:26 on 21 January 1988 (CRSP day 6424) for Merck. From the quotes surrounding this suspect quote—bid and ask prices (in ticks) of 1184 and 1200 (6 seconds earlier) and 1188 and 1196 (10 seconds later)—it appears reasonable to assume that the correct ask price (in ticks) was 1196 rather than 1276. The discrepancy for the ask in this case would have been 80 ticks or exactly $10.

[4]These are quotes that are ineligible for inclusion in the National and NASD Best Bid & Offer calculations.

	Quote Condition Codes									Trade Condition Codes									
	B	C	D	I	L	N	O	P	Z	A	C	L	N	O	R	S	Z	?	total
AA		557			194		258				15	26	18		17		683	1	1,769
ALD		671			206		256			1	27	31	25		16		865	3	3,101
AXP		804	2		209		262				84	111	90		30		1,279	1	2,872
BA		556			208		260				44	14	28		29		1,077		2,216
BS		696			213		255	3		1	31	52	17	1	5		1,024	3	2,301
CHV		773		3	181		270				60	32	70		28		931	1	2,340
DD		725			181	20	254				44	71	22	4	20	1	1,507	4	2,933
EK		836		3	191	51	257				70	58	68	4	21	2	3,504	4	5,139
GE		911			187		262				213	70	108	11	44		3,470	3	5,279
GM		696		4	203	143	257				81	59	44	5	54	1	2,407	4	3,958
GT		582		8	212	3	254				9	41	14	6	5		410		1,544
IBM		856			187	166	255				251	73	207	7	79		5,759	5	7,845
IP		591			209	27	254				19	42	23	1	23		672		1,861
KO		844			190	2	358				62	31	74		19	1	1,417	2	3,000
MCD		667			232		254				42	16	34	2	14		905		2,166
MMM		671			191	6	254				37	49	53	5	25		965	1	2,257
MO		545		6	211		255				58	101	50	3	32		1,613		2,874
MRK		915		17	206	126	255				50	392	54	9	27		1,893	3	3,947
NAV		796			233	1	481				42	17	25	12	3		2,018	1	3,632
PA		589		3	201	5	250	5			22	14	18		13		362	1	1,483
PG		630			185	1	273				35	36	38	1	20		653	1	1,873
S		778	7	4	178	159	297				41	37	59	4	24		2,333	5	3,926
T	1	845			217	30	271				410	278	199	5	49	1	4,028	9	6,343
TX		662	3	10	184	817	277	7	1		68	30	45	2	18	1	5,053	2	7,170
UK		776		2	213	64	288				34	25	24	2	47		1,404		2,969
UTX		640			196		253				15	38	13	3	17		460	1	1,636
WX		695			194		263				69	31	3	4	10		820	1	2,090
X		868			212	11	321				31	32	24	1	21		1,141	2	2,664
XON		809			191		304				252	47	191	4	49	1	3,314	6	5,168
Z		694		7	192		279				21	18	17	2	8		738	2	2,008

Table B.4: Condition Code Occurrence

Counts of quote and trade (sale) condition codes for the sample of 30 Dow Jones Industrials in 1988. The meanings of these condition codes are summarized in Table B.5. The column labeled '?' applies to those condition codes that are missing (value '_') in the ISSM data files.

Code	Description
A	Depth on the Offer Side
B	Depth on the Bid Side
C	Closing
D	News Dissemination
F	Fast Trading
G	Pre-Opening Indication
I	Order Imbalance
L	Market Maker Quotes Closed (NASD)
N	Non-Firm
O	Opening
P	News Pending
S	Due to Related Security
V	In View of Common
X	Order Influx
Z	No Open Or No Resume
A	Acquisition (auction market is temporarily suspended while a buyer, with the specialist's assistance, acquires a large block of stock, NYSE Rule 392)
C	Cash Sale (same day clearing)
D	Distribution (similar to acquisition, only for a block being sold)
L	Sold Last (a trade in sequence but reported late)
N	Next Day (next day clearing)
O	Opened (delayed opening or opening out of sequence)
R	Seller (an unusual number of days may elapse before delivery of the stock)
S	Split Trade (not applicable to NYSE trades)
Z	Sold (reported out of sequence)

Table B.5: Condition Code Description

Description of quote condition codes (upper panel) and trade condition codes (lower panel).

time, the phenomenon of having identical time stamps should probably be attributed to the way the quotation data are collected and time-stamped. To avoid having zero interarrival times between quotes, I eliminated all but the last quote from any set of consecutive NYSE quotes having the same time stamp.

B.2.2 Quotes: Data Description

Table B.6 gives some summary statistics for all NYSE quotes that remain after carrying out all the screens described in the previous section. The spread distribution is given for spreads equal to 1 tick, 2 ticks, 3 ticks, 4 ticks, and 5 ticks or more, all expressed as a percentage of the total number of quotes, n_q. Note that with the exception of Merck (MRK) and Procter & Gamble (PG), at least 95 percent of all bid-ask spreads is 3 ticks or less.

Also given in Table B.6 are the average spread (in ticks) and the average depth on both the bid side and the ask side (in round lots of 100 shares).[5]

B.3 Trades

B.3.1 Trades: Data Screening

The first screen is again for data errors and missing data. For trades the requirement is that both price and volume are positive.

The second screen is for certain unusual trades marked by trade condition codes, see Table B.4 for a summary of the trade condition codes encountered for the sample of DJIA stocks and Table B.5 for a brief explanation of the various condition codes. Screened out are trade condition codes A, C, D, O, R, and Z.

The third screen is for the first and last NYSE transaction of every trading day. The reason for screening out these transactions is that the mechanism by which they are generated is fundamentally different from the mechanism governing the remaining

[5] Note that the average spread is remarkably similar across stocks, despite the fact that different stocks may be trading at quite different absolute price levels. A comparison of percentage spreads would reveal a much greater variability across stocks.

		ticks					mean	bid	ask
ticker	n_q	1	2	3	4	≥ 5	spread	depth	depth
AA	17,768	32.58	59.29	6.92	1.17	0.04	1.77	41.76	48.72
ALD	14,924	32.14	50.03	16.51	1.25	0.07	1.87	45.37	44.12
AXP	30,525	58.75	40.79	0.44	0.01	0.02	1.42	157.11	156.79
BA	27,805	43.04	50.86	5.94	0.12	0.03	1.63	75.83	85.44
BS	27,262	67.60	31.77	0.59	0.03	0.01	1.33	123.95	131.91
CHV	35,326	43.42	44.34	10.71	1.45	0.08	1.71	71.15	84.71
DD	50,486	30.32	45.73	19.18	4.19	0.57	1.99	33.37	37.81
EK	28,199	41.77	56.26	1.69	0.22	0.07	1.61	93.24	96.71
GE	94,219	70.37	28.73	0.80	0.08	0.01	1.31	150.29	183.04
GM	30,566	39.65	59.22	1.05	0.08	0.00	1.62	50.23	48.83
GT	15,729	23.61	45.19	27.32	3.77	0.11	2.12	48.47	49.60
IBM	150,506	63.39	33.16	3.12	0.29	0.04	1.40	60.23	75.94
IP	42,527	45.35	45.42	8.79	0.40	0.04	1.64	61.51	63.14
KO	29,262	46.62	47.58	5.72	0.07	0.01	1.59	99.96	109.96
MCD	26,398	41.51	49.58	8.60	0.30	0.02	1.68	66.16	72.05
MMM	54,744	39.49	40.04	16.48	3.61	0.39	1.86	46.15	52.84
MO	39,586	31.90	50.08	15.35	2.58	0.09	1.89	49.59	54.55
MRK	41,681	27.98	38.31	15.17	14.97	3.58	2.33	36.83	45.13
NAV	5,639	84.80	15.11	0.07	0.02	0.00	1.15	661.01	702.41
PA	21,616	54.38	37.48	7.89	0.24	0.02	1.54	60.93	61.76
PG	41,921	32.26	40.58	20.36	6.42	0.38	2.02	37.94	40.27
S	21,420	54.45	43.80	1.65	0.09	0.01	1.48	105.74	121.55
T	23,209	66.34	32.92	0.71	0.03	0.01	1.34	373.00	403.32
TX	23,151	55.22	44.13	0.50	0.10	0.06	1.46	132.78	139.38
UK	39,395	67.16	30.95	1.75	0.10	0.04	1.35	143.26	183.20
UTX	28,959	35.47	53.04	11.29	0.17	0.02	1.76	69.57	64.52
WX	39,340	32.91	48.14	14.84	3.95	0.16	1.91	43.18	42.97
X	30,292	58.65	37.49	3.60	0.20	0.06	1.46	122.04	131.44
XON	21,831	53.67	45.49	0.77	0.06	0.01	1.47	125.36	151.11
Z	35,278	42.96	50.97	5.83	0.20	0.05	1.64	43.62	48.13

Table B.6: Frequency Distribution Spreads

Frequency distribution of the spread for the sample of 30 Dow Jones Industrials over 1988 for 1 tick, 2 ticks, 3 ticks, 4 ticks, and 5 ticks or more. n_q is the number of spreads remaining after imposing the screens of section B.2.

transactions (a continuous auction market). For example, the opening transaction is a single call auction with all trades occurring at the same price.

The fourth screen is for trades in excess of 3,276,000 shares ('big' trades in the ISSM parlance), which may have been arranged away from the exchange floor. The number of these big trades is very small. Of the 30 stocks in my sample, there were only 10 such trades: ALD (1), IBM (2), NAV (1), S (2), TX (1), and UK (3).

The fifth screen is for all trades that originated from an exchange other than the NYSE, since the primary focus here is on NYSE-based trading. However, I will keep information on whether or not any trades occurred away from the NYSE.[6]

Finally, trades that do not have prior quotes are screened out. The reason for screening out those trades is that I want to distinguish between buyer-initiated and seller-initiated transactions, and the classification scheme used requires outstanding quotes, see section B.4.

B.3.2 Trades: Data Description

Table B.7 gives some summary statistics for the trades remaining after all the screening procedures of the previous section. The column labeled n_t gives the number of such NYSE trades, which ranges from 19,313 for Goodyear (GT) to 163,852 for IBM. Also given in Table B.7 are selected percentiles ξ_p of the observed dollar volume distributions. In addition to the quartiles $\xi_{0.25}$, $\xi_{0.50}$, and $\xi_{0.75}$, the 99.5th percentile $\xi_{0.995}$ is included. The last column of this table gives the mean dollar volumes for all stocks. It is clear by inspection that the dollar-volume distribution is highly skewed. In 23 out of 30 cases, the mean is even larger than the 75th percentile.

B.4 Buy-Sell Classification of Trades

It is well known that quotes are often recorded ahead of the trade that triggered them, see Lee and Ready (1991). They show that most of the mis-sequencing can be avoided

[6]See the variable OFFEXCH in subsection 4.3.2.

ticker	n_t	$\xi_{0.25}$	$\xi_{0.50}$	$\xi_{0.75}$	$\xi_{0.995}$	mean
AA	28,874	15,450	43,000	112,812	1,825,425	122,550
ALD	24,525	6,775	20,625	58,950	982,500	64,677
AXP	63,237	7,387	24,862	71,100	1,303,050	80,474
BA	48,068	9,750	29,812	97,000	1,922,800	123,001
BS	32,579	5,000	12,675	42,250	924,000	56,705
CHV	43,265	9,500	37,300	96,337	1,579,375	104,007
DD	55,499	17,475	69,700	160,000	1,498,500	141,110
EK	79,899	8,625	25,950	90,500	1,740,000	106,719
GE	106,895	8,875	32,700	106,500	1,338,750	107,723
GM	67,761	17,200	64,875	162,525	1,850,000	173,856
GT	19,313	17,100	47,750	135,937	2,017,800	147,292
IBM	163,852	24,525	95,500	226,000	1,925,000	190,407
IP	50,390	12,712	34,800	78,750	1,283,400	87,024
KO	44,259	11,475	38,000	104,287	1,819,125	110,483
MCD	39,718	9,275	33,500	91,750	1,424,100	101,766
MMM	50,692	17,775	57,625	122,750	1,529,600	119,716
MO	67,321	19,725	85,125	190,000	2,412,500	189,433
MRK	65,028	23,000	74,100	165,250	1,677,375	155,990
NAV	51,533	1,325	3,700	9,600	498,750	17,028
PA	30,390	5,650	12,937	30,750	744,250	48,146
PG	45,235	16,350	52,675	97,050	1,360,800	104,309
S	63,108	7,275	21,900	78,650	1,308,125	88,748
T	111,842	4,975	10,850	35,525	1,450,000	74,276
TX	85,970	9,375	28,050	99,250	2,250,000	127,563
UK	76,564	4,550	10,625	25,987	1,045,450	52,222
UTX	29,621	8,100	25,125	64,800	1,248,750	79,450
WX	38,587	10,550	32,025	88,187	1,047,500	84,488
X	53,929	6,350	25,000	63,000	1,174,750	75,672
XON	74,703	9,125	32,287	116,200	2,068,750	138,696
Z	36,606	16,875	45,625	107,000	1,339,875	104,149

Table B.7: Dollar-Volume Statistics for Individual Trades

Quantiles and means of the dollar-volume distribution for individual trades. ξ_p is the pth quantile of the dollar-volume distribution. n_t is the number of trades remaining after imposing the screens of section B.3.

by requiring a time separation of at least 5 seconds between quotes and trades. In other words, any quote that is matched to a trade (for reasons described below) is required to be at least 5 seconds old, as determined from the time stamps.

The reason we match trade and quote data is that we want to classify trades as either "buyer-initiated" or "seller-initiated" to see whether buy transactions and sell transactions have different effects on the bid-ask spread. For example, Allen and Gorton (1991) suggest that the information content for buys and sells is different.

If all trades occur either at the bid price or at the ask price, it is straightforward to do this buy-sell classification: a trade at the ask will be considered a buy and a trade at the bid will be considered a sell. In practice, however, trades can also be outside the quotes or inside the quotes. Indeed, Lee and Ready (1991) show that as many as 30% of NYSE trades take place within the quotes.

To classify all trades, I adopt the same algorithm as used by Blume, MacKinlay, and Terker (1989). This algorithm defines a trade to be *indeterminate* or *neutral* if it occurs at the midpoint of the bid-ask spread. To classify a transaction with price P_k whose matched quotes are B_k and A_k (see above), the buy-sell indicator IBS_k is defined by

$$\mathrm{IBS}_k = \begin{cases} 1 & \text{if } P_k > \frac{1}{2}(B_k + A_k) \\ 0 & \text{if } P_k = \frac{1}{2}(B_k + A_k) \\ -1 & \text{if } P_k < \frac{1}{2}(B_k + A_k). \end{cases}$$

It is possible to classify all trades as either a buy ($\mathrm{IBS}_k = 1$) or a sell ($\mathrm{IBS}_k = -1$) by using the "tick test," see Lee and Ready (1991), but this rule necessarily involves a certain arbitrariness. It appears reasonable to treat trades at the midquote different from trades that are closer to either the bid or the ask. Incidentally, the classification of buys and sells is somewhat delicate, since there is obviously a buyer and a seller for every transaction. The distinction is made to distinguish whether the buyer or seller was more eager to execute the trade and, therefore, willing to pay (part of) the bid-ask spread.

Table B.8 gives the proportions of trades classified as buy, sell, or indeterminate for the sample of 30 stocks. For almost all stocks, the fraction of buys is roughly the same as the fraction of sells. With the exception of outlier Navistar (NAV), which trades at a low price and whose bid-ask spread is typically 1 tick (see Table B.6), the fraction of trades classified as indeterminate is generally between 20 and 30 percent.

ticker	n_t	buy (%)	sell (%)	neutral (%)
AA	28,874	36.70	36.15	27.16
ALD	24,525	32.20	39.44	28.36
AXP	63,237	32.90	43.54	23.55
BA	48,068	34.48	40.59	24.93
BS	32,579	41.95	38.63	19.42
CHV	43,265	39.24	39.99	20.78
DD	55,499	43.83	36.33	19.85
EK	79,899	32.11	26.75	41.15
GE	106,895	40.23	39.54	20.23
GM	67,761	30.31	32.39	37.30
GT	19,313	39.88	38.05	22.07
IBM	163,852	39.83	42.18	17.99
IP	50,390	43.53	36.60	19.87
KO	44,259	29.23	35.76	35.01
MCD	39,718	41.61	39.30	19.09
MMM	50,692	43.61	36.45	19.94
MO	67,321	37.75	34.19	28.05
MRK	65,028	43.76	32.63	23.61
NAV	51,533	52.16	43.82	4.01
PA	30,390	38.10	42.16	19.73
PG	45,235	44.40	40.47	15.13
S	63,108	34.58	35.56	29.86
T	111,842	28.75	49.74	21.52
TX	85,970	37.40	39.39	23.21
UK	76,564	44.66	37.85	17.49
UTX	29,621	33.37	39.10	27.52
WX	38,587	34.82	36.15	29.03
X	53,929	40.46	40.62	18.92
XON	74,703	22.79	47.03	30.18
Z	36,606	36.13	36.31	27.56

Table B.8: Classification of Trades (Buy, Sell, or Neutral)

Fraction of trades classified as a buy ($\text{IBS}_k = 1$), sell ($\text{IBS}_k = -1$), and neutral ($\text{IBS}_k = 0$) for the sample of 30 Dow Jones Industrials over 1988 using the rule

$$\text{IBS}_k = \begin{cases} 1 & \text{if } P_k > \frac{1}{2}(B_k + A_k) \\ 0 & \text{if } P_k = \frac{1}{2}(B_k + A_k) \\ -1 & \text{if } P_k < \frac{1}{2}(B_k + A_k), \end{cases}$$

where P_k is the transaction price of the kth trade and A_k and B_k are the ask and the bid matched to the kth trade, respectively. n_t is the total number of trades that were thus classified.

Appendix C

Closing Quotes

As noted in the Introduction, several studies have examined the behavior of bid-ask spreads around stock splits. Copeland (1979) and Conroy, Harris, and Benet (1990) find that post-split percentage bid-ask spreads increase significantly, whereas Murray (1985) finds that stock splits do not have an effect on the percentage bid-ask spread. Copeland (1979) and Murray (1985) use representative bid-ask quotes in their studies, whereas Conroy et al. (1990) use closing bid-ask quotes, an improvement over the earlier studies. I use the complete record of all intraday bid-ask quotes to investigate this issue.

There are several reasons why using the entire record of all spreads is more informative than using just closing spreads. First, closing bid-ask quotes are not firm, since no trades will occur at these quotes. Second, even if closing spreads accurately reflect trading opportunities at the end of the trading day, the spread at the end of the day only gives partial information about the "typical" bid-ask spread of a stock. Indeed, McInish and Wood (1992) have shown that bid-ask spreads exhibit a U-shape pattern over the course of the trading day.

This appendix explores in detail how significant the first issue is by making a comparison between spreads derived from the last intraday quotes and closing spreads. The results indicate that closing spreads are on average narrower than spreads quoted at the very end of the trading day. Moreover, in cases where there are multiple closing spreads for a security according to the ISSM database, "later" closing spreads are on average narrower than "earlier" closing spreads. Thus my findings for spreads complement the findings by

Harris (1989) that closing transaction prices may not consistently represent stock values.

Closing quotes on the ISSM tapes are marked with quote condition 'C.'[1] These quotes were previously screened out and they were therefore not part of my earlier analyses. Selecting those quotes corresponding to NYSE quotes, I found that there may be more than one closing NYSE quote for a given stock on a given day. Moreover, in case of multiple quotes, these quotes may or may not be identical.[2] For example, the ISSM database reports two closing NYSE quotes for Van Dorn Co (ticker symbol VDC) on the first trading day of 1988 (January 4). The first one carries the timestamp 16:01:28 and the bid and ask are $\$33\frac{5}{8}$ and \$34, respectively. The second one is timestamped 16:03:16 and its bid and ask are $\$33\frac{3}{4}$ and \$34, respectively. Thus the first reported closing spread is 3 ticks (\$0.375) and the last reported closing spread is 2 ticks (\$0.25).

Given this ambiguity about the "true" closing spread, there are at least four ways to define "the" closing bid-ask spread from the ISSM database: (1) the *first* reported closing spread, (2) the *last* reported closing spread, (3) the *minimum* of the reported closing spreads, and (4) the *maximum* of the reported closing spreads.

I have calculated these four quantities for the two samples of stocks that split in 1988 and 1991. To make the comparison with intraday spreads, I actually calculated those quantities in percentage terms. The results are reported in Tables C.1 and C.2 for 1988 and 1991, respectively. From these tables, we see that for the overwhelming majority of the stocks considered, earlier closing spreads tend to be higher than later closing spreads. Indeed, based on averages over all trading days, this condition holds for all 26 stocks of the 1988 sample and for 42 out of 46 stocks of the 1991 sample.

To determine whether closing spreads closely mirror spreads prevailing at the end of the trading day, I retrieved the last intraday quote from the ISSM database for each day and each stock in my sample. Summary statistics on the corresponding spreads are reported in Tables C.3 and C.4. A comparison of Tables C.3 and C.4 suggests that final intraday spreads are on average higher than closing spreads—strongly so when "last"

[1]The ISSM documentation of June 1992 describes this quote condition as follows: "C – Closing: This condition will be indicated upon dissemination of the last quotation from a Participant for that security during the trading day or at the market price."

[2]I only looked at the quoted bid and ask prices and not at the quoted depth at these prices.

closing spreads are considered as "the" closing spreads. For example, the final closing spread is on average 8.9% smaller than the last intraday spread for the 1988 sample, whereas the corresponding decline for the 1991 sample is 4.8%.

To investigate this in more detail, I carry out formal tests of the null hypothesis that for a given stock the final intraday spread is on average equal to the closing spread—the "first" closing spread in the first test and the "last" closing spread in the second test. For each stock, I calculate the series of daily differences between the final intraday spread and the closing spread.[3] The test statistic q defined as the average difference divided by its standard deviation has an asymptotic standard normal distribution. The values of q_1 (closing spread is "first" closing spread) and q_2 (closing spread is "last" closing spread) are reported in Tables C.5 and C.6. To summarize these tables, I counted the number of times that the values of q_1 and q_2 are significant at the 95 percent and the 99 percent significance levels (corresponding to the standard normal quantiles $z_{0.975} = 1.960$ and $z_{0.995} = 2.576$). The results are as follows:

Sample Period	α	$q_1 > z_{1-\alpha/2}$	$q_1 < -z_{1-\alpha/2}$	$q_2 > z_{1-\alpha/2}$	$q_2 < -z_{1-\alpha/2}$
1988 (26 stocks)	0.05	17	1	25	0
	0.01	12	0	25	0
1991 (46 stocks)	0.05	14	3	22	2
	0.01	12	2	20	0

Thus at the 1 percent significance level, the last closing spread is *smaller* than the immediately preceding intraday spread for 25 out of 26 stocks for the 1988 sample and for 20 out of 46 stocks for the 1991 sample. Moreover, for none of the stocks in the two samples is the last closing spread significantly *larger* than the immediately preceding intraday spread at the 1 percent level. The results are slightly less impressive for the comparison of the first closing spread versus the immediately preceding intraday spread. All in all, these findings confirm that the use of closing data may be misleading. There appears to be a systematic downward bias in the reported closing spread. This complements earlier

[3]If on a particular day there is no observation available for one or both of these two quantities, that day is excluded from the analysis.

empirical findings by Harris (1989) that closing transaction prices may not consistently represent stock values.

Ticker	Nobs	First	Last	Min	Max
ASH	253	0.7321	0.6823	0.6718	0.7447
BCR	253	0.7467	0.7119	0.7070	0.7526
BMS	253	1.0954	1.0780	1.0634	1.1086
CCK	253	0.4566	0.4282	0.4277	0.4571
CTB	253	1.0315	1.0123	1.0078	1.0360
DCI	252	1.3919	1.3759	1.3759	1.3919
DOV	253	0.7567	0.7393	0.7351	0.7609
F	253	0.3976	0.3705	0.3638	0.4052
FBO	253	0.9841	0.9728	0.9703	0.9866
FCB	253	1.2936	1.2802	1.2802	1.2936
GW	253	0.3727	0.3480	0.3466	0.3787
HLT	253	0.5173	0.4977	0.4971	0.5179
IMD	253	1.2105	1.1880	1.1880	1.2105
MRK	253	0.3387	0.3081	0.2971	0.3517
NHY	253	0.8046	0.7821	0.7783	0.8094
NWL	252	0.7783	0.7616	0.7605	0.7795
PC	253	0.9034	0.8840	0.8840	0.9035
SBL	235	0.9100	0.8950	0.8878	0.9172
SC	253	0.4090	0.4018	0.4011	0.4097
SDW	253	1.1239	1.0949	1.0934	1.1239
SPP	253	0.6801	0.6614	0.6560	0.6879
STH	253	1.2143	1.1990	1.1977	1.2156
TKR	253	0.6838	0.5983	0.5983	0.6838
UB	253	1.2934	1.2735	1.2705	1.2974
VDC	253	1.2619	1.2420	1.2405	1.2633
WWY	253	0.7142	0.5947	0.5919	0.7238

Table C.1: Average Quoted Closing Spreads (1988)

Nobs is the number of trading days with closing quotes. First, Last, Min, and Max refer to the algorithm used to define the closing quote for a given day. For "First" ("Last") the first (last) closing quote is used. For "Min" ("Max") the minimum (maximum) closing quote is used. All candidate closing quotes are NYSE quotes marked with the quote condition 'C' in the ISSM database. Missing data: DCI (6634), NWL (6415), and SBL (6411–6428).

Ticker	Nobs	First	Last	Min	Max
ASC	250	0.5762	0.5534	0.5465	0.5832
AUD	250	0.6669	0.6528	0.6126	0.7183
BOL	250	0.5493	0.5214	0.5132	0.5581
BV	250	1.2102	1.1828	1.1743	1.2187
CIR	250	0.5482	0.5207	0.5166	0.5537
CL	250	0.4021	0.4000	0.3961	0.4067
CML	249	1.0373	1.0153	1.0153	1.0373
CNC	250	0.6425	0.5655	0.5598	0.6482
CPB	250	0.3674	0.3523	0.3433	0.3790
GLX	249	0.4287	0.4161	0.4119	0.4354
GPS	250	0.4961	0.4358	0.4276	0.5081
GPU	250	0.6260	0.6247	0.6188	0.6319
GS	176	0.3914	0.3772	0.3727	0.3976
GWW	250	0.4668	0.4437	0.4390	0.4697
HF	249	0.8198	0.8104	0.8072	0.8427
HKF	250	0.9771	0.9784	0.9522	1.0132
HRB	250	0.5956	0.5747	0.5587	0.6120
IGT	228	0.7882	0.7474	0.7401	0.8038
IMA	250	0.4734	0.4597	0.4508	0.4855
JEC	250	1.1070	1.0642	1.0609	1.1090
K	250	0.3332	0.2788	0.2718	0.3529
LC	250	0.8279	0.8291	0.8244	0.8326
LDL	249	1.1445	1.1345	1.1230	1.1586
LGN	250	1.2915	1.2906	1.2866	1.2954
LOC	250	0.6476	0.6413	0.6401	0.6500
MDT	250	0.3876	0.3641	0.3480	0.4083
MS	250	0.4436	0.4218	0.4162	0.4486
NLC	250	0.6324	0.5979	0.5807	0.6545
NME	250	0.6199	0.6012	0.5852	0.6496
PFE	248	0.3728	0.3516	0.3316	0.4007
PVH	248	1.3327	1.2858	1.2672	1.3513
RAD	249	0.6760	0.6555	0.6398	0.6917
RAL	249	0.3793	0.3683	0.3623	0.3853
RBD	249	0.4658	0.4445	0.4318	0.4836
RPR	249	0.4223	0.4178	0.4122	0.4295
SCR	249	0.9893	0.9753	0.9700	0.9979
SHW	249	0.8524	0.8265	0.8170	0.8617
SRR	250	0.6415	0.6219	0.6194	0.6489
STR	249	0.9917	0.9964	0.9761	1.0168
SYN	250	0.4193	0.4167	0.3929	0.4466
UNP	250	0.3595	0.3488	0.3454	0.3647
USS	250	0.3901	0.3322	0.3245	0.4006
VCD	140	1.1508	1.1601	1.1439	1.1707
VNO	245	1.4286	1.4055	1.4054	1.4286
WAG	250	0.5546	0.5351	0.5295	0.5602
WTI	250	0.9800	0.9535	0.9506	0.9865

Table C.2: Average Quoted Closing Spreads (1991)

Ticker	Nobs	Spread
ASH	253	0.7466
BCR	253	0.7877
BMS	253	1.1875
CCK	253	0.5395
CTB	253	1.0694
DCI	252	1.4153
DOV	253	0.7742
F	253	0.4135
FBO	253	1.0343
FCB	253	1.2973
GW	253	0.4828
HLT	253	0.5184
IMD	253	1.2279
MRK	253	0.3634
NHY	253	0.7650
NWL	253	0.8783
PC	253	0.9454
SBL	235	0.9259
SC	253	0.4550
SDW	253	1.1636
SPP	253	0.6959
STH	253	1.2505
TKR	253	0.7656
UB	253	1.3203
VDC	253	1.2747
WWY	253	0.8160

Table C.3: Average Final Intraday Spreads (1988)

Average percentage spread of the last intraday quotes for the stocks included in the 1988 sample. Missing data for the following stocks and CRSP days: DCI (6634) and SBL (6411 through 6428).

Ticker	Nobs	Spread	Ticker	Nobs	Spread
ASC	251	0.5825	LGN	251	1.3040
AUD	251	0.6585	LOC	251	0.6568
BOL	251	0.5347	MDT	251	0.3738
BV	251	1.1439	MS	251	0.4634
CIR	251	0.5529	NLC	251	0.6788
CL	251	0.3913	NME	251	0.6357
CML	251	1.0653	PFE	250	0.3344
CNC	251	0.7314	PVH	250	1.3693
CPB	251	0.3653	RAD	250	0.7040
GLX	251	0.3927	RAL	250	0.3767
GPS	251	0.5094	RBD	250	0.4585
GPU	251	0.6099	RPR	250	0.4140
GS	177	0.3908	SCR	250	1.0418
GWW	251	0.5094	SHW	250	0.8887
HF	251	0.8278	SRR	251	0.6439
HKF	251	0.9840	STR	251	0.9943
HRB	251	0.5827	SYN	251	0.4276
IGT	230	0.8388	UNP	251	0.3516
IMA	251	0.4726	USS	251	0.4734
JEC	251	1.1654	VCD	141	1.1650
K	251	0.3398	VNO	251	1.4211
LC	251	0.8167	WAG	251	0.5984
LDL	251	1.1836	WTI	251	1.0004

Table C.4: Average Final Intraday Spreads (1991)

Average percentage spread of the last intraday quotes for the stocks included in the 1991 sample. For a discussion of missing data, see Appendix D.

Ticker	Nobs	q_1	q_2
ASH	253	2.10	5.38
BCR	253	2.37	4.58
BMS	253	4.84	5.63
CCK	253	9.40	11.58
CTB	253	3.03	4.20
DCI	252	1.97	3.07
DOV	253	2.26	3.37
F	253	1.34	3.28
FBO	253	3.05	3.66
FCB	253	1.37	2.99
GW	253	8.22	11.35
HLT	253	0.15	3.14
IMD	253	1.77	3.48
MRK	253	1.74	4.42
NHY	253	−2.37	−1.63
NWL	252	7.93	8.54
PC	253	3.15	4.37
SBL	235	1.97	2.62
SC	253	5.59	6.19
SDW	253	3.05	4.75
SPP	253	1.64	3.19
STH	253	2.88	3.49
TKR	253	5.50	10.90
UB	253	1.75	2.89
VDC	253	1.45	2.63
WWY	253	5.35	11.56

Table C.5: Comparison of Last Intraday Spread and Closing Spread (1988)

Results of the test whether the last intraday spread is different from the first (last) closing spread for the stocks included in the 1988 sample. For each stock, the series of daily differences between the final intraday spread and the closing spread is constructed. The test statistic q defined as the average difference divided by its standard deviation has an asymptotic standard normal distribution. The values of q_1 (closing spread is "first" closing spread) and q_2 (closing spread is "last" closing spread) are reported above. Nobs is the number of pairwise observations for each stock.

Notes: (1) If on a particular day there is no observation available for one or both of the spreads, that day is excluded from the analysis; (2) All spreads are calculated in percentage terms.

Ticker	Nobs	q_1	q_2	Ticker	Nobs	q_1	q_2
ASC	250	0.72	2.80	LGN	250	1.21	1.29
AUD	250	−0.72	0.31	LOC	250	1.28	2.05
BOL	250	−1.44	1.07	MDT	250	−1.29	1.00
BV	250	−2.47	−1.63	MS	250	2.82	5.16
CIR	250	0.61	3.01	NLC	250	3.47	5.79
CL	250	−1.47	−1.08	NME	250	1.22	2.23
CML	249	2.89	4.26	PFE	248	−4.81	−2.02
CNC	250	6.48	11.56	PVH	248	1.98	3.93
CPB	250	−0.26	1.68	RAD	249	2.54	3.43
GLX	249	−3.72	−2.29	RAL	249	−0.30	1.56
GPS	250	1.50	7.69	RBD	249	−0.73	1.37
GPU	250	−1.78	−1.42	RPR	249	−1.55	−0.72
GS	176	−0.28	1.12	SCR	249	3.16	3.65
GWW	250	5.07	7.41	SHW	249	2.94	4.30
HF	249	0.58	1.18	SRR	250	0.24	2.95
HKF	250	0.69	0.35	STR	249	0.48	0.04
HRB	250	−0.94	0.54	SYN	250	0.71	0.94
IGT	228	4.03	5.91	UNP	250	−1.28	0.38
IMA	250	−0.20	1.19	USS	250	7.03	11.60
JEC	250	4.03	6.43	VCD	140	0.83	0.42
K	250	0.78	7.33	VNO	245	−1.04	1.41
LC	250	−1.80	−1.77	WAG	250	3.83	5.55
LDL	249	3.63	3.83	WTI	250	1.37	3.60

Table C.6: Comparison of Last Intraday Spread and Closing Spread (1991)

Results of the test whether the last intraday spread is different from the first (last) closing spread for the stocks included in the 1991 sample. For each stock, the series of daily differences between the final intraday spread and the closing spread is constructed. The test statistic q defined as the average difference divided by its standard deviation has an asymptotic standard normal distribution. The values of q_1 (closing spread is "first" closing spread) and q_2 (closing spread is "last" closing spread) are reported above. Nobs is the number of pairwise observations for each stock.

Notes: (1) If on a particular day there is no observation available for one or both of the spreads, that day is excluded from the analysis (see Appendix D for a discussion of missing data); (2) All spreads are calculated in percentage terms.

Appendix D

Missing 1991 Data

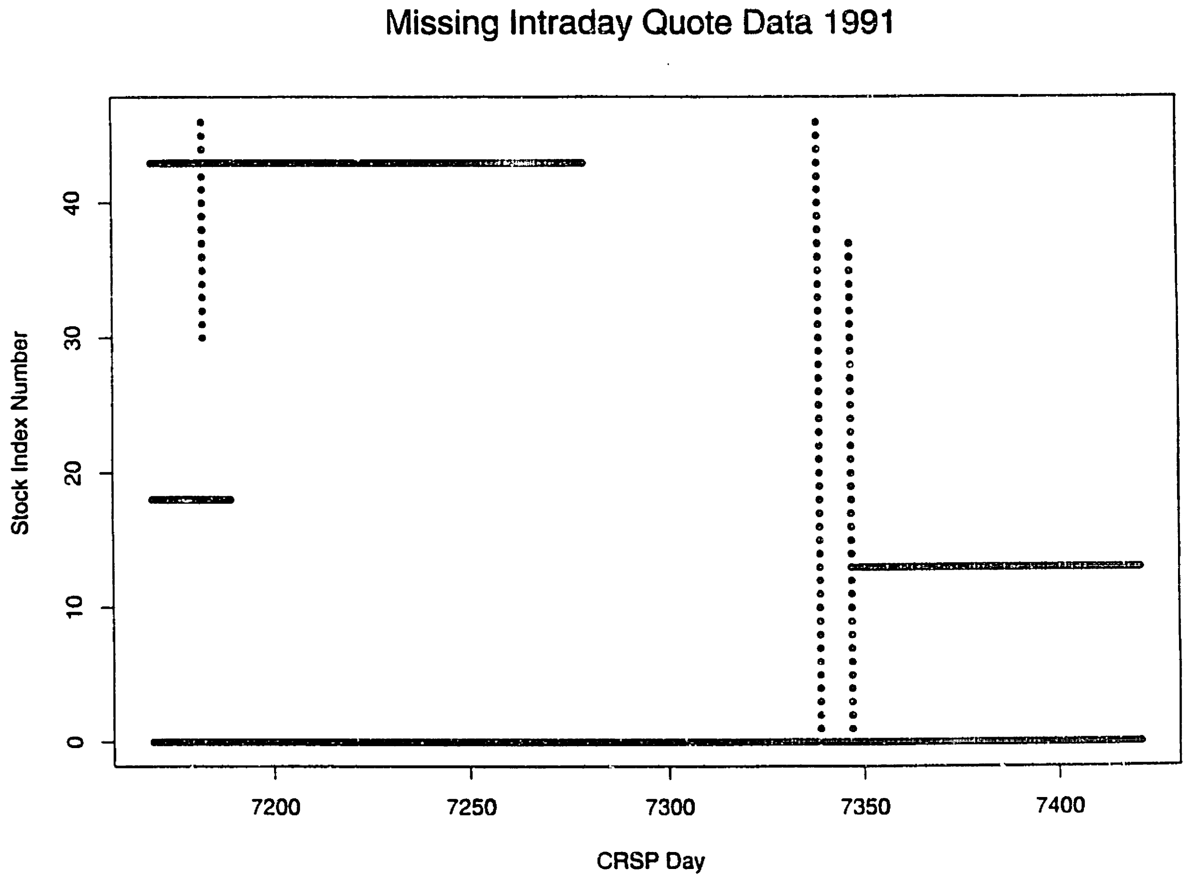
It is perhaps not surprising that the ISSM database for 1991 is not fully complete. However, it turns out that non-trivial amounts of data are unavailable. In this appendix, I discuss the extent of missing data focusing on (1) missing intraday quote data, (2) missing closing quote data, and (3) missing transactions data for the year 1991.

D.1 Missing Intraday Quote Data (1991)

Figure D.1 shows the extent of missing data on intraday quotes for my sample of 46 stocks. In this graph, each stock is represented by its index (1 to 46, see Table 5.2) on the vertical axis. Each stock has one or more days with missing data, indicated by dots in the graph. The year 1991 had 253 trading days, indexed from CRSP day 7169 (January 2, 1991) to CRSP day 7421 (December 31, 1991) on the horizontal axis.

As can be seen from Figure D.1, no quote information is available for any of my sample stocks on CRSP day 7339 (September 4, 1991). In addition, CRSP day 7347 (September 16, 1991) has data missing for stocks 1 through 37 and CRSP day 7182 (January 21, 1991) has data missing for stocks 30 through 46. Several stocks have large segments of missing data: stock 13 (GS) for 75 days (CRSP days 7347–7421), stock 18 (IGT) for 21 days (CRSP days 7169–7189), and stock 43 (VCD) for 111 days (CRSP days 7169–7279).

The absence of the three trading days 7182, 7339, and 7347 poses problems for a cross-sectional analysis of quote characteristics (spreads and depths) once the various stocks



In this graph, each stock included in the 1991 sample is represented by its index (1 to 46, see Table 5.2) on the vertical axis. Each stock has one or more days with missing intraday quote data, indicated by dots in the graph. The year 1991 had 253 trading days, indexed from CRSP day 7169 (January 2, 1991) to CRSP day 7421 (December 31, 1991) on the horizontal axis.

Figure D.1: Missing Intraday Quote Data 1991

are aligned by their split date. The split dates of the 46 stocks are plotted in Figure D.2. If stocks are aligned such that date 0 corresponds to the split date (the first day of the post-split period), Figure D.3 shows that any reasonable time window around the split date will contain days for which data are unavailable for at least one of the stocks.

To avoid this problem, I deleted missing days from the time series of a given stock. For example, the exdate of Lydall Inc (LDL, stock 23) is 7348 and data for days 7339 and 7347 are missing for this stock. Relative to the split date, therefore, days −9 and −1 are missing. Dropping days with missing data and shifting days over to eliminate gaps in the 10-day pre-split period results in the following sequence:

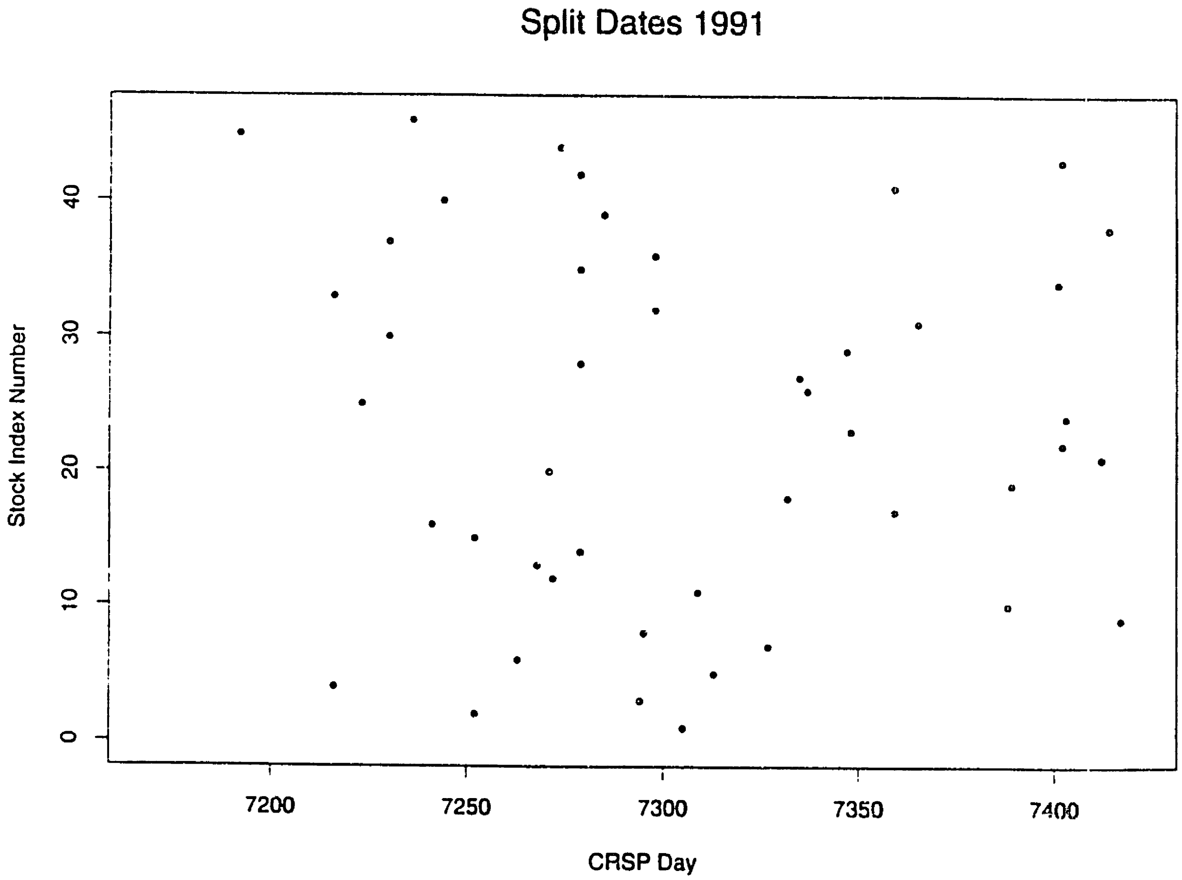
	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1
Raw Data	7338	NA	7340	7341	7342	7343	7344	7345	7346	NA
Shifted Data	7336	7337	7338	7340	7341	7342	7343	7344	7345	7346

The same procedure is applied for the post-split period: if we are interested in the K-day post-split period, the first K post-split days are selected for which data are available. Thus a cross-sectional analysis can be carried out in the usual way.

D.2 Missing Closing Quote Data (1991)

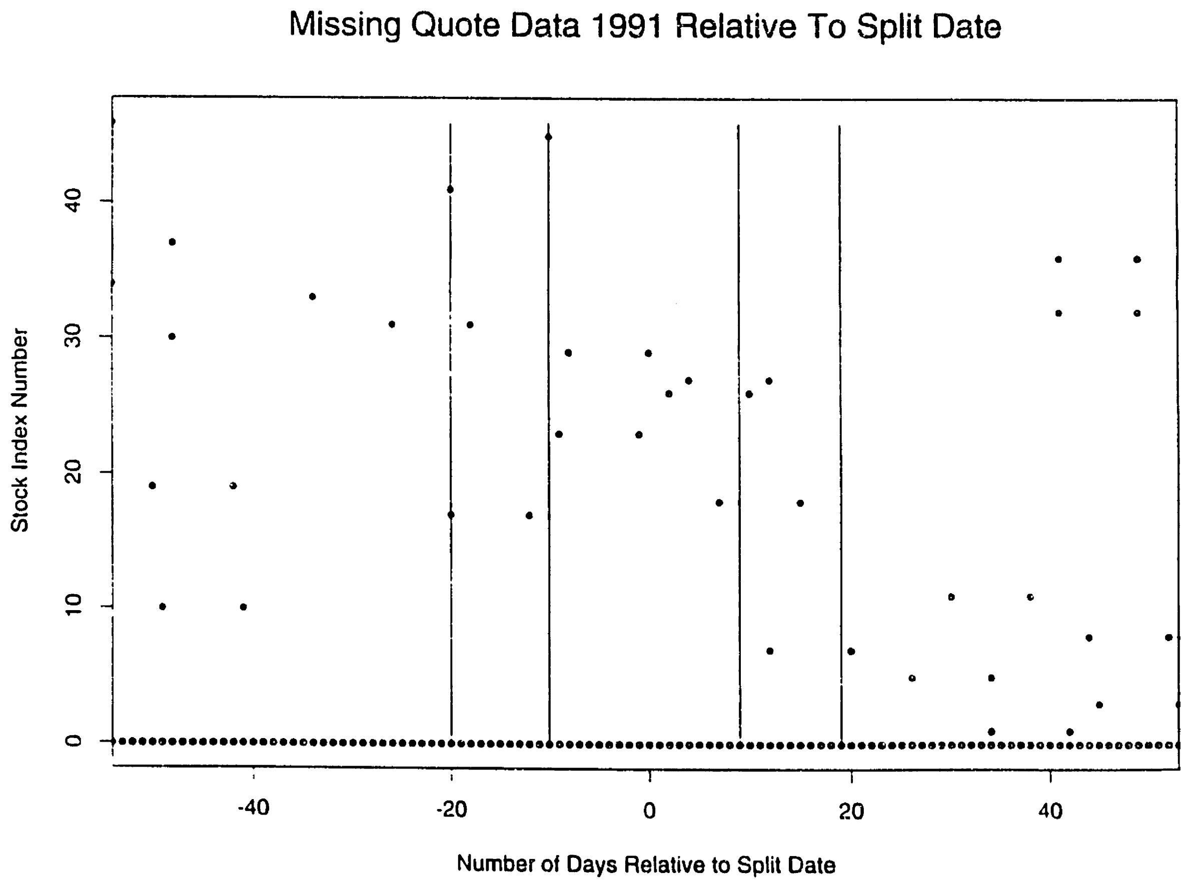
In this section I report on the absence of closing quotes for the 46 stocks in the 1991 sample. Recall that a closing quote is a quote marked with the quote condition 'C' on the ISSM tapes. Contrary to intraday quotes, specialists are not required to transact at the closing quotes.

The days on which no closing quotes are available are summarized in Figure D.4. A comparison of Figures D.1 and D.4 reveals that there are more missing closing quotes than missing intraday quotes. It is not surprising that the absence of intraday quotes implies an absence of closing quotes. However, it turns out that there are a few days on which there were intraday quotes but no closing quotes for some of the stocks. For example, it turns out that the ISSM database reports closing quotes for none of the 46 stocks on CRSP day 7295 (July 2, 1991), although all stocks have intraday quotes on that day.



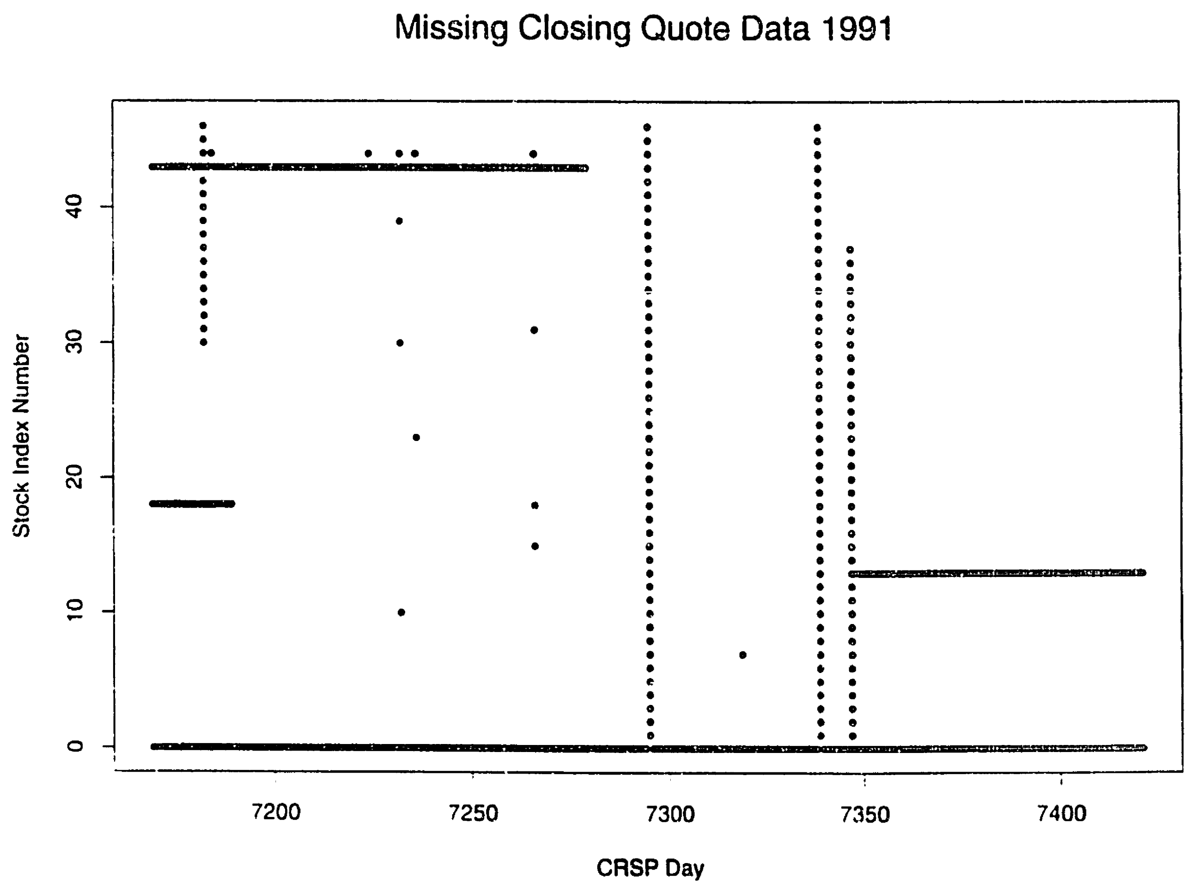
In this graph, each stock included in the 1991 sample is represented by its index (1 to 46, see Table 5.2) on the vertical axis. The split dates for all stocks are represented by dots in the graph. The year 1991 had 253 trading days, indexed from CRSP day 7169 (January 2, 1991) to CRSP day 7421 (December 31, 1991) on the horizontal axis.

Figure D.2: Split Dates 1991



In this graph, each stock included in the 1991 sample is represented by its index (1 to 46, see Table 5.2) on the vertical axis. All stocks are aligned such that date 0 corresponds to the split date (the first day of the post-split period). This graph shows that any reasonable time window around the split date will contain days for which data are unavailable for at least one of the stocks.

Figure D.3: Missing Quote Data Relative to Split Date



In this graph, each stock included in the 1991 sample is represented by its index (1 to 46, see Table 5.2) on the vertical axis. Each stock has one or more days with missing closing quote data, indicated by dots in the graph. The year 1991 had 253 trading days, indexed from CRSP day 7169 (January 2, 1991) to CRSP day 7421 (December 31, 1991) on the horizontal axis.

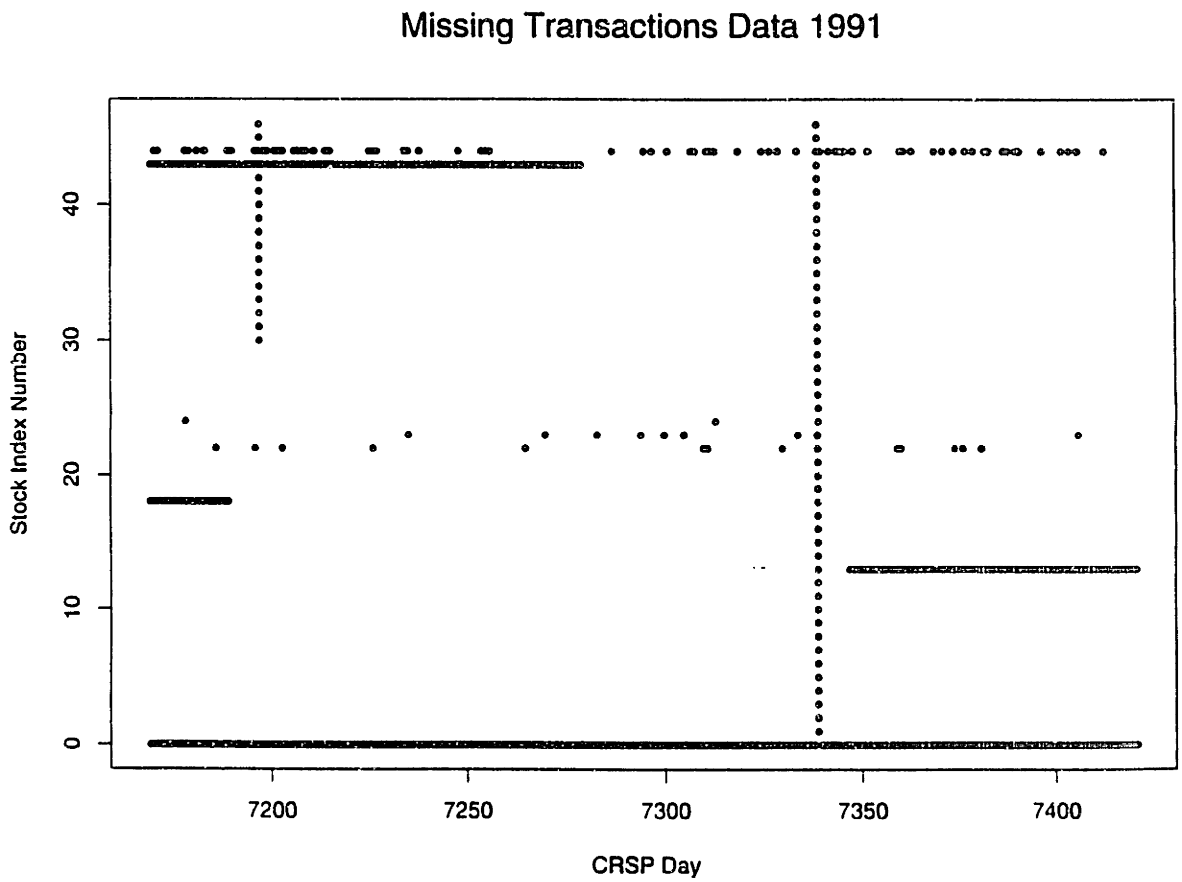
Figure D.4: Missing Closing Quote Data 1991

D.3 Missing Transactions Data (1991)

Figure D.5 shows the extent of trading days in 1991 on which there were no transactions reported on the ISSM tapes for the stocks in my sample. No transactions are reported for any of the stocks on CRSP day 7339 (September 4, 1991) and no transactions are reported for stocks indexed 30 (PFE) through 46 (WTI) (see Table 5.2) on CRSP day 7197 (February 11, 1991).

Also, there are some stocks for which there are many days with no trades. International Game Technology (IGT, index 18) and Value City Dept Stores Inc (VCD, index 43) have no transactions activity at the beginning of the year and Gillette (GS, index 13) has no transactions at the end of the year.[1] Some of the other stocks appear to be very thinly traded. Liberty Corp (LC, index 22) has no trades reported on 14 days, Lydall Inc (LDL, index 23) has no trades on 9 days, and Vornado Incorporated (VNO, index 44) has no trades reported on as many as 73 days.

[1] Given the absence of quotes these stocks were probably not traded at all during those periods.



In this graph, each stock included in the 1991 sample is represented by its index (1 to 46, see Table 5.2) on the vertical axis. Each stock has one or more days with missing transactions data, indicated by dots in the graph. The year 1991 had 253 trading days, indexed from CRSP day 7169 (January 2, 1991) to CRSP day 7421 (December 31, 1991) on the horizontal axis.

Figure D.5: Missing Transactions Data 1991

Bibliography

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

[2] Allen, F. and G. Gorton (1991), "Stock Price Manipulation, Market Microstructure and Asymmetric Information," Working Paper No. 3862, National Bureau of Economic Research.

[3] Amihud, Y. and H. Mendelson (1980), "Dealership Market: Market-Making with Inventory," *Journal of Financial Economics*, 8 (March): 31–53.

[4] Anderson, T. and L. Goodman (1957), "Statistical Inference About Markov Chains," *Annals of Mathematical Statistics*, 28: 89–110.

[5] Bagehot, W. (1971), "The Only Game in Town," *Financial Analysts Journal*, 27 (March/April): 12 ff.

[6] Baker, H. and P. Gallagher (1980), "Management's View of Stock Splits," *Financial Management*, 9 (Summer): 73–77.

[7] Bernstein, P. (1987), "Liquidity, Stock Markets, and Market Makers," *Financial Management*, 16 (Summer): 54–62.

[8] Black, F. (1971), "Towards a Fully Automated Exchange," *Financial Analysts Journal*, 27 (July/August): 29 ff.

[9] Blume, M., A. MacKinlay, and B. Terker (1989), "Order Imbalances and Stock Price Movements on October 19 and 20, 1987," *Journal of Finance*, 44 (September): 827–848.

[10] Brennan, M. and T. Copeland (1988), "Stock Splits, Stock Prices, and Transaction Costs," *Journal of Financial Economics*. 22 (October): 83–101.

[11] Brock, W. and A. Kleidon (1992), "Periodic Market Closure and Trading Volume," *Journal of Economic Dynamics and Control*, 16 (July/October): 451–489.

[12] Conroy, R., R. Harris, and B. Benet (1990), "The Effects of Stock Splits on Bid-Ask Spreads," *Journal of Finance*, 45 (September): 1285–1295.

[13] Copeland, T. (1979), "Liquidity Changes Following Stock Splits," *Journal of Finance*, 34 (March): 115–141.

[14] Copeland, T. and D. Galai (1983), "Information Effects on the Bid-Ask Spread," *Journal of Finance*, 38 (December): 1457–1469.

[15] Demsetz, H. (1968), "The Cost of Transacting," *Quarterly Journal of Economics*, 82 (February): 33–53.

[16] Dravid, A. (1987), "A Note on the Behavior of Stock Returns around Ex-Dates of Stock Distributions," *Journal of Finance*, 42 (March): 163–168.

[17] Dubofsky, D. (1991), "Volatility Increases Subsequent to NYSE and AMEX Stock Splits," *Journal of Finance*, 46 (March): 421–431.

[18] Easley, D. and M. O'Hara (1987), "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19 (September): 69–90.

[19] Easley, D. and M. O'Hara (1992), "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47 (June): 577–605.

[20] Fama, E., L. Fisher, M. Jensen, and R. Roll (1969), "The Adjustment of Stock Prices to New Information," *International Economic Review*, 10 (February): 1–21.

[21] Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. 1 (Third Edition), New York: Wiley.

[22] Garbade, K. and Z. Lieber (1977), "On the Independence of Transactions on the New York Stock Exchange," *Journal of Banking and Finance*, 1 (October): 151–172.

[23] Garman, M. (1976), "Market Microstructure," *Journal of Financial Economics*, 3 (June): 257–275.

[24] George, T., G. Kaul, and M. Nimalendran (1991), "Estimation of the Bid-Ask Spread and Its Components: A New Approach," *Review of Financial Studies*, 4: 623–656.

[25] Glosten, L. (1987), "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42 (December): 1293–1307.

[26] Glosten, L. and L. Harris (1988), "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21 (May): 123–142.

[27] Glosten, L. and P. Milgrom (1985), "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14 (March): 71–100.

[28] Grossman, S. and M. Miller (1988), "Liquidity and Market Structure," *Journal of Finance*, 43 (July): 617–633.

[29] Harris, L. (1989), "A Day-End Transaction Price Anomaly," *Journal of Financial and Quantitative Analysis*, 24 (March): 29–45.

[30] Harris, L. (1990a), "Liquidity, Trading Rules, and Electronic Trading Systems," *New York University Monograph Series in Finance and Economics*, Monograph No. 1990–4.

[31] Harris, L. (1990b), "Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator," *Journal of Finance*, 45 (June): 579–590.

[32] Harris, L. (1992), "Minimum Price Variations, Discrete Bid/Ask Spreads, and Quotation Transparency," Working Paper 91–42, University of Southern California.

[33] Hasbrouck, J. (1991), "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46 (March): 179–207.

[34] Hasbrouck, J. and T. Ho (1987), "Order Arrival, Quote Behavior, and the Return-Generating Process," *Journal of Finance*, 42 (September): 1035–1048.

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

[36] Ho, T. and H. Stoll (1981), "Optimal Dealer Pricing Under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9 (March): 47–73.

[37] Kempthorne, P. and T. Marsh (1990), "Stock Price Subordination in a State Space Specification," Working Paper.

[38] Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82 (December): 1032–1063.

[39] Kyle, A. (1985), "Continuous Auctions and Insider Trading," *Econometrica*, 53 (November): 1315–1335.

[40] Lakonishok, J. and B. Lev (1987), "Stock Splits and Stock Dividends: Why, Who, and When," *Journal of Finance*, 42 (September): 913–932.

[41] Lamoureux, C. and P. Poon (1987), "The Market Reaction to Stock Splits," *Journal of Finance*, 42 (December): 1347–1370.

[42] Lee, C., B. Mucklow, and M. Ready (1993), "Spreads, Depths and the Impact of Earnings Information: An Intraday Analysis," *Review of Financial Studies*, 6: 345–374

[43] Lee, C. and M. Ready (1991), "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46 (June): 733–746.

[44] Madhavan, A. and S. Smidt (1991), "A Bayesian Model of Intraday Specialist Pricing," *Journal of Financial Economics*, 30 (November): 99–134.

[45] McInish, T. and R. Wood (1992), "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks," *Journal of Finance*, 47 (June): 753–764.

[46] Mendelson, H. (1982), "Market Behavior in a Clearing House," *Econometrica*, 50 (November): 1505–1524.

[47] Mood, A., F. Graybill, and D. Boes (1974), *Introduction to the Theory of Statistics* (Third Edition), McGraw-Hill.

[48] Murray, D. (1985), "Further Evidence on the Liquidity Effects of Stock Splits and Stock Dividends," *Journal of Financial Research*, 8 (Spring): 59–67.

[49] Niederhoffer, V. and M. Osborne (1966), "Market Making and Reversal on the Stock Exchange," *Journal of the American Statistical Association*, 61 (December): 897–916.

[50] O'Hara, M. and G. Oldfield (1986), "The Microeconomics of Market Making," *Journal of Financial and Quantitative Analysis*, 21 (December): 361–376.

[51] Ohlson, J. and S. Penman (1985), "Volatility Increases Subsequent to Stock Splits," *Journal of Financial Economics*, 14 (June): 251–266.

[52] Roll, R. (1984), "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39 (September): 1127–1139.

[53] Smidt, S. (1979), "Continuous versus Intermittent Trading on Auction Markets," *Journal of Financial and Quantitative Analysis*, 14 (November): 837–866.

[54] Stoll, H. (1978), "The Supply of Dealer Services in Securities Markets," *Journal of Finance*, 33 (September): 1133–1151.

[55] Stoll, H. (1989), "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance*, 44 (March): 115–134.

[56] Tiniç, S. (1972), "The Economics of Liquidity Services," *Quarterly Journal of Economics*, 86 (February): 79–93.

[57] Tiniç, S. and R. West (1972), "Competition and the Pricing of Dealer Service in the Over-the-Counter Stock Market," *Journal of Financial and Quantitative Analysis* (June): 1707–1727.

[58] Venkatesh, P. and R. Chiang (1986), "Information Asymmetry and the Dealer's Bid-Ask Spread: A Case Study of Earnings and Dividend Announcements," *Journal of Finance*, 41 (December): 1089–1102.

[59] Working, H. (1954), "Price Effects of Scalping and Day Trading," in A.E. Peck, ed., *Selected Writings of Holbrook Working* (Chicago Board of Trade, Chicago, IL): 181–193.