Market Microstructure Invariants *

Albert S. Kyle  
Robert H. Smith School of Business  
University of Maryland  
akyle@rhsmith.umd.edu

Anna A. Obizhaeva  
Robert H. Smith School of Business  
University of Maryland  
obizhaeva@rhsmith.umd.edu

May 8, 2011

Abstract

The hypothesis of “market microstructure invariance”—based on the intuition that stocks vary on a time scale according to which “bets” generate trading activity—is tested using a database of 400,000+ portfolio transition trades. Defining trading activity \( W \) as the product of dollar volume and returns standard deviation, microstructure invariance predicts that order size, market impact costs, and bid-ask spread costs—as fractions of daily volume—are proportional to \( W^{-2/3} \), \( W^{1/3} \), and \( W^{-1/3} \), respectively. Estimated exponents are -0.63, 0.33, and -0.39, respectively, with order order size conforming to a log-normal distribution. Calibration leads to a simple transaction cost formula based on trade size, volume, and volatility.

---

*We are grateful to Georgios Skoulakis, Mark Loewenstein, and Vish Viswanathan for helpful comments. Obizhaeva is also grateful to the Paul Woolley Center the at London School of Economics for its hospitality as well as Ross McLellan, Simon Myrgren, Sébastien Page, and especially Mark Kritzman for their help.
1 Introduction

This paper uses the concept of time scale to propose an empirical hypothesis we call “market microstructure invariance.” Market microstructure invariance leads to precise predictions concerning how bet size, market impact costs, and bid-ask spread costs vary across stocks with different levels of trading activity. These predictions are tested statistically using a database of more than 400,000 portfolio transition orders provided by a leading vendor of portfolio transition services. Although not in perfect agreement with the theory, the statistical results are remarkably consistent with the predictions of market microstructure invariance.

When portfolio managers trade stocks, they can be modeled as playing trading games. Trading game models typically make specific assumptions about the consistency of beliefs across traders, the flow of public and private information which informed traders use to trade, the flow of orders from liquidity traders, and auction mechanisms in the context of which market makers compete to take the other side of trades. Some models emphasize adverse selection, such as Treynor (1971), Kyle (1985), and Glosten and Milgrom (1985); some models emphasize inventory dynamics, such Grossman and Miller (1988) and Campbell and Kyle (1993); and some models emphasize both, such as Grossman and Stiglitz (1980) and Jiang Wang (1993).

The purpose of this paper is to develop and test a model of order flow dynamics, market depth, and bid-ask spreads based on simple invariance principles, without making specific assumptions about how informed traders, liquidity traders, and market makers interact in a specific game-theoretic context. Instead, our emphasis concerns how to infer a measure of order flow imbalance from price, volume, and volatility data, not on the game theoretic principles that derive endogenously an appropriate level of price volatility using concepts of competition and market efficiency. For example, the hypothesis of market microstructure invariance provides an approach for testing how the market depth parameter $\lambda$ of Kyle (1985) varies cross-sectionally across stocks, under the maintained hypothesis that risk neutral market makers make markets semi-strong form efficient.

Market microstructure invariance is based on the intuition that trading in individual stocks is a game played at different speeds for different stocks. Trading time is different from calendar time; the trading game clocks ticks more quickly for actively traded stocks than for inactively traded stocks. The hypothesis of market microstructure invariance is based on the idea that empirically invariant market microstructure properties can be derived by examining trading at a time scale appropriate to the stock’s level of trading activity. We propose that an appropriate time scale implies a trading game clock that ticks at a rate proportional to the arrival rate of “bets” in the market. We think of a bet as a strategic decision to change risk exposure by buying or selling individual stocks, likely implemented through numerous orders, trades, and prints spread out over time. The idea that financial markets operate in trading time that is different from calendar time is not new. Rather, we continue a long-standing discussion by Mandelbrot and Taylor (1967), Clark (1973), Hasbrouck(1999, 2007), and Dufour and Engle (2000) concerning whether “time” in financial markets should be measured relative to “calendar time,” “volume time,” or “transaction time.” Our main hypothesis is that all trading games are fundamentally the same, when the time clock is adjusted appropriately for the level of trading activity.
The hypothesis of market microstructure invariance leads to precise predictions about how bet size, market impact costs, and bid-ask spread costs vary across stocks with different levels of trading activity. Since trading in financial assets exchanges risks and the size of the risk depends on both notional value and returns volatility, our measure of “trading activity” is the product of daily dollar volume and daily percentage standard deviation of returns. We think of market microstructure invariance as a combination of three distinct but related conjectures, which we call “trading game invariance,” “market impact invariance,” and “bid-ask spread invariance,” respectively. Each of these conjectures leads directly to empirically testable hypotheses:

- **Trading game invariance** hypothesizes that between each tick on the re-scaled clock, the distribution of risks transferred by a bet does not vary across stocks or across time. Trading game invariance predicts that when trading activity increases by one percent, the arrival rate of bets increases by $\frac{2}{3}$ of one percent and the expected size of bets increases by $\frac{1}{3}$ of one percent, but the shape of the distribution of bet sizes does not change. Bet size as a fraction of average daily volume is reduced by $\frac{2}{3}$ of one percent.

- **Market impact invariance** hypothesizes that the expected market impact cost of a bet is constant across stocks and across time. Market impact invariance predicts when trading activity increases by one percent, the market impact costs (per dollar traded) incurred by executing a bet equal to a given fraction of average daily volume (say one percent) increases by $\frac{1}{3}$ of one percent.

- **Bid-ask spread invariance** hypothesizes that the expected bid-ask spread cost of a bet is constant across stocks and across time. Bid-ask spread invariance predicts that when trading activity increases by one percent, the bid-ask spread costs (per dollar traded) incurred by executing a bet equal to a given fraction of average daily volume (say one percent) decreases by $\frac{1}{3}$ of one percent.

The intuition for these results derives from the fact that speeding up a trading game clock increases proportionally both the order arrival rate and the variance of the risk transferred by a bet. Since standard deviation is the square root of variance, standard deviation increases half as fast as order arrival rate. Since our measure of the risk transferred by a bet is based on standard deviation, the risk transferred by a bet increases half as fast as the arrival rate of bets. Therefore, as trading activity increases, $\frac{2}{3}$ of the increase represents the increased arrival rate of bets and remaining $\frac{1}{3}$ of the increase represents the increased risk transferred by each bet per unit of calendar time.

We test these predictions against two alternative models based on different invariance principles. We call these alternatives the model of “invariant bet frequency” and the model of “invariant bet size.” The model of invariant bet frequency, which is similar to Amihud (2002) and captures the conventional wisdom of traders, is based on the assumption that as trading activity increases, the number of bets per day remains constant but the size of bets scales up proportionally. The model of invariant bet size, which captures some features of Hasbrouck (2009), is based on the assumption that as trading activity increases, the size of bets remains constant but the number of bets per day increases. In comparison with our proposed model of market microstructure invariance, these alternative models make dramatically different
predictions about how bet size, market impact costs, and bid-ask spread costs vary with the level of trading activity. All three models nest conveniently into common specifications amenable to statistical testing.

We test the predictions of the three alternative models using a database of more than 400,000 portfolio transition orders executed over the period 2001-2005 by a leading vendor of portfolio transition services. In a portfolio transition, a legacy portfolio manager is replaced with a new manager. To implement the trades necessary to convert the legacy portfolio into the new target portfolio, a skilful third-party transition manager is often hired to replace the legacy portfolio with the new one. The transition manager knows precisely the sizes of the trades needed to implement the transition, and the trades are typically implemented over an easily identified, tight time frame.

We use the portfolio transition orders to test both the order size predictions, market impact predictions, and bid-ask spread predictions of our hypothesis of market microstructure invariance against the two alternative competing theories. Our tests support the model of microstructure invariance.

Tests Based on Order Sizes. Consistent with the predictions of market microstructure invariance, we find that the shape of the distributions of order sizes adjusted for trading activity is almost the same across stocks in different volume and volatility groups. Although not predicted by the microstructure invariance, we find that the shape of these distributions closely resembles a log-normal. This empirical finding is far different from the normal distribution assumed in Kyle (1985). While differences in the distribution of order sizes across volume and volatility groups are large enough to be statistically significant, the differences are barely noticeable based on visual inspection, except for a handful of outliers. A handful of the very largest orders in the most inactive stocks are noticeably smaller than predicted by a log-normal, consistent with the interpretation that portfolio managers seek to hold less than the regulatory reporting threshold of 5% of outstanding shares.

We test the predictions of market microstructure invariance concerning mean bet size by comparing how the sizes of portfolio transition orders vary with the level of trading activity, under the identifying assumption that portfolio transition orders are proportional to typical bets. A simple log-linear OLS regression shows that when trading activity increases by one percent, the size of a portfolio transition order as a fraction of average daily volume decreases by 0.63 percent (standard error 0.008). This result is close to the predicted decrease of 2/3 of one percent from market microstructure invariance, and far from the predictions of the models of constant bet frequency (a decrease of zero percent) and constant bet size (a decrease of one percent). While our hypothesis that the predicted decrease is equal to 2/3 is rejected due to the low standard error, the alternative models are rejected by much more significant margins. Quantile estimates provide similar results. Estimating order size from the individual components of trading activity—price, volume, and volatility—results in an economically modest but statistically significant increase in explanatory power.

Tests Based on Implementation Shortfall We use implementation shortfall to test the predictions of market microstructure invariance concerning how market impact costs and bid-ask spread costs vary across stocks with different levels of trading activity. Perold (1988)
defines implementation shortfall as the difference between a “paper trading” benchmark and actual trading results, marking to market unexecuted shares at post-trade prices. For our purposes, the paper trading benchmark for a given transition is defined to be the price which would have been obtained if all shares were executed at the market closing price the day before any trades implementing a given transition began to take place. This benchmark is compared against the actual prices at which the transition trades are later executed. The difference is measured in basis points. Implementation shortfall includes the effect of both market impact and bid-ask spread, as well as random changes in the stock price between the benchmark date and the time when the trades are executed. We make the identifying assumption that the returns on the stock would otherwise have had a mean return of zero, which implies that the mean of the implementation shortfall is an unbiased estimate of expected transaction costs.

Implementation shortfall as a measure of transactions costs typically has problems associated with selection bias and lack of statistical power. Both of these problems are mitigated in the database of portfolio transition trades. Selection bias is avoided because the orders sizes needed to implement a portfolio transition are known precisely at the beginning of the transition and transition managers do not cancel orders when the price runs away from them. As discussed in Obizhaeva (2009b), this selection bias problem can otherwise be substantial, because orders canceled when the price runs away from order have an opportunity cost which may be inappropriately not taken into account. The lack of statistical power is also mitigated in the portfolio transition database due to the large number degrees of freedom arising from 400,000+ orders, the large sizes of orders relative to average daily volume, and the tight time frame over which the orders are executed. In particular, large portfolio transition orders magnify price impact relative to other trading noise, thus overcoming the empirical difficulties Black (1986) points out.

Transactions costs are estimated in a non-linear regression in which the transactions cost is assumed to be the sum of a price impact term (with assumed linear price impact) and a bid-ask spread term. For market impact, the power parameter estimate is 0.33 (with standard error 0.024), almost exactly equal to the value of 1/3 predicted by market microstructure invariance and far different from the value of zero predicted by the model of invariant bet frequency and the value of 1/2 predicted by the model of invariant bet size. For bid-ask spread, the power parameter estimate is -0.39 (with standard error 0.025), almost close enough to the predicted value of -1/3 to avoid statistical rejection, and far different from the value of zero predicted by the model of invariant bet frequency and the value of -1/2 predicted by the model of invariant bet size.

The transaction cost model is also estimated using the quoted spread rather than an implicit spread estimate. Under the assumption that the realized spread cost is proportional to the quoted spread, the model of market microstructure invariance implies that a regression of the log of the quoted spread on the log of trading activity should have a coefficient of -1/3. The estimated coefficient is -0.36 (standard error 0.003). When the transactions cost estimation procedure is modified to require spread costs to be proportional to the quoted spread rather than a power of trading activity, the estimated exponent of impact costs changes from 0.33 to 0.31 (standard error 0.025), and the adjusted R-squared is reduced from 0.0123 to 0.0110. These results suggest that our implicitly estimated spread cost is consistent with quoted spreads, but estimating the spread cost implicitly from trading
activity is more accurate than using quoted spreads.

We interpret these empirical results to imply that, to a first approximation, variation in order size, market impact costs, and spread costs across stocks with different levels of trading activity are consistent with the model of market microstructure invariants. The fit is not, however, precise enough to rule out the possibility that other factors also influence variation in these variables. Whether taking into account these other factors would strengthen or weaken the consistency of the data with the predictions of microstructure invariance poses an interesting set of issues for further research.

Having confirmed from empirical estimates using portfolio transition data that the predictions of the model of microstructure invariants are reasonable, we calibrate the results by investigating implications of implied parameter values concerning the way markets for individual stocks work. The calibration requires making identifying assumptions. We assume that portfolio transition orders are equal to the size of typical bets. We also assume that trading volume can be decomposed into bet volume and non-bet volume which intermediates among bets. To calibrate results, we assume that half of trading volume is bet volume and the other half is non-bet intermediation volume, consistent with the interpretation that a monopolistic specialist intermediates all trade.

Using these identification assumptions, we estimate that for the benchmark stock (trading volume of 1 million shares per day, stock price of $40 per share, and volatility of 2% per day), the expected number of bets per day is 85. The bets are log-normally distributed in size with mean bet size of $471,838 and median bet size of $135,184. These estimates scale up and down for stocks with different levels of trading activity, using elasticities of 2/3 and 1/3 as implied by microstructure invariance. For example, increasing trading activity by a factor of 8 increases the number of bets per day by a factor of 4 and increases the size of bets by a factor of 2. The expected market impact cost and spread costs of a bet are $1960 and $373, respectively. The expected market impact cost and spread cost of the median bet are only $13.20 and $107, respectively. Consistent with conventional trader wisdom, market impact costs dominate spread costs for large trades, but spread costs dominate market impact costs for small trades. The invariance hypothesis implies that the log-standard deviation of bet size (as well as other moments related to the shape of the distribution of bet size), the expected market impact cost of a bet, and the expected bid-ask spread cost of a bet are unchanged as a result of increasing the level of trading activity. The log-standard deviation of bet size is 1.58, implying that a one standard deviation change in order size represents a factor of 4.86. Large bets dominate small bets in terms of their contribution to trading volume and return variance. The larger half of bets, for example, accounts for 94.29% of trading volume and generates 99.92% of total variance.

The invariance hypothesis implies a simple, easily implementable formula estimating transactions costs, taking as its only input expected daily dollar volume and expected daily percentage returns volatility (see equations (65) and (66) below). For the benchmark stock, a trade of one percent of average daily volume is estimated to incur a market impact cost of 2.89 basis points and a spread cost of 7.91 basis points. The invariance hypothesis predicts how to extrapolate these results for the benchmark stock to stocks with different levels of dollar volume and volatility. Specifically, the hypothesis implies that a one percent increase in dollar volume increases the market impact cost by 1/3 of one percent and decreases the spread cost by 1/3 of one percent, while a one percent increase in volatility increases the
market impact cost by $4/3$ of one percent and decreases the spread cost by $2/3$ of one percent. The assumption of linear price impact implies that increasing trade size increases the market impact cost (in basis points) proportionally but has no effect on spread costs. It is an interesting question for further research to investigate whether our formula—estimated from U.S. stock data—generalizes to bond markets, commodity markets, futures markets, foreign exchange markets, and markets in other countries.

From the invariance hypothesis, we derive two two liquidity indices for the markets for individual stocks (see equations (68) and (69) below). The first index is based on the expected cost of converting an asset to cash (scaled in basis points). The second index is based on the expected cost of transferring risks of the sizes exchanged in the market (scaled like a Sharpe ratio). Both measures are related to market “velocity”—the number of bets per calendar day. Our insights about the structure of the trading process may also help to improve existing measures of liquidity such as, for example, the probability of informed trading (PIN) introduced by Easley et al. (1996).

The predictions of microstructure invariance are broadly consistent with established empirical regularities summarized in Bouchaud, Farmer, and Lillo (2009). In the results that they summarize, stock size is measured by market capitalization and trade size is measured by the size of TAQ “prints.” To compare their results with ours, we have to assume that market capitalization is proportional to trading activity, TAQ print size is proportional to bet size, and percentage returns variance does not vary much across stocks. They report that the number of trades per day has been estimated to be proportional to stock size raised to powers ranging between 0.44 to 0.86. The midpoint of this range is 0.65, close to our predicted power of $2/3$. They report that market impact is proportional to stock size raised to the power -0.30, close to our predicted power of $-1/3$. They report that the bid-ask spread is proportional to the standard deviation of percentage returns between trades. Microstructure invariance predicts the same proportionality, since both percentage returns variance between trades and percentage bid ask spread are proportional to trading activity raised to same the power of $-1/3$.

The remainder of this paper, section 2 presents market microstructure invariance as an empirical hypothesis, section 3 nests the invariance hypothesis into a framework with two alternatives, section 4 describes the portfolio transitions data, section 5 presents empirical results based on order size predictions, section 6 presents empirical results based on impact and spread cost predictions, and section 7 presents a calibration exercise confirming that the model-implied estimates for various market microstructure variables have reasonable, internally consistent values.

2 Market Microstructure Invariance: An Empirical Hypothesis

It is well-known that microstructure characteristics such as order size, order arrival rate, price impact, and bid-ask spreads vary across stocks and across time. We refer to “market microstructure invariance” as the hypothesis that this variation almost disappears if the market microstructure of the trading processes is examined at an appropriate time scale.
In order to examine trading at different time scales, we avoid making explicit assumptions about the structure of information, the motivations of traders, and the consistency of their beliefs. Instead, we take a reduced form approach, focusing on the effects of different time scales.

2.1 Trading Activity as a Sequence of Bets

We think of trading a stock as playing a trading game, in which traders buy and sell shares to implement bets in the market. As a first approximation, we assume that “bets” — representing decisions to buy or sell shares — arrive approximately randomly with an arrival rate of $\gamma$ bets per day and that bets are approximately independently distributed, with positive quantities representing buying and negative quantities representing selling. Let $P$ denote the price of the stock in dollars per share. Let $\sigma$ denote the percentage standard deviation of returns on the stock per day.

Since financial markets are places where risks are transferred, a good measure of the size of a bet should be the size of the risk transferred. This includes both the notional value of the bet, $P \cdot \tilde{Q}$, and the risk per unit of notional value. Therefore, a good measure of the risk transferred by a bet per calendar day is the product of the notional value $P \cdot \tilde{Q}$ and the daily volatility $\sigma$:

$$\tilde{B} = \tilde{Q} \cdot P \cdot \sigma.$$  \hspace{1cm} (1)

For bonds, this definition of bet risk is consistent with the idea that a given dollar value of volatile long-duration Treasury bonds is much riskier than the same dollar value of short duration Treasury bills. For stocks, this definition is consistent with the idea that the size of a bet of given dollar size in low-volatility stock might be the same as the size of a bet of smaller dollar size in a higher volatility stock.

Our definitions of trading activity and bets are consistent with the Modigliani-Miller irrelevance of leverage. For example, if a company leverages up its equity by paying a debt-financed cash dividend equal to fifty percent of the value of the equity, then the volatility of the remaining equity, ex dividend, should double, while the price should halve. If share volume remains constant, then dollar volume halves, but our measures of trading activity and bet size remain unaffected by the change in leverage. This is consistent with the intuition that each share of leveraged stock still represents the same pro rata share of firm risk as a share of un-leveraged stock.

Our definitions of trading activity and bets are also consistent with the Modigliani-Miller irrelevance of stock splits. For example, a two-for-one stock split should theoretically double the share volume of bets without affecting volatility. Since each share is worth one-half of its pre-split value, then our measures of bets and trading activity are not affected by a two-for-one stock split which doubles the share size of bets.

We assume that not all trading volume comes from bets. In addition to “primary” bet volume, we also assume there is “secondary” non-bet volume endogenously generated as a response to bets. Non-bet volume includes trading by market makers, high frequency traders, and other arbitragers. Bets represent long-term decisions to bear risk. Non-bet volume represents short-term decisions to bear risk temporarily, in a manner which intermediates among long-term bets.
Let $V$ denote expected daily trading volume, measured in shares per day. Since each unit of share volume has a buy-side and a sell-side, the expected share volume from matching buy bets and sell bets is given by $(1/2) \cdot \gamma \cdot E\{|Q|\}$. Expected total daily share volume $V$ is a multiple of this quantity. Letting $\zeta$ denote this multiple, expected daily volume is given by

$$V = \zeta \cdot \gamma \cdot E\{|\tilde{Q}|\}. \tag{2}$$

If all trades were bets, then we would have $\zeta = 1$, since each unit of trading volume would match a buy-bet with a sell-bet. To the extent that buy and sell bets do not match against each other perfectly, because intermediaries like market makers and arbitragers are involved in the trading process, the parameter $\zeta$ will be large than 1. For example, if the NYSE specialist were to intermediate all bets without involvement of other intermediaries, we would have $\zeta = 2$. If each bet were intermediated by different NASDAQ market makers, who each laid off inventories by trading with another intermediary who was not a market maker, then we would have $\zeta = 3$. If positions are passed around among multiple intermediaries, we would could have $\zeta = 4$ or more.

We do not expect the bet generating process to conform precisely to the definition of a compound poisson process, which requires random bet arrivals and independent bets. Instead, over short periods of time, we expect some variation in bet arrival rates, while over long periods of time, we expect there to be small negative serial correlation in bets, so that markets might clear “in the long run” without the presence of intermediaries. The assumption of a compound poisson process generating bets is a first-order approximation.

Bets are difficult for researchers to observe. Bets represent almost-statistically-independent decisions to take risky positions. They are different from orders, trades, or prints, which are the results of potentially correlated decisions to implement bets. For example, after performing weeks of research on a target stock, an asset manager might make a decision to purchase 100,000 shares. This decision represents a bet with $\tilde{Q} = 100,000$. The trader might implement this bet by placing a sequence of orders to purchase 20,000 shares of stock per day for five days in a row. Each of these orders might be broken into smaller pieces for execution. For example, on day one there may be trades of 2,000, 3,000 5,000, and 10,000 shares executed at different prices. Each of these smaller trades may each show up in TAQ data as multiple “prints.” Since the various orders, trades, and prints are positively correlated because they implement a common bet, it would not be appropriate to think of them as almost-statistically-independent bets.

A bet is like an idea, which can be shared. Thus, if an analyst recommendation to buy a stock is followed by buy orders from multiple customers, all of these orders are part of the same bet. For example, if an analyst makes a recommendation to ten different customers that they buy a stock, and all of the customers together buy 100,000 shares as a result of the recommendation, it is reasonable to think of this purchase of 100,000 shares as one bet with $\tilde{Q} = 100,000$, rather than as a sequence of ten different bets. Since the ten purchases are all based on the same information—the recommendation of the analyst—the decisions probably lack statistical independence, even to a first approximation.

Consistently with the idea that the risk of one transaction is measured by the product of dollar value and volatility, we define “trading activity” $W$ for a risky asset as the product of the daily dollar trading volume and the daily returns standard deviation of the asset.
\[ W = V \cdot P \cdot \sigma. \]  
\[(3)\]

Trading activity measures the standard deviation of one day’s close-to-close dollar gains or losses resulting from the purchase or sale of one day’s trading volume. In this sense, trading activity \( W \) measures the aggregate risk transfer expected to take place in one calendar day.

Since trading volume, stock price, and volatility can all be observed or estimated empirically, trading activity is empirically observable. Observable trading activity can be equivalently defined as the product of three quantities which are difficult to observe: one-half the volume multiplier \( \zeta \), the expected number of bets per day \( \gamma \), and expected (unsigned) bet size \( \mathbb{E}\{|\tilde{B}|\} \):

\[ W = \frac{\zeta}{2} \cdot \gamma \cdot \mathbb{E}\{|\tilde{B}|\}. \]  
\[(4)\]

To interpret the empirical results in this paper, we make the identifying assumption that the bet volume multiplier \( \zeta \) is a constant which is the same for all stocks. Investigating whether this assumption is true is an interesting subject for future research. For example, it is reasonable to investigate whether high frequency trading is a larger percentage of trading volume for stocks with high levels of trading activity than for stocks with low levels of trading activity. Under the identifying assumption of constant \( \zeta \), it is clear from equation (4) that trading activity \( W \) can increase because either the arrival rate of bets \( \gamma \) increases or expected bet size \( \mathbb{E}\{|\tilde{B}|\} \) increases, or both.

### 2.2 Three Principles of Market Microstructure Invariance

This paper examines an hypothesis we call “market microstructure invariance.” Market microstructure invariance is the hypothesis that if time is re-scaled so that trading activity occurs at a constant rate in re-scaled units of time, then microstructure characteristics like sizes of risks transferred during trading, order arrival rates, market impact, and bid-ask spreads will be invariant across stocks, when these variables are measured in re-scaled units of time. When re-scaled time is replaced by calendar time, however, market microstructure invariance implies that these invariant microstructure characteristics appear to vary across stocks in a precisely predictable manner.

The invariance hypothesis captures the intuition that trading games for stocks with different levels of trading activity are the same, except that are played at different speeds. Like the game of chess, the game of trading financial assets can be played quickly or slowly. Trading an active stock is like playing chess with a fast time clock. Trading an inactive stock is like playing chess with a slow time clock. Market microstructure invariance captures the intuition that except for the speed of the time clock, the trading games look the same across stocks.

To develop the implications of microstructure invariance, we re-scale the time so that one bet is expected to arrive in the market for each tick of the re-scaled clock. This implies that the calendar clock operates at a speed faster than the re-scaled clock by a factor equal to the arrival rate of bets \( \gamma \). Note that since the concept of a bet is different from a concept of transaction, the time clock implied by our model is different from the approach of Mandelbrot and Taylor (1967), who modeled trading in transaction time. The main intuition
of microstructure invariance is the idea that re-scaling time not only creates time units in which bets arrive at the same rate but also creates time units in which variance or bet risk is constant as well.

Using the idea of microstructure invariance, we test three invariance conjectures concerning (1) the size of bets, (2) the price impact costs of bets, and (3) the bid-ask spread costs of bets.

Trading Game Invariance: Between each tick on the re-scaled clock, the distribution of the risks transferred by a bet does not vary across different stocks and across time. The principle of trading game invariance leads to the empirical prediction that as trading activity increases, the arrival rate of bets increases twice as fast as bet size. Intuitively, this prediction results from the fact that speeding up time speeds up both bet arrival rate and bet variance per calendar day proportionately. Bet risk, however, is based on standard deviation, not variance. Taking a square root to convert variance to standard deviation implies that bet risk increases half as fast as the bet arrival rate. Thus, when trading activity increases by one percent, trading game invariance predicts that the the arrival rate of bets increases by 2/3 of one percent and the size of bets increases by 1/3 of one percent. This intuition can be made precise with some simple math.

The quantity \( \tilde{\beta}/\gamma^{1/2} \) measures the signed standard deviation of the dollars gained or lost on a bet over a time interval representing the average waiting time between the arrival of consecutive bets, with positive values for buys and negative values for sells. Trading game invariance states that the risk transferred by a bet between each tick on the re-scaled clock has a probability distribution which does not vary across stocks or across time. Letting \( \tilde{I} \) denote this “invariant” distribution, trading game invariance can be stated as follows:

\[
\tilde{I} \approx \frac{\tilde{\beta}}{\gamma^{1/2}} = \frac{P \cdot \tilde{Q} \cdot \sigma}{\gamma^{1/2}},
\]

where the notation “\( \approx \)” is interpreted to mean “has the same probability distribution as.” The hypothesis of trading game invariance states that the probability distribution for \( \tilde{I} \) is invariant across stocks and across time, even though the distributions of \( \tilde{\beta} \) and \( \tilde{Q} \) and the values of \( P, V, \sigma, \) and \( \gamma \) vary across stocks and across time.

Trading game invariance implies that the probabilities distributions of \( \tilde{I}, \tilde{\beta}, \) and \( \tilde{Q} \) all have the same shape, but the distributions of \( \tilde{\beta} \) and \( \tilde{Q} \) differ from \( \tilde{I} \) by proportionality coefficients \( \gamma^{-1/2} \) and \( P \cdot \sigma \cdot \gamma^{-1/2} \), respectively. For example, this hypothesis implies that as the bet arrival rate \( \gamma \) increases, the distribution of unsigned bet sizes \( |\tilde{\beta}| \) shifts upwards half as fast as the bet arrival rate changes.

Using equations (4) and (5), we can express both the bet arrival rate \( \gamma \) and bet size \( \tilde{\beta} \) as powers of trading activity \( W \):

\[
\gamma = \left[ \frac{\zeta}{2} \cdot E\{|\tilde{I}|\} \right]^{-2/3} \cdot W^{2/3},
\]

\[
\tilde{\beta} \approx \left[ \frac{\zeta}{2} \cdot E\{|\tilde{I}|\} \right]^{-1/3} \cdot W^{1/3} \cdot \tilde{I}
\]

In these equations, \( W \) varies across stocks and across time, but the coefficients multiplying \( W \), powers of \( (\zeta/2) \cdot E\{|\tilde{I}|\} \), are constants. These equations imply that if \( W \) increases by
one percent, the arrival rate of bets $\gamma$ increases by $2/3$ of one percent and the distribution of bet size $\tilde{B}$ shifts upwards by $1/3$ of one percent.

In practice, the size of an institutional bet is often expressed as a fraction of average daily volume, which in our notation is $\tilde{Q}/V$. Since $\tilde{Q}/V = \tilde{B}/W$, expressing $\tilde{Q}/V$ as a function of trading activity yields

$$\tilde{Q}/V \approx \left[ \frac{\zeta}{2} \cdot E\{\tilde{I}\} \right]^{-1/3} \cdot W^{-2/3} \cdot \tilde{I}. \quad (8)$$

Note that in equations (6), (7), and (8), parameter $\zeta$ effectively deflates both trading volume $V$ and trading activity $W$ by eliminating non-bet volume from these variables.

**Market Impact Invariance:** The expected market impact cost of a bet is the same across stocks and across time. Let $\lambda$ denote the linear market impact cost of buying or selling one share of stock, measured in units of dollars per share per share, or “dollars per share-squared.” This notation for $\lambda$ is consistent with Kyle [1985]. Assuming linear price impact costs, the expected impact cost of a bet is a quadratic function of bet size given by $(\lambda/2) \cdot E\{\tilde{Q}^2\}$, with $\lambda$ and $\tilde{Q}$ varying across stocks and across time. Let $C_L$ denote a constant dollar impact cost which does not vary across stocks or across time. Then market impact cost invariance can be expressed as

$$C_L = \frac{\lambda}{2} \cdot E\{\tilde{Q}^2\}, \quad (9)$$

where $\lambda$ and $\tilde{Q}$ vary across stocks and across time, but $C_L$ does not.

Using equation (9), the percentage market impact of trading $\tilde{Q}$ shares is given by,

$$\frac{\lambda}{P} \cdot |\tilde{Q}| = 2 \cdot C_L \cdot \left[ \frac{\zeta}{2} \right]^{2/3} \cdot \left[ E\{\tilde{I}\} \right]^{2/3} \cdot \frac{E\{\tilde{I}^2\}^{2/3} \cdot \sigma \cdot |\tilde{Q}|}{V}. \quad (10)$$

Under the combined assumptions of trading game invariance and market impact invariance, the variables $\tilde{I}$ and $C_L$ in equation (10) do not vary across stocks or across time. This equation leads to the empirical hypothesis that the expected market impact cost of trading a fixed percentage of daily volume accounts for a fraction of daily volatility proportional to $W^{1/3}$, with a proportionality constant $2 \cdot C_L \cdot \left[ \zeta/2 \right]^{2/3} \cdot E\{\tilde{I}^2\}^{-2/3}$ which does not vary across stocks or across time (assuming $\zeta$ is also constant).

**Bid-Ask Spread Invariance:** The expected bid-ask spread cost of a bet is the same across stocks and across time. Let $\kappa$ denote the bid-ask spread, measured in dollars per share. The expected bid-ask half-spread cost of a bet, expressed in dollars, is given by $(\kappa/2) \cdot E\{\tilde{Q}\}$. Let $C_K$ denote the invariant dollar half-spread cost of a bet, which does not vary across stocks or across time. Then bid-ask spread invariance implies

$$C_K = \frac{\kappa}{2} \cdot E\{\tilde{Q}\}, \quad (11)$$

where $\kappa$ and $\tilde{Q}$ vary across stocks but $C_K$ does not.
Using equation (11), the percentage expected bid-ask spread cost of a one-share round-trip trade is given by,

\[
\frac{K}{P} = 2 \cdot C_K \cdot \left[\frac{1}{2} \zeta \right]^{1/3} \cdot \left[ E\{|\tilde{I}|\} \right]^{-2/3} \cdot W^{-1/3} \cdot \sigma.
\]  

(12)

This equation leads to the empirical hypothesis that the expected bid-ask spread cost of trading a share accounts for a fraction of daily volatility proportional to \( W^1 \), where the proportionality constant, given by \( 2 \cdot C_K \cdot \left[ \zeta/2 \right]^{1/3} \cdot \left[ E\{|\tilde{I}|\} \right]^{-2/3} \), does not vary across stocks or across time (assuming \( \zeta \) constant).

Market microstructure invariance is a combination of the three hypotheses of (1) trading game invariance, (2) market impact invariance, and (3) bid-ask spread invariance.

3 Alternative Hypotheses

We examine two specific alternatives to our hypothesis of microstructure invariance. These alternative hypotheses serve three purposes. First, they provide empirical alternatives which can be easily tested against our proposed model because they nest easily into a common specification. Second, the two alternatives are similar in spirit to the models of Amihud (2002) and Hasbrouck (2009), respectively. In this sense, the alternative models capture the thinking behind the current literature. Third, we believe that the alternative models capture the intuition traders use to think about trading costs.

**Alternative Models and Bet Size**  To specify alternative hypotheses we generalize the invariance equation (5). Let \( \alpha \) denote a parameter such that \( 0 \leq \alpha \leq 1 \). The generalized invariance equation is

\[
\tilde{I} \approx B \cdot \tilde{Q} \cdot \sigma \cdot \gamma^{\alpha \cdot (1-\alpha)} \cdot \left[ \frac{\zeta}{\zeta^*} \right]^{\alpha \cdot (1-\alpha)} \cdot \left[ \frac{W}{W^*} \right]^{1-\alpha}.
\]  

(13)

For \( \alpha = 1 \), we interpret this equation to mean that \( \gamma \) is a constant. Generalized versions of equations (6), (7), and (8) can be easily derived by solving equations (4) and (13) for \( \gamma, \tilde{B}, \) and \( \tilde{Q}/V \) in terms of \( \zeta, W, \) and moments of \( \tilde{I} \).

Let \( Q^*, V^*, \zeta^*, \) and \( W^* \) denote trade size, daily expected share volume, the volume deflator, and daily trading activity for a hypothetical benchmark stock which conforms to equations (1) through (4). Dividing the generalizations of equations (6), (7), and (8) for some stock by corresponding equations for the benchmark results in moments of \( \tilde{I} \) canceling and yields the following equations similar to equations (6), (7), and (8):

\[
\gamma = \gamma^* \cdot \left[ \frac{\zeta}{\zeta^*} \right]^{-(1-\alpha)} \cdot \left[ \frac{W}{W^*} \right]^{1-\alpha},
\]  

(14)

\[
\tilde{B} \approx \tilde{B}^* \cdot \left[ \frac{\zeta}{\zeta^*} \right]^{-\alpha} \cdot \left[ \frac{W}{W^*} \right]^\alpha,
\]  

(15)

\[
\frac{\tilde{Q}}{V} \approx \frac{\tilde{Q}^*}{V^*} \cdot \left[ \frac{\zeta}{\zeta^*} \right]^{-\alpha} \cdot \left[ \frac{W}{W^*} \right]^{-(1-\alpha)}.
\]  

(16)
The ratio \( \zeta / \zeta^* \) here adjusts both trading volume \( V \) and trading activity \( W \) for non-bet volume. The three equations (14), (15), and (16) are all equivalent ways of expressing the same empirical predictions. Assuming \( \zeta = \zeta^* \) for all stocks, these equations state that as trading activity increases, a fraction \( 1 - \alpha \) of the increase results from increased bet arrival rate and therefore a fraction \( \alpha \) of the increase results from increased bet size, implying that bet size as a fraction of trading activity declines at rate \( 1 - \alpha \) as trading activity increases. Market microstructure invariance is equivalent to the hypothesis

\[
\alpha = 1/3, \tag{17}
\]

implying bet frequency increases twice as fast as bet arrival rate as trading activity increases.

Each value of \( \alpha \) in the range \( 0 \leq \alpha \leq 1 \) corresponds to a different alternative model. We focus on two alternative models, corresponding to \( \alpha = 1 \) and \( \alpha = 0 \).

Our first alternative model, which we call the model of invariant bet frequency, is based on the hypothesis that the number of bets per day is invariant across stocks with different levels of trading activity. This model implies that as trading activity increases, bet size increases proportionally but the expected number of bets per day is constant. Thus, bet size as a fraction of volume is constant across stocks and across time. The model of invariant bet frequency is equivalent to the hypothesis

\[
\alpha = 1. \tag{18}
\]

Our second alternative model, which we call the model of invariant bet size, is based on the hypothesis that the size of bets is invariant across stocks with different levels of trading activity. This model implies that as trading activity increases, the number of bets increases proportionally but bet size does not change. The model of invariant bet size is equivalent to the hypothesis

\[
\alpha = 0. \tag{19}
\]

Figure 1 illustrates differences between our model of market microstructure invariance and the two alternatives by comparing two hypothetical stocks. One stock has four equal-size bets per day \( (\gamma^* = 4) \). Its expected daily volume is 1 million shares \( (V^* = 10^6) \). Thus, each bet of 250,000 shares \( (Q^* = 250,000) \) contributes one-fourth of average daily volume. Consider another stock that has the same volatility, price, and volume deflator, but daily volume of 8 million shares \( (V = 8 \cdot 10^6) \).

The three model differ in how they decompose the eight-fold difference in trading activity into differences in bet size and bet arrival rate. The model of invariant bet frequency assumes that there are still four bets, but these bets are eight times larger \( (\gamma = 4, \tilde{Q} = 8 \cdot 250,000) \). The model of invariant bet size assumes that bets are still of 250,000 shares, but they are eight times more frequent \( (\gamma = 8 \cdot 4, Q = 250,000) \). Market microstructure invariance implies that both bet arrival rate and bet size increase. Since bet arrival rate increases twice as fast as bet size, bet arrival rate increases by a factor of four and bet size increases by a factor of two \( (\gamma = 8^{2/3} \cdot 4, Q = 8^{1/3} \cdot 250,000) \). Note that the bet size in bet time remains the same across two games. Specifically, while the number of shares traded doubles, the standard deviation of gains between bet arrivals halves because of time interval between bet arrivals decreases by a factor of four. In this sense, both trading games look the same, with one game played four times faster than the other one.
Alternative Models and Market Depth In order to generate predictions for market depth from the alternative models, it is useful first to make a connection between market impact costs and returns volatility. We continue to assume that market impact is linear in bet size. Since each of \( \gamma \) expected bets generates price impact \( \lambda \cdot \tilde{Q} \), the variance of the price impact of \( \gamma \) bets per day is \( \gamma \cdot \lambda^2 \cdot E\{\tilde{Q}^2\} \). The variance of price changes per day is \( \sigma^2 P^2 \).

We assume that a fraction \( \psi^2 \) of price variance results from the price impact of trades. This immediately implies the following relationship between the market impact costs of bets and price volatility:

\[
\psi^2 \cdot \sigma^2 \cdot P^2 = \gamma \cdot \lambda^2 \cdot E\{\tilde{Q}^2\}. \tag{20}
\]

If prices change as a result of announcements that incorporate information into prices without trading, we tend to have \( \psi < 1 \). If the price impact of trading has a transitory component, we tend to have \( \psi > 1 \). The continuous model of Kyle (1985), where all price changes result from trading and all price impact is permanent, is consistent with \( \psi = 1 \).

We conjecture that \( \psi \) is constant across stocks and across time. Under the assumptions of market microstructure invariance, the invariance of \( \psi \) is equivalent to the invariance of the expected market cost of a bet \( C_L \). Indeed, the connection between the two variables can be easily derived from (9), (20), and (5):

\[
C_L = \frac{1}{2} \cdot \psi \cdot \left[E\{\tilde{I}^2\}\right]^{1/2}. \tag{21}
\]

Under alternative models, where \( \alpha \neq 1/3 \), \( C_L \) may vary across stocks and time, even if \( \psi \) does not vary. The variable \( \psi \), however, is an intuitive market impact invariant which makes it possible to derive simple formula for market depth applicable to our model of microstructure invariance and to the alternative models corresponding to different values of \( \alpha \).

Using equation (20), price impact \( \lambda \) is given by

\[
\lambda = \frac{\psi \cdot \sigma \cdot P}{\gamma^{1/2} \cdot \left[E\{\tilde{Q}^2\}\right]^{1/2}}. \tag{22}
\]

Note that the derivation of equation (22) does not depend on an assumed value for \( \alpha \); it is valid for the our model of microstructure invariance (\( \alpha = 1/3 \)) and for alternative models (e.g. \( \alpha = 0 \) and \( \alpha = 1 \)). Equations (13) through (16) provide different ways of calculating the denominator \( \gamma^{1/2} \cdot \left[E\{\tilde{Q}^2\}\right]^{1/2} \) for different assumptions about the parameter \( \alpha \).

Equation (22) is very similar to the formula \( \lambda = \sigma_V / \sigma_U \) from the continuous model of Kyle (1985). In Kyle (1985), \( \sigma_V \) denotes the standard deviation of an informed trader’s private information. In the continuous equilibrium, \( \sigma_V \) also equals the dollar standard deviation of changes in the stock price over one trading day, assuming that market makers correctly anticipate the adverse selection in the order flow. In equation (22), the numerator \( \psi \cdot \sigma \cdot P \) also measures the dollar standard deviation of price changes resulting from trading. In Kyle (1985), \( \sigma_U \) denotes the standard deviation of the net quantity traded by noise traders in one trading day. In the continuous equilibrium, \( \sigma_U \) also equals the standard deviation of the combined order flow from both informed and noise traders, since the informed trader smooths out his trading so that it does not contribute to order flow variance. In equation (22), the denominator measures the standard deviation of the net quantity traded by all bets placed in one calendar day. To conform to the intuition of Kyle (1985), these bets should
contain a mix of informed trading and noise trading. Under the maintained assumptions that rational expectations by market makers results in linear impact with the appropriate amount of market depth (captured by price volatility $\sigma$), market microstructure invariance ($\alpha = 1/3$) provides the correct way to calculate $\sigma_U$. In this sense, microstructure invariance provides a way to test the market depth implications of Kyle (1985).

Kyle (1985) also assumes that noise trades are normally distributed with the continuous version of Kyle (1985) assuming that the order flow looks like a Brownian motion. Microstructure invariance assumes that bets are generated by something similar to a compound Poisson process. Furthermore, we show below that the estimated bet sizes are consistent with a log-normal—not a normal—distribution. Whether the execution of bets by breaking trades into many pieces results in something similar to the Brownian motion of Kyle (1985) is an interesting subject for future research.

Let $P^*, \sigma^*, \lambda^*$ denote share price, daily volatility, and price impact for a hypothetical benchmark stock which conforms to the invariance equation (20). Dividing equation (20) for some stock by equation (20) for the benchmark stock and plugging into equations (14) and (16), we obtain the percentage expected impact cost of executing $|Q|$ shares,

$$\frac{\lambda}{P} \cdot |Q| = \frac{\lambda^* \cdot V^*}{P^*} \cdot \left[ \frac{\psi}{\psi^*} \right] \cdot \left[ \frac{\zeta}{\zeta^*} \right]^{(1+\alpha)/2} \cdot \left[ \frac{W}{W^*} \right]^{1/2} \cdot \left[ \frac{\sigma}{\sigma^*} \right] \cdot |Q| V.$$

This equation leads to the empirical hypothesis that the expected market impact cost of trading a fixed percentage of daily volume $V$ accounts for a fraction of daily volatility $\sigma$ proportional to $W^{1/2}$, where the proportionality constant $[\lambda^* \cdot V^*/P^*] \cdot [\psi/\psi^*] \cdot [\zeta/\zeta^*]^{(1+\alpha)/2}$ does not vary across stocks or across time (assuming $\psi = \psi^*$ and $\zeta = \zeta^*$).

Equation (23) shows why alternative models capture the thinking behind the current literature. In the model of invariant bet frequency, the assumption $\alpha = 0$ implies that market impact is proportional to a ratio of daily volatility and daily dollar trading volume, as in Amihud (2002). In the model of invariant bet size, the assumption $\alpha = 1$ implies market impact is inversely proportional to the square root of daily dollar trading volume, as in the empirically motivated specification of price impact in Hasbrouck (2009). A difference with Hasbrouck (2009) is that the daily volatility $\sigma$ enters our formula in a non-linear manner, while it is absent from his specification. To see why these claims are true, one needs to express trading activity in equation (23) as the product of dollar volume and volatility. Similar thinking underlies specifications in Cooper, Groth, and Avera (1985) and Gabaix et al. (2006).

Equation (23) also shows why the alternative models conform to the way in which traders think about trading costs. In the model of invariant bet frequency, the market impact of a given fraction of daily volume is a constant fraction of volatility. This thinking is consistent with a practical rule of thumb of executing orders at a rate which keeps one’s participation rate in the market below a given threshold, for example, executing not more than one percent of trading volume.

**Alternative Models and Bid-Ask Spread** In order to generate predictions for bid-ask spread from the alternative models, we need to introduce a bid-ask spread invariant. Indeed, since the market impact cost $C_L$ is not constant in alternative models, the spread cost $C_K$
is unlikely to be constant either. To make a connection between market impact costs and bid-ask spread costs, we assume that the bid-ask spread cost of a bet is some fraction $\phi$ of expected market impact costs, i.e.,

$$C_K = \phi \cdot C_L.$$  \hspace{1cm} (24)

We conjecture that $\phi$ is constant across stocks and across time. Under the assumptions of trading game invariance and market impact invariance, the invariance of $\phi$ is equivalent to the invariance of the bid-ask spread cost of a bet $C_K$. Indeed, the relationship between the two variables can be easily derived from (9), (20), and (5) to be

$$C_K = \frac{1}{2} \cdot \phi \cdot \psi \cdot [E\{\tilde{I}^2\}]^{1/2}. \hspace{1cm} (25)$$

Under the alternative models, where $\alpha \neq 1/3$, $C_K$ may vary across stocks and time, even if both $\phi$ and $\psi$ remain the same. The variable $\phi$ is therefore an intuitive bid-ask spread invariant that allows us to connect our model to both alternatives.

Using equations (24), (9), and (11), we can express the bid-ask spread $\kappa$ as,

$$\kappa = \frac{\phi \cdot \lambda \cdot E\{\tilde{Q}^2\}}{\gamma^{1/2} \cdot E\{|Q|\}}. \hspace{1cm} (26)$$

Let $\kappa^*$ denote the bid-ask spread for a hypothetical benchmark stock which conforms to the invariance equation (24). Dividing equation (26) for the benchmark stock by equation (26) for another stock and plugging into equations (14) and (16), we obtain the percentage bid-ask spread cost

$$\frac{\kappa}{P} = \frac{\kappa^*}{P^*} \cdot \left[\frac{\phi}{\phi^*}\right] \cdot \left[\frac{\psi}{\psi^*}\right] \cdot \left[\frac{\zeta}{\zeta^*}\right]^{1-\alpha} \cdot \left[\frac{W}{W^*}\right]^{-\frac{(1-\alpha)}{2}} \cdot \left[\frac{\sigma}{\sigma^*}\right]. \hspace{1cm} (27)$$

This equation leads to the empirical hypothesis that the bid-ask spread cost accounts for a fraction of daily volatility $\sigma$ proportional to $W^{-(1-\alpha)/2}$, where the proportionality constant $[\kappa^*/P^*] \cdot [\phi/\phi^*] \cdot [\psi/\psi^*] \cdot [\zeta/\zeta^*]^{(1-\alpha)/2}$ does not vary across stocks or across time (assuming $\zeta = \zeta^*$, $\psi = \psi^*$ and $\phi = \phi^*$).

Summary of Predictions. Under the identifying assumptions $\zeta = \zeta^*$, $\psi = \psi^*$ and $\phi = \phi^*$, equations (16), (23), and (27) make precise nested predictions about how bet size (as a fraction of average daily volume), market impact costs (controlling for trade size and volatility) and bid-ask spread costs (controlling for volatility) vary with the level of trading activity as a function of the single “deep” parameter $\alpha$.

Below we use a database of portfolio transitions to estimate the parameter $\alpha$ in three different ways, using each of the three equations (16), (23), and (27). We do this in two steps. First, under the identifying assumption that portfolio transition orders are proportional to bets, we use the size of portfolio transition orders to test whether the predicted exponent $-(1-\alpha)$ of trading activity $W$ in equation (16) is consistent with the prediction of market microstructure invariance that $\alpha = 1/3$ against the alternative models’ predictions of $\alpha = 0$ and $\alpha = 1$. Second, using implementation shortfall to estimate trading costs as the sum of market impact costs and bid-ask spread costs, we test whether the prediction from the impact
equation (23) that the exponent of trading activity $W$ is $(1 - \alpha)/2$ and the prediction from spread equation (27) that the exponent of trading activity $W$ is $-(1 - \alpha)/2$ are consistent with the implications of market impact invariance and bid-ask spread invariance that $\alpha = 1/3$ (against the alternative models’ predictions of $\alpha = 0$ and $\alpha = 1$).

4 Data

4.1 Portfolio Transitions Data

The empirical implications of all of the three theoretical models are tested using a proprietary database of portfolio transitions from a leading vendor of portfolio transition services. This database is derived from the post-transition reports prepared by transition managers for their U.S. clients. During the evaluation period, this portfolio transition vendor supervised more than 30 percent of outsourced U.S. portfolio transitions. The sample includes 2,680 portfolio transitions executed over the period 2001 to 2005. This is the same database used by Obizhaeva (2009a, 2009b).

The portfolio transitions database contains data on individual transactions. Each observation has the following fields: a trade date, an identifier of a portfolio transition, its starting and ending dates, the name of the stock traded, the number of shares traded, buy or sell indicator, the average execution price, the pre-transition benchmark price, commissions, and fees. The data is given on separate lines for three trading venues: internal crossing networks, external crossing networks, and open market transactions. It is also given separately for each trading day in a trading package executed over several days. Old “legacy” and new “target” portfolios usually overlap. For example, both portfolios may have positions in some large and therefore widely held securities. Instead of first selling overlapping holdings from legacy portfolios and then acquiring them into target portfolios by buying them back, these positions are transferred from one account to another as “in-kind” transfers which do not incur transaction costs. Thus, if the legacy portfolio holds 10,000 shares of IBM stock and the new portfolio holds 4,000 shares of IBM in a portfolio transition, then 4,000 shares are transferred in-kind and recorded as in-kind transfers. The balance of 6,000 shares is sold. If the transition manager sells these shares in two days with open market trades on the first day and both external crosses and open market trades on the second day, then there will be four lines in the database corresponding to IBM stock for this portfolio transition: a 4,000 share in-kind transaction, an open market trade the first day, an open market trade the second day, and an external cross the second day. Our empirical results do not depend at all on in-kind transfers. Instead, our empirical results are based on open market trades, external crosses, and internal crosses.

The original data is further grouped at order level. For example, aforementioned transactions are combined into one line corresponding to the order for IBM stock in a given portfolio transition. This observation contains the name of the stock, the pre-transition benchmark price, buy or sell indicator, the number of shares executed over different trading venues, the average execution price for each of them, as well as the data on portfolio transition such as its beginning and ending dates.

The portfolio transition data are then merged with the CRSP data to add data on stock
prices, returns, and volume. Only common stocks (CRSP share codes of 10 and 11) listed on the New York Stock Exchange (NYSE), the American Stock Exchange (Amex), and NASDAQ in the period of January 2001 through December 2005 are included in the sample. ADRs, REITS, and closed-end funds were excluded. Also excluded were stocks with missing CRSP information necessary to construct variables used for empirical tests, low-priced stocks defined as stocks with prices less than 5 dollars, and transition observations which appeared to contain typographical errors and obvious inaccuracies. Since it was unclear from the data whether adjustments for dividends and stock splits were made in a consistent manner across all transitions, all observations with non-zero payouts during the first week following the starting date of portfolio transitions were excluded from statistical tests.

After exclusions, the number of daily observations was 441,865 orders (204,780 buy orders and 237,085 sell orders).

Portfolio Transitions as Bets. The three proposed models make very different assumptions about how the sizes of bets vary across stocks with different levels of trading activity. To test these different assumptions empirically, it is necessary to identify the theoretical concept of a bet \( Q \) with actual data.

It is challenging to design a reasonable test of these predictions. For empirical implementation, it is important to remember that they are formulated for independent orders \( \tilde{Q} \), arriving to the market. More precisely, one must think about \( \tilde{Q} \) as the innovation to the target holdings of all investors who got the same “trading idea.” Of course, both independent orders \( \tilde{Q} \) and their arrival rates \( \gamma \) are unobservable. In actual trading, one independent trading decision may be shared between a number of investors who submit perfectly correlated orders. Each of these orders may be broken down into smaller pieces for execution, and an execution of a trade may have several different counter-parties and prices. Thus, independent trading decisions usually generate multiple reports in the data. This complicates testing of trading game invariance using the sample of usual trades. In contrast, the dataset of portfolio transitions allows to observe entire transition orders.

If the portfolios put together by professional asset managers result from bets, then the differences in these portfolios represent the results of numerous bets in many different stocks. Therefore, we make the identifying assumption that the differences in professionally managed portfolios, while not exactly a random sample of bets themselves, vary in a manner proportional to the size of bets.

Note that the quantities traded often do not match the levels of positions in legacy and target portfolios, but rather the quantities traded match the differences in positions. When legacy and target portfolios overlap, the overlapping positions are transferred from one account to another one as “in-kind” transactions. These in-kind transactions are transfers of positions, not trades, we therefore exclude in-kind transfers from the empirical tests. As a result, our proxies for bets represent differences in portfolios across two different asset managers. We focus on transactions rather than positions because our models are designed to explain the cross-sectional differences in the execution data. The models are not meant to explain the absolute levels of holdings.
Portfolio Transitions and Implementation Shortfall. To estimate transaction costs $C(X)$, we use the concept of implementation shortfall as developed by Perold (1988). Specifically, we estimate costs by comparing the average execution prices of portfolio transition trades with closing prices the evening before the transition trades begin executing. In many applications, the use of implementation shortfall to measure transaction costs is problematic. Portfolio transitions avoid the usual problems associated with implementation shortfall and therefore provide one of the rare situations where implementation shortfall works well as a measure of transaction costs.

The fundamental problem with using implementation shortfall to measure transaction costs is that the actual quantities traded may not be known at the start date due to order cancelations or changes in trading intentions which occur after the start date and affect actual quantities traded. Statistically, the resulting selection bias problem can lead to significant underestimation of transaction costs if orders tend to be either canceled when prices move in an unfavorable direction or increased when prices move in a favorable direction. Implementation shortfall can also lead to biased estimates of transaction costs if the trading decisions are based on short-lived private information which is incorporated into prices during the period when the trades occur. Portfolio transition data has several important properties which make it particularly advantageous for estimating transaction costs using implementation shortfall.

For each stock in a portfolio transition, the quantities to be traded are known precisely at a specific time before the trades are actually executed. The composition of legacy and target portfolios is fixed in the mandates that transition managers receive the night before portfolio transitions begin. These managers then execute orders regardless of the unfolding price dynamics. This makes it reasonable to assume that the initial orders or trading intentions are exactly equal to the quantities subsequently traded. Thus, portfolio transition data tend not to be affected by the selection bias problem that would affect databases of trades where the quantities traded change in a manner correlated with price changes between the time orders are placed and the time they are executed, canceled, or increased.

The timing of portfolio transitions is likely determined by a schedule of investment committee meetings of institutional sponsors, who make decisions to undertake transitions. The investment committee meets regularly on schedules set well in advance of the meetings. Among the issues boards discuss are the replacement of fund managers and the changes of asset mix. If a decision is made to replace a portfolio manager, then a portfolio transition is arranged shortly after the meeting. These decisions are unlikely to be correlated with short-term price dynamics of individual securities during the period of the transition. This makes it possible to obtain estimates of price impact and spread that are not affected by short-lived information likely to be incorporated into prices during the period the transition trades are executed.

These properties of portfolio transitions are not often shared by other data. Consider a database built up from trades by a mutual fund, a hedge fund, or a proprietary trading desk at an investment bank. In such samples, the trading intentions of traders may not be recorded in the database. For example, a database might time stamp a record of a trader placing an order to buy 100,000 shares of stock but not time stamp a record of the trader's secret intention to buy another 200,000 shares after the first 100,000 shares is bought. Furthermore, trading intentions before traders begin trading may not coincide with
realized trades because the trader changes his mind as market conditions change. Traders often condition their trading strategies on prices by using limit orders or by canceling parts of their orders, thus hard-wiring into their strategies a selection bias problem for using such data to estimate transaction costs. The trading intentions themselves can be significantly affected by overall price dynamics, e.g., traders may be following trends or playing contrarian strategies. This dependence of actually traded quantities on prices, consequently, makes it impossible to use implementation shortfall in a meaningful way to estimate market depth and bid-ask spreads from data on trades only.

Portfolio transitions also have properties that may distort our results. Indeed, we implicitly assume that bets are executed in a “natural” trading-game time, with bets in active stocks being executed faster than bets in inactive stocks. Portfolio transitions are usually unwound over a short period of time. If short-term deadlines make transition managers significantly speed up executions of positions in inactive stocks comparing to those in active stocks, then the implementation shortfall for inactive stocks will appear to be “too high.” We will revisit this concern when discussing our tests based on the implementation shortfall in Section 6.

\section{4.2 CRSP Data: Prices, Volume, and Volatility}

Our three models use trading activity to explain how transaction costs and expected trade size vary across stocks. Trading activity is the product of trading volume (in shares), share price (in dollars), and volatility (percentage standard deviation of daily returns). To measure implementation shortfall, a pre-trade benchmark price is needed. The components of trading activity and the pre-trade benchmark are calculated from CRSP data.

As a pre-trade price, denoted \( P_{0,i} \) for \( i \)th order, we use the closing price of the corresponding security on the evening before the portfolio transition trades begin. More precisely, a portfolio transition involves trades in numerous stocks. Typically, many of the stocks are traded on the first day of the transition. For each stock in the transition, the benchmark price \( P_{0,i} \) is the closing price the evening before the first trade is made in any of the stocks in the portfolio transition, even if a particular stock itself is not traded on the first day.

As expected trading volume during portfolio transitions, denoted \( V_i \) for \( i \)th order, we use the average daily trading volume (in the number of shares) of the corresponding security in the pre-transition month.

We estimate the expected volatility of daily returns, denoted \( \sigma_i \) for \( i \)th order, using past daily CRSP returns for the stock involved in the \( i \)th trade. We use two different estimates of volatility, a simple estimate equal to average daily volatility from the past month and a more complicated estimate from an ARIMA model.

For each security, we first calculate the monthly standard deviation of returns from daily CRSP returns data. Let \( r_{i,t,k} \) denote the CRSP return for the \( k \)th day of month \( t \) for stock involved in the \( i \)th trade. Letting \( N_{i,t} \) denote the number of CRSP trading days in month \( t \), then the standard deviation for month \( t \) for the stock in \( i \)th trade, denoted \( \sigma^m_{i,t} \), is

\[
\sigma^m_{i,t} = \left( \sum_{k=1}^{N_{i,t}} r_{i,t,k}^2 \right)^{1/2}
\]  

(28)
We do not de-mean the returns data since the mean return in a month is very small relative to the standard deviation. We also do not adjust the estimates for autocorrelation of returns by adding a cross-product of adjacent returns, since this might result in negative estimates of volatility for some stocks.

One simple estimate of daily volatility for the stock in trade $i$ for month $t$, denoted $\sigma_{i,t}^h$, is the monthly standard deviation converted to daily units:

$$\sigma_{i,t}^h = \frac{1}{N_{i,t}^{1/2}} \sigma_{i,t}^m.$$  \hspace{1cm} (29)

We also estimate an ARIMA model to obtain another forecast of the daily return standard deviations for each stock $j$ and month $t$. To reduce effects from the positive skewness of the standard deviation estimates, we use a logarithmic transformation for the volatility. We estimate a third-order moving average process for the changes in $\ln \sigma_{i,t}^m$ over the whole sample from 2001 to 2005:

$$(1 - L) \ln \sigma_{i,t}^m = \Theta_0 + (1 - \Theta_1 L - \Theta_2 L^2 - \Theta_3 L^3) u_t$$  \hspace{1cm} (30)

The conditional forecast for the volatility of daily returns is

$$\sigma_{i,t}^e = \frac{1}{N_{i,t}^{1/2}} \exp \left[ \ln \sigma_{i,t}^m + \frac{1}{2} \hat{V}(u) \right]$$  \hspace{1cm} (31)

where $\hat{V}(u)$ is the variance of the prediction errors of the ARIMA model.

In the empirical tests below, both $\sigma_{i,t-1}^e$ and $\sigma_{i,t-1}^h$ are used as proxies for $\sigma_i$ in the $ith$ transition trade. It is possible that using these proxies in our regressions may introduce an error-in-variables problem due to the volatility estimates themselves having errors. The empirical results are quantitatively similar for both proxies. Thus, only results for the estimates based on $\sigma_{i,t-1}^e$ are reported. We use the pre-transition variables known before portfolio transition trades in order to avoid any spurious effects from using contemporaneous variables, except to the extent that the ARIMA model uses in-sample data to estimate model parameters.

### 4.3 Descriptive Statistics

Table 1 reports statistical characteristics of both securities traded and individual transition trades. Statistics are calculated for all securities in aggregate as well as separately for ten groups of stocks sorted by average daily dollar volume. Instead of dividing the securities into ten deciles with the same number of securities, volume break points are set at the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of trading volume for the universe of stocks listed on the NYSE with CRSP share codes of 10 and 11. Group 1 contains stocks in the bottom 30th percentile by dollar trading volume. Group 10 contains stocks in the top 5th percentile. Smaller percentiles for the more active stocks make it possible to focus on the stocks which are most important economically. For each month, the thresholds are recalculated and the stocks are reshuffled across bins.
Panel A of table 1 reports statistical properties of the securities in the sample. There is a column for each of the ten groupings as well as a column which reports aggregate statistics. For the entire sample of stocks, the median trading volume is $19.99 million per day, ranging from $1.22 million for the lowest volume decile to $212.55 million for the highest volume decile. Since the average dollar volume ranges over more than two orders of magnitude, this variation in the data should create statistical power helpful in determining how transaction costs and trade size vary with dollar volume. Panel A reports that the median volatility for all stocks is a standard deviation in returns of 1.85 percent per day. Volatility tends to be slightly higher in the lower volume deciles than the higher ones. The volatility for the lowest volume decile is 2.04 percent, and it is 1.76 percent for the highest volume group. This implies that the amount of variation in volatility across stocks is somewhat small.

Panel A reports that the median bid-ask spread, a quoted spread obtained from the transition database, is 11.54 basis points. Its mean is 23.67 basis points. From lowest volume grouping to highest volume grouping, the median bid-ask spread declines monotonically across groups from 38.16 basis points in the lowest volume group to 4.83 basis points in the highest volume group. This monotonic decline of almost one order of magnitude in reported bid-ask spreads is so large that significant statistical power should be generated to differentiate the predictions of the three models for bid-ask spreads. This, of course, assumes that the spreads reported in Panel A, which are quoted spreads not estimated from implementation shortfall, also show up in statistical estimates based on implementation shortfall.

For example, our proposed model of market microstructure invariance predicts that spreads should decrease 1/3 of one percent for each increase of one percent in trading volume, holding volatility constant. From lowest to highest decile, volume increases by a factor of $212.55/1.22 = 174.22$. A back-of-the-envelope prediction for the decrease in spreads across these groups is 1/3 power of the increase in volume, i.e., $174.22^{1/3} = 5.58$. The actual decrease in spreads is a factor of $38.16/4.83 = 7.90$. While this back-of-the-envelope calculation suggests that spreads decrease more than our model predicts, the difference between 5.58 and 7.90 is small enough to warrant further statistical investigation. It is possible that the estimates of effective spreads obtained from implementation shortfall are different from quoted spreads, and thus results based on implementation shortfall will be somewhat different.

Panel B of table 1 reports properties of order sizes in the portfolio transition data. The mean portfolio transition order is 3.90 percent of average daily volume of the stock traded. The means decline monotonically across the ten volume groups from 15.64 percent in the smallest group to 0.49 percent in the largest. The median portfolio transitions order is 0.56 percent of average daily volume. The median also declines monotonically across the ten volume groups, from 3.48 percent in the smallest to 0.14 percent in the largest. The fact that the medians are much smaller than the means indicates that the order size is skewed to the right. This is to be expected, since the order size is a non-negative number, and there may be some very small trades from highly diversified portfolios involving smaller transitions as well as very large trades from less diversified portfolios involving larger transitions. Below we show that the distribution of order size is very close to a log-normal distribution, for which the medians are smaller than the means.

The significant variation in mean and median order size as a fraction of average daily volume across dollar volume deciles is expected to have several important effects on statistical
estimates.

On the one hand, the larger order size in the lower deciles generates more statistical power for using implementation shortfall to estimate market impact in the lower deciles than in the higher ones (holding constant market impact of an order with the size being a constant percentage of trading volume). On the other hand, our proposed model predicts that price impact increases as volume increases, holding order size as a percentage of volume constant. Of the other two models, one predicts no change in impact while the other predicts an even larger increase. Since each of the three models makes very different predictions concerning how market impact varies with trading activity, all three models will try to extrapolate the statistical power from one volume group to another, but the extrapolation will operate differently for each model.

Second, the variation in the order size across volume groups makes it possible to test the assumptions of the three models concerning how the size of liquidity trades varies across stock with different trading activity levels. From highest to lowest group, daily volume increases by a factor of $212.55/1.22 = 174.22$. According to our proposed invariance hypothesis, average trade size as a percent of volume should decrease by $2/3$ of one percent for every one percent increase in volume, holding volatility constant. As a back-of-the-envelope calculation, this implies that the decrease in order size from lowest to highest decile should be $2/3$ power of the increase in dollar volume, i.e., $174^{2/3} = 31.2$. The actual median order size decreases as a fraction of average volume by a factor of $3.48/0.14 = 24.86$, from lowest group to highest group. While the back-of-the-envelope calculation of 31.2 does not exactly match the factor of 24.86, the numbers are close enough to suggest that further statistical investigation is warranted.

Table 1 reports that, on average, portfolio transitions represent 0.29% of overall monthly trading volume. They account for 0.41% of trading in small stocks and 0.13% of trading in large stocks. This is consistent with the findings of Blume and Keim (2010) that recently institutions have significantly increased their holdings in small stocks and now tend to underweight the largest stocks relative to the smaller stocks. Since the order size as a fraction of volume drops by a factor of 24.86 and the contribution to overall volume decreases only by a factor of $0.41/0.13 = 3.15$ as the volume increases, the implied number of transition orders has to increase by a factor of 8. The number of orders seems to increase with volume more slowly than a factor of $174^{2/3} = 31.2$ predicted by invariance hypothesis.

## 5 Empirical Results: Order Size

The three theoretical models make distinctly different predictions concerning how bet sizes vary with the level of activity. We use portfolio transition orders to test these implications.

**Identifying Assumptions.** We make two identifying assumptions in our tests.

Our first identifying assumptions is that the that volume multiplier $\zeta$ is constant across stocks and across time. For simplicity of exposition, we interpret our results under the assumption that $\zeta = 2$, i.e., each portfolio transition order is intermediated by a specialist without involvement of other intermediaries. Of course, to the extent that intermediation is more significant in active stocks relative to inactive stocks, the deflator can potentially vary
with trading activity. This variable is fundamentally unobservable in our data; its calibration
for different stocks is an interesting issue for future research.

Our second identifying assumption is that portfolio transition trades are proportional in
size to bets. Let $X_i$ denote the unsigned number of shares transacted in a given security
during a given portfolio transition (excluding shares transferred in-kind). The index $i$ ranges
across 441,685 stock-transition pairs. We assume that the distribution of portfolio transition
orders $X_i$ differs from the distribution of typical bets by a constant factor $\theta$. We express this
identifying assumption informally as

$$|\tilde{Q}| \approx \theta \cdot X_i. \quad (32)$$

While it is possible that our portfolio transition orders may be on average larger or smaller
than bets, we believe it is reasonable to assume that the factor $\theta$ by which they may be
larger or smaller than bets does not vary across stocks with higher and lower levels of
trading activity. Our second identification assumptions is that $\theta$ is constant across stocks
and across time. For simplicity of exposition, we interpret our results under the assumption
that $\theta = 1$, i.e., transition orders are drawn from the same size distribution as bets.

Portfolio transitions are examples of institutional trades. Since it is known that institu-
tional investors used to trade in big stocks more intensively than retail investors and used
to avoid trading in small stocks, one may suspect that transition orders are not representa-
tive of bets in a market where large institutions and small retail investors are both making
bets. Blume and Keim (2010) report, however, that institutions started to switch towards
trading in small stocks in 1990s, and the trading activity of institutional and retail investors
had somewhat balanced across markets during the 2001-2005 time period of this study. In-
vestigating further the relationship between portfolio transition trades and others empirical
proxies for bets is an interesting topic for further research.

5.1 Invariant Order Size Distribution as Log-Normal

Trading game invariance predicts that the distribution of order sizes $X_i$ is invariant across
stocks when the mean is adjusted to reflect differences in trading activity. Plugging equation
(13) for the benchmark stock with trading activity $W^*$ and equation (32) into equation (16),
we get,

$$\ln \left( \frac{X_i}{V_i} \right) + (1 - \alpha) \cdot \ln \left( \frac{W_i}{W^*} \right) \approx E\{ \ln \left( \frac{|\tilde{Q}|}{V^*} \right) \} - \alpha \cdot \ln \left( \frac{\zeta}{\zeta^*} \right) - \ln(\theta) + \tilde{\epsilon}, \quad (33)$$

where $\tilde{\epsilon}$ is the zero mean random variable $\ln(|\tilde{I}|) - E\{\ln(|\tilde{I}|)\}$. Trading game invariance
 corresponds to the hypothesis $\alpha = 1/3$ in equation (33).

The invariance can be seen by observing that the right-hand-side of equation (33) does
not vary across portfolio transition trades because of the invariance of the distribution of
$\tilde{I}$ and the identifying assumptions that neither the volume multiplier $\zeta$ nor the portfolio
transition size multiplier $\theta$ vary across portfolio transition trades. We also examine the
additional hypothesis—not implied by microstructure invariance—that the left-hand side of
equation (33) is normally distributed, implying an invariant log-normal distribution for order
size adjusted for trading activity.
To examine this hypothesis visually, figure 2 shows the empirical distribution of the invariant distribution in equation (33) for selected volume deciles and volatility quintiles for 400,000+ portfolio trades in stocks sorted into ten volume groups and five standard deviation groups. As before, volume groups are based on average dollar trading volume with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar volume. Standard deviation groups are quintiles based on thresholds corresponding to the 20th, 40th, 60th, and 80th percentiles. These thresholds are obtained for the sample of NYSE-listed common stocks. On each plot, we superimpose the bell-shaped density function of a normal distribution with mean -5.69 and variance 2.50 matching the pooled mean and variance of the invariant distribution (33) across all volume deciles and volatility quintiles. If trading game invariance describes the data well and order sizes are distributed as a log-normal, then these empirical density functions should appear to be almost identical to a normal distribution with common mean and variance.

Figure 2 shows a plot of the empirical distributions for volume groups 1, 4, 7, 9, and 10 and for volatility groups 1, 3 and 5. These 15 distributions of \( \overline{W} \)-adjusted order sizes are strikingly similar. Results for the remaining 35 subgroups also look very similar and therefore are not presented in this paper. The visual similarity of distributions is consistent with the similarity of their first four moments. The means range from -6.06 to -5.42, close to the mean of -5.69 for the entire sample. The variances range from 2.19 to 2.82, also close to the variance of 2.50 for the entire sample. The skewness ranges from -0.21 to 0.05, close to skewness of zero for the normal distribution. The kurtosis ranges from 2.66 to 3.48, also close to the kurtosis of 3 for a normal random variable. These results suggest that the assumption that order sizes are distributed as log-normal random variables is a reasonable one. Moreover, imposing on ln(\( W/W^* \)) a coefficient of \(- (1 - \alpha) = -2/3\) based on the hypothesis of trading game invariance (\( \alpha = 1/3 \)) adjusts the means of the distributions so that they appear visually to be very similar.

Despite the visual similarity, a Kolmogorov-Smirnov test rejects the hypothesis that all fifty empirical distributions are generated from the same normal distribution. The standard deviation of the means across bins is larger than implied by a common normal distribution. Indeed, if we assume that all 441,865 observations are spread between fifty groups uniformly, then there will be about 9,000 observations in each bin. For a normal distribution with the mean of -5.69 and the variance of 2.50, the standard error of the mean is equal to about 5.69/(\( \sqrt{2.50 \cdot \sqrt{9,000}} \)) = 0.04. This number is much smaller than the standard deviation of about 0.15 that we observe in the data. This reveals additional variation in order sizes relative to the model of trading game invariance. Of course, we do not claim that our model perfectly describes the data. Rather it seems to capture the first-order effect, making it worthwhile to investigate more carefully why the data is different from the model and what explains these differences.

Figure 3 further examines the log-normality of distributions focusing on their tails. Panel A shows the quantile-quantile plots for the empirical distributions of log-order sizes based on trading game invariance (\( \alpha = 1/3 \)) versus a normal distribution for different volume groups. The more similar the empirical distributions is to a normal distribution, the closer the plot should correspond to the 45-degree. Panel B depicts the logarithm of ranks based on re-scaled order sizes. Both panels show that the empirical distributions are very similar to normal distributions, except in the positive and negative tails of the distributions.
The smallest orders in the left tails tend to be smaller than implied by a normal distribution. These observations are, however, economically unimportant. A closer examination reveals that most of them represent one-share transactions in low-price stocks (perhaps the result of coding errors in the data). Furthermore, there are too few such orders to have a meaningful effect on our statistical results.

The largest orders in the right tails are much more important economically but not statistically. On each subplot, there are a handful of positive outliers (out of 400,000+ observations) that do not appear fit a normal distribution almost perfectly based on visual inspection. The largest orders in low-volume stocks seem to be smaller than implied by a normal distribution, and the largest orders in high-volume stocks seem to be larger than implied by a normal distribution. Finding explanations for these deviations remains an interesting question.

The finding that the largest orders in low-volume stocks are smaller than implied by a log-normal small can be explained by market regulations. According to Section 13(d) of the 1934 Act and Regulation 13D, institutions are required to report their holdings whenever they own more than 5 percent of publicly traded companies. Some small companies may also have adopted “poison-pill” provisions with triggers being set much lower than the typical levels of 15%. To avoid reporting, large institutional investors may adjust their holdings downward by acquiring fewer shares when holdings would otherwise exceed the 5% reporting threshold. Indeed, all 400,000+ portfolio transition orders are for amounts smaller than 4.5% of shares outstanding. A closer examination reveals that the five largest orders for low-volume stocks accounts for 2%, 3%, 4%, 4%, and 4% of shares outstanding, respectively. Based on a log-normal distribution, in the sample of 65,081 observations in the first low-volume group, several orders should be larger than 5%. Indeed, if the median order size as a fraction of shares outstanding is one basis point (from table 1) and the variance of the normal variable underlying a log-normal distribution of the invariant ˜I is about 2.50, then a four-standard deviation order will be equal to 6.25% of shares outstanding (one basis point times exp(4·√2.50) = 625). In expectation, there should be at least two observations larger than 6.25% in the low-volume group. One reason we do not observe these large observations in our sample may be that institutions deliberately curtail their demand to avoid regulatory reporting requirements.

In summary, we conclude that when order size is adjusted according to the prediction of trading game invariance in (33), then order sizes appear to have a size distribution quite similar to—but not exactly equal to—a log-normal distribution. Next, we report the results of several econometric tests.

**OLS Estimates of Order Size.** Defining the regression coefficient α₀ by α₀ = −(1 − α), the order size predictions from equation (33) can be expressed in a simple log-linear OLS regression

\[
\ln \left[ \frac{X_i}{V_i} \right] = \ln \left[ \tilde{q} \right] + α₀ \cdot \ln \left[ \frac{W_i}{W^*} \right] + \tilde{\epsilon},
\]

where the constant term \( \ln(\tilde{q}) \) is given by \( \ln(\tilde{q}) = E\{\ln(|\tilde{Q}^*|/V^*)\} - α \ln(\zeta/\zeta^*) - \ln(θ) \) and where \( \tilde{\epsilon} \) denotes a zero-mean random variable \( \ln(|\tilde{I}|) - E\{\ln(|\tilde{I}|)\} \) drawn from the same distribution for all portfolio transition trades. Under the log-normality assumption and the
identifying assumptions that $\theta$ and $\zeta$ are constants, $\bar{q}$ represents the median order size for the benchmark stock as a fraction of daily volume. This equation relates the size of the trade $X_i$ as a fraction of average volume $V_i$ to the level of trading activity $W_i$, defined as the product of benchmark price $P_{0,i}$, last month’s trading activity $V_i$, and estimated volatility $\sigma_i$:

$$W_i = P_{0,i} \cdot V_i \cdot \sigma_i.$$  \hspace{1cm} (35)

The scaling constant $W^* = (40)(10^6)(0.02)$ measures trading activity for a hypothetical arbitrary benchmark stock with price of $40 per share, trading volume of one million shares per day, and volatility of 2 percent per day. Table 1 implies that this stock would belong to the bottom tercile of S&P 500. The scaling constant has no effect on estimates for $\alpha_0$ but does affect the meaning of the constant term $\ln(\bar{q})$.

We estimate $\ln(\bar{q})$ and $\alpha_0$. We have replaced $\alpha - 1$ by $\alpha_0$ to emphasize the distinction from other tests of $\alpha$ based on market impact and bid-ask spread. The model of market microstructure invariance ($\alpha = 1/3$) predicts $\alpha_0 = -2/3$, the model of invariant bet frequency ($\alpha = 1$) predicts $\alpha_0 = 0$, and the model of invariant bet size ($\alpha = 0$) predicts $\alpha_0 = -1$. Since the three models make dramatically different predictions concerning $\alpha_0$, it should be possible to test the models by estimating this parameter.

A potential econometric difficulty with the specification in equation (34) is that taking the log of trade size as a fraction of average daily volume has the potential to create large negative outliers out of tiny, economically meaningless trades, with the inordinately large influence on reported results. Since we have shown in the previous section that the shape of the distribution of re-scaled order sizes closely match a log-normal distribution, however, the error term in the regression above has a distribution close to a zero-mean normal distribution; therefore, tiny orders have a negligible distorting effect on estimates.

Table 2 presents estimates for the OLS coefficients in equation (34). The first column of the table reports the results of a regression pooling all the data. The four other columns in the table report results for four separate OLS regressions in which the four parameters are estimated separately for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells.

The estimate for $\alpha_0$ is $\hat{\alpha}_0 = -0.63$ with standard error of 0.008. Economically, the point estimate for $\alpha_0$ is close to the value predicted by the model of market microstructure invariance $\alpha_0 = -2/3$, but this model is strongly rejected ($F = 17.03, p < 0.0001$) because the standard error is very small. This point estimate is so different from the predictions of the two other models that they are rejected by overwhelming margins.

When sample is broken down into NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, the point estimates $-0.63, -0.60, -0.71, and -0.61$ are consistently close to the value of $-2/3$ predicted by trading game invariance ($\alpha = 1/3$), but the model of trading game invariance is rejected in all cases due to the low standard error.

**Quantile Estimates of Order Sizes.** So far, our tests have provided evidence that the invariance hypothesis holds reasonably well for the means of distributions. There is also evidence that it holds for various percentiles as well. Table 3 shows the estimates for the coefficients in the quantile regressions in equation (34). The estimate for $\alpha_0$ varies, almost monotonically, from $-0.69$ for the 1st percentile to $-0.62$ for the 99th percentile. All estimates are economically close to the value of $-2/3$ predicted by trading game invariance.
The decrease in the absolute value of coefficients across percentiles is consistent with the hypothesis that a 100-share round lot trades size is a more binding constraint for inactive stocks than active stocks. This is consistent with table 1, which shows that the percentage of trades for 100 shares or less declines across volume groups from 5% to 1%.

**Separate Coefficients for Price, Volume, and Volatility.** The model of trading game invariance predicts that bet size is explained by trading activity. Time clock irrelevance and Modigliani-Miller irrelevance suggest that the individual components of trading activity—price, volume, and volatility—should have no additional explanatory power beyond the explanatory power provided by trading activity \( W \).

To test whether the individual components of trading activity—price, volume, and volatility—explain bet size, table 4 presents results for the OLS regression

\[
\ln \left( \frac{X_i}{V_i} \right) = \ln(\bar{q}) - \frac{2}{3} \ln \left( \frac{W_i}{W_s} \right) + b_1 \cdot \ln \left( \frac{\sigma_i}{(0.02)} \right) + b_2 \cdot \ln \left( \frac{P_{0,i}}{(40)} \right) + b_3 \cdot \ln \left( \frac{V_i}{(10^6)} \right) + \epsilon. \tag{36}
\]

This regression imposes on \( \ln(W_i/W_s) \) the coefficient \( \alpha_0 = -2/3 \) predicted by the model of market microstructure invariance. It then allows the coefficient on the three components of \( W_i \) to vary freely. Thus, the model of trading game invariance predicts \( b_1 = b_2 = b_3 = 0 \). The model of invariant bet frequency predicts \( b_1 = b_2 = b_3 = 2/3 \), and the model of invariant bet size predicts \( b_1 = b_2 = b_3 = -1/3 \).

A comparison of table 4 with table 2 shows that increasing the number of parameters estimated from two to four increases the adjusted \( R^2 \) of the regression from 0.3188 to 0.3211.

The table reports point estimates for the coefficient on volatility of \( \hat{b}_1 = 0.25 \), the coefficient on price of \( \hat{b}_2 = 0.16 \), and the coefficient on share volume of \( \hat{b}_3 = 0.01 \), with corresponding standard errors of 0.031, 0.014, and 0.009, respectively (t-values of 8.17, 11.05, and 0.86). While the regression fails to reject the hypothesis \( b_3 = 0 \) (thus supporting the model of microstructure invariance), the coefficients on volatility and price are significantly positive, indicating that trade size, as a fraction of average daily volume, does not decrease with increasing volatility and volume as fast as predicted by the model of microstructure invariance. This is consistent with the intuition that the 100-share trade size minimum might impose a binding constraint on some orders.

**Comparison of Three Models.** Table 5 estimates the constant term in the regression under the assumption that the coefficient \( \alpha_0 \) of \( \ln(W_i) \) in equation (34) is fixed at the values implied by the three models. For the model of market microstructure invariance, fixing the coefficient at \( \alpha_0 = -2/3 \) results in a constant term estimate of log order size as a fraction of average daily volume equal to \( \ln(\bar{q}) = -5.69 \). We have seen that the distribution of order sizes is close to being log-normal. With a log-normal distribution, the log of the median is the mean of the log. This implies a median order size of 33.75 basis points of volume, or 0.3375% of average daily volume, for the benchmark stock.

For each sample, the reported maximum likelihood function is many orders of magnitude greater for the model of microstructure invariance than for the two alternative models. For prior probabilities assigning a non-trivial probability to the model of market microstructure invariance, a Bayesian posterior distribution will assign probabilities of almost 100% to the proposed model and conclude that alternative models are highly unlikely.

28
Economic Interpretation. From the perspective of meaningful economic magnitudes, the $R^2$ results in tables 2, 4, and 5 provide strong support for the model of microstructure invariance. When the coefficient on $\ln(W/W^*)$ is fixed at the model-implied value of $\alpha_0 = -2/3$ and only one parameter (constant term) is estimated, we obtain an adjusted $R^2$ of $R^2 = 0.3177$. By contrast, when the values of $\alpha_0$ implied by the two alternative models are fixed at either $\alpha_0 = 0$ or $\alpha_0 = -1$, then we obtain adjusted $R^2$ values of $R^2 = 0.0000$ (by definition) and $R^2 = 0.2105$, respectively. Clearly, the model of microstructure invariance has superior explanatory power. The results reflect the fact that portfolio transition orders for less active stocks tend to be for a much greater percentage of average daily volume than portfolio transition orders for more active stocks, in a manner consistent with the hypothesis $\alpha = 1/3$.

When the parameter $\alpha_0$ is estimated rather than held fixed, changing $\alpha_0$ from the model value of $\alpha_0 = -2/3$ to the estimated value of $\hat{\alpha}_0 = -0.63$ increases the adjusted $R^2$ only modestly, from $R^2 = 0.3177$ reported in table 5 to $R^2 = 0.3188$ reported in table 2. Although statistically significant, the addition of the extra variable does not add much explanatory power. When the specification is further relaxed to estimate coefficients on price, volume, and volatility separately, the addition of two extra parameters again increases the adjusted $R^2$, from $R^2 = 0.3188$ reported in table 2 to $R^2 = 0.3211$ reported in table 4. Although statistically significant, the improvement in $R^2$ is again modest. We interpret these results as implying that portfolio transition order size as a fraction of average daily volume varies across stock with different levels of trading activity in a manner consistent with our model of microstructure invariance.

**Dummy Variables as a Robustness Check for Linearity.** Figure 4 uses ten dummy variables for volume groups to estimate the three versions of the regression

$$
\ln \left[ \frac{X_i}{V_i} \right] = \sum_{j=1}^{10} \mathbb{1}_{j;i} \cdot \ln \left[ \bar{\tilde{q}}_j \right] + \alpha_0 \cdot \ln \left[ \frac{W_i}{W^*} \right] + \tilde{\epsilon},
$$

in which the value of $\alpha_0$ is fixed at the value predicted by the three theoretical models, and the ten dummy variables are estimated. Dummy variables $\bar{\tilde{q}}_j, j = 1, \ldots, 10$ quantify the average size of liquidity order for the benchmark stock based on data for $j$th volume group. If the linear regression is well-specified, then the values of the dummy variables resulting from fixing $\alpha_0$ at the corresponding level should be constant across volume groups. In figure 4, the ten dummy variables are plotted, along with their 95% confidence bounds. The value of the constant term from the one-parameter regression is shown as a horizontal line. If the regression is well-specified, the values of the dummy variables $\bar{\tilde{q}}_j$ should line up along the horizontal line.

In the first graph in the figure, the ten dummy variables resulting from fixing $\alpha_0 = -2/3$, as implied by the model of microstructure invariance, are plotted. Visually, it can be seen that these dummy variables line up nicely along a horizontal line. Upon close inspection however, it is possible to notice that the 95% confidence bounds are so narrow that some of the points lie outside the 95% confidence bound.

In the second graph in the figure, the ten dummy variables resulting from fixing $\alpha_0 = 0$, as implied by the model of invariant trade frequency, are plotted. Instead of lining up on the
horizontal line, the dummy variables decline monotonically from a level very far above the line to a level very far below it, far outside the 95% confidence bounds. This is consistent with strong rejection of this model.

In the third graph in the figure, the ten dummy variable resulting from fixing \(\alpha_0 = -1\), as implied by the model of invariant trade frequency, are plotted. Instead of lining up on the horizontal line, the dummy variables increase monotonically from a level far below the line to a level far above it, far outside the 95% confidence bounds. This is consistent with strong rejection of this model as well.

What is the intuition underlying these patterns? Why does the model of invariant bet frequency significantly overestimate trade size for high-volume stocks? Why does the model of invariant bet size significantly underestimate it? The reason is that the first model incorrectly attributes large trading volume entirely to large bet size whereas the second model mistakenly explains this volume entirely by high bet frequency. If our model of microstructure invariance is true, then 2/3 of this volume comes from higher frequency of bets placed and 1/3 from their larger size. Note that consistent with the two-to-one ratio between bet frequency and bet size, the deviation from a horizontal line is twice as large for the first alternative model as for the second one.

The data on the average size of portfolio transition orders support market microstructure invariance and soundly reject the two alternative models. The reason for the rejection is that variations in trading activity are associated with variations in both bet frequency and bet size; neither remains constant.

6 Empirical Results: Market Impact and Bid-Ask Spread

The three theoretical models make distinctly different predictions concerning how transaction costs vary with the level of trading activity. We test these predictions using the data on implementation shortfall from executions of portfolio transition trades.

**Nested Log-Linear Specification.** The three theoretical models make predictions about market impact cost and bid-ask cost in (23) and (27). Let \(X_i\) denote the unsigned quantity traded. Let \(C(X_i)\) denote the transaction costs, measured in basis points per share and associated with trading the quantity \(X_i\), with market impact costs and spread costs both included. Both predictions can be then conveniently expressed in a simple non-linear form that relates implementation shortfall to levels of trading activity \(W_i\):

\[
C(X_i) = \frac{1}{2} \bar{\lambda} \cdot \frac{\sigma_i}{\sigma^*} \left( \frac{W_i}{W^*} \right)^{\alpha_1} \cdot \frac{X_i}{(0.01)V_i} + \frac{1}{2} \bar{\kappa} \cdot \frac{\sigma_i}{\sigma^*} \left( \frac{W_i}{W^*} \right)^{\alpha_2}.
\]

(38)

The first term on the right-hand side measures the linear component of transaction cost due to market impact, and the second term measures the fixed component of transaction costs due to bid-ask spread. We have replaced \((1 - \alpha)/2\) by \(\alpha_1\) and \(- (1 - \alpha)/2\) by \(\alpha_2\) to emphasize the difference of tests based on market impact and bid-ask spread from tests based on order sizes. The model of market microstructure invariance (\(\alpha = 1/3\)) predicts \(\alpha_1 = 1/3\) and
\( \alpha_2 = -1/3 \), the model of invariant bet frequency (\( \alpha = 1 \)) predicts \( \alpha_1 = 0 \) and \( \alpha_2 = 0 \), and the model of invariant bet size (\( \alpha = 0 \)) predicts \( \alpha_1 = 1/2 \) and \( \alpha_2 = -1/2 \).

The variables \( \bar{\lambda} \) and \( \bar{\kappa} \) are defined as

\[
\bar{\lambda} = \left[ \frac{\lambda^* \cdot 0.01 \cdot V^*}{P^*} \right] \cdot \left[ \frac{\psi/\psi^*}{\theta/\theta^*} \right]^{-1} \cdot \left[ \frac{\zeta/\zeta^*}{\theta/\theta^*} \right]^{1+\alpha} \cdot 10^4, \tag{39}
\]

\[
\bar{\kappa} = \left[ \frac{\kappa^*}{P^*} \right] \cdot \left[ \frac{\phi/\phi^*}{\psi/\psi^*} \right] \cdot \left[ \frac{\zeta/\zeta^*}{\psi/\psi^*} \right]^{1-\alpha} \cdot 10^4, \tag{40}
\]

The variable \( \bar{\lambda}/2 \) quantifies the market impact costs, measured in basis points, associated with trading one percent of average daily volume in the benchmark stock. The impact parameter \( \lambda \) is divided by 2 based on the intuition that the bet is divided into many small trades executed as the trader moves up along a linear demand or supply schedule. The variable \( \bar{\kappa}/2 \) quantifies the percentage bid-ask spread cost for the benchmark stock as a half-spread, also measured in basis points. Note that \( \bar{\lambda} \) is different from \( \lambda \) in that the units have been changed to basis points (cents per hundred dollars) rather than dollars per share per share. Similarly, \( \bar{\kappa} \) is different from \( \kappa \) in that units have been changed to basis points from dollars per share. In each model, under the identifying assumptions \( \theta = \theta^* \), \( \zeta = \zeta^* \), impact cost invariance \( \psi = \psi^* \), and spread cost invariance \( \phi = \phi^* \), the variables \( \bar{\lambda} \) and \( \bar{\kappa} \) do not vary across stocks or across time.

Equation (38) is a non-linear function of the four parameters \( \bar{\lambda} \), \( \bar{\kappa} \), \( \alpha_1 \), and \( \alpha_2 \), all of which can be estimated. For a trade in the benchmark stock equal to one percent of average daily volume, the two parameters \( \bar{\lambda} \) and \( \bar{\kappa} \) represent the market impact and bid-ask spread in basis points. The two remaining parameters, the exponents \( \alpha_1 \) and \( \alpha_2 \), describe how the models extrapolate market impact and spread costs across stocks with different levels of activity. Since the three models make dramatically different predictions concerning \( \alpha_1 \) and \( \alpha_2 \), it should be possible to test the models by estimating all four parameters.

In particular, we can make the identifying assumption that, in a correctly specified model, the implementation shortfall from the portfolio transition database is an unbiased estimate of the transaction cost \( C(X) \). We can think of implementation shortfall as representing the sum of two components: (1) the transaction costs incurred as a result of market impact and bid-ask spread and (2) the effect of other random price changes between the time the benchmark price is set and the time the trades are executed. Since implementation shortfall is an unbiased estimate of transaction costs, we can think of adding an error term to \( C(X) \) and modeling the random price changes as an error in a regression.

The bid-ask spread costs can be estimated in two ways: explicitly from the quoted spread and implicitly using implementation shortfall. In the explicit approach, we use the quoted spread to quantify the bid-ask component of implementation shortfall and estimate \( \bar{\lambda} \) and \( \alpha_1 \) using a non-linear regression. In the implicit approach, we estimate all four parameters \( \bar{\lambda} \), \( \alpha_1 \), \( \bar{\kappa} \), and \( \alpha_2 \) from the non-linear implementation shortfall regression. Both specifications are non-linear because the exponent parameters \( \alpha_1 \) and \( \alpha_2 \) appear in \( C(X) \) in a non-linear manner.

**Additional Adjustments.** To implement this strategy, two adjustments are made, one based on statistics and one based on economics.
First, since the errors in the regression are likely to be proportional in size to the return volatility of the stock, both the right-hand-side and left-hand-side variables are divided by return volatility \( \sigma_i \). This has the effect of making a crude correction for a heteroscedasticity problem which would otherwise occur. Furthermore, the imperfectly observed volatility \( \sigma_i \) is replaced by its estimate \( \sigma_{i,t-1}^e \). To the extent that \( \sigma_{i,t-1}^e \) is an imperfect estimate of \( \sigma_i \), the problem of bias associated with errors in variables is reduced by placing this variable on the right-hand-side.

Second, we adjust our estimation procedure for the fact that transition managers have access to different pools of liquidity. Transition orders can be executed through internal crossing networks, through external crossing networks, or in open market transactions. Market impact and bid-ask spread may be different across trading venues. Some of the portfolio transitions are the result of internal crosses. In an internal cross, one of the transition manager’s customers buys from the other at some price. In fact, it is possible that both the buyer and the seller represent different portfolio transitions being implemented simultaneously. Internal crosses with other types of customers also occur, for example, crosses against flows from a passive investment management unit affiliated with the same firm as the transition management unit. Since the buyer and the seller pay the same price, it seems reasonable to assume that there is no effective spread incurred for internal crosses but there is spread for external crosses and open market transactions.

Concerning market impact for crosses and open market transactions, it is assumed that the transition manager optimally chooses the percentages of the orders to execute via these trading venues. To the extent that crosses are cheaper than open market transactions, this is expected to show up as a larger percentage of the orders being crossed than executed in open markets, not as lower market impact and spread costs on crosses. The fact that both crosses and open market transactions are used in a significant proportion of orders suggests that there are significant pools of liquidity in both crossing networks and open markets, i.e., neither dominates the other. We therefore assume that there is market impact associated with internal crosses that is equal in magnitude to the impact of external crosses and open market trades. To check this assumption, we will later relax the assumptions of our benchmark specification and allow transaction costs vary across trading platforms.

**OLS Estimates for Quoted Spread.** We first start with testing the invariance theories using the data on quoted spreads, supplied in the dataset of portfolio transitions. Let \( k_i \) denote the pre-trade dollar quoted spread for the \( i \)th order. All three models have different predictions concerning how spread varies with trading activity. Their predictions about the fixed-cost component in (38) can be nested in a simple log-linear regression of the form

\[
\ln \left( \frac{k_i}{P_i \cdot \sigma_i} \right) = \ln \left( \frac{\bar{k}}{40 \cdot 0.02} \right) + \alpha_2 \cdot \ln \left( \frac{W_i}{W^*} \right) + \tilde{\epsilon}.
\]  

(41)

This equation relates the percentage quoted spread \( k_i/P_i \) as a fraction of volatility \( \sigma_i \) to the level of trading activity \( W_i \). The scaling constant \( W^* = (40)(10^6)(0.02) \) corresponds to \( W_i \) for the hypothetical benchmark stock with price $40 per share, trading volume of one million shares per day, and volatility of 2% per day. In this regression, the model of market microstructure invariance predicts \( \alpha_2 = -1/3 \), the model of invariant bet frequency predicts \( \alpha_2 = 0 \), and the model of invariant bet size predicts \( \alpha_2 = -1/2 \).
Table 6 presents the estimates. For the entire sample, the estimate for $\alpha_2$ is $\hat{\alpha}_2 = -0.31$ with standard error of 0.025. For samples of NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, the estimate ranges from $-0.32$ to $-0.22$. All estimates are economically close to the level of $-1/3$ predicted by the model of market microstructure invariance. For all samples, except for NYSE Buys, statistical tests do not reject market microstructure invariance ($\alpha_0 = -1/3$), but strongly reject both alternatives. Also, the estimated intercept implies that the percentage spread for a benchmark stock is equal to $\exp(-3.07) \cdot 0.02$, which equals about 9 basis points or 3.6 cents per share for the benchmark stock with a share price of $40$. This estimate is consistent with the declining pattern of spreads across volume groups in table 1.

**Impact Estimates in Nested Non-Linear Regression with Quoted Spread.** Let $X_i$ denote the number of shares in the $i$th order. Let $X_{omt,i}$ and $X_{ec,i}$ denote the number of these shares executed in open market transactions and external crosses, respectively. We use the percentage quoted spread $k_i$ as a bid-ask component of implementation shortfall and estimate the three parameters $\lambda, \alpha_1$ and $s$ in the following non-linear regression:

$$
\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot (0.02) \cdot \frac{X_i}{\sigma_i} = \frac{1}{2} \cdot \frac{X_i}{(0.01)V_i} + \frac{1}{2} \cdot \frac{X_{omt,i} + X_{ec,i}}{X_i} \cdot (0.02) \cdot 10^4 + \epsilon_i,
$$

where the coefficient $s$ is introduced to reflect the possibility that portfolio transition managers achieve price improvement by paying only a fraction of the quoted spread. In this non-linear regression, the observed data items have subscript $i$: $P_{ex,i}$, $P_{0,i}$, $I_{BS,i}$, $W_t$, $X_i$, $V_i$, $k_i$, and $\sigma_i$. The indicator variable $I_{BS,i}$ is +1 for buy orders and -1 for sell orders. Since $P_{0,i}$ denotes the benchmark price established the night before the transition begins and $P_{ex,i}$ denotes the average execution price, the expression $I_{BS,i}(P_{ex,i} - P_{0,i})/P_{0,i} \cdot 10^4$ is the implementation shortfall measured in basis points. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity variable $W_t$ is defined as the product of benchmark price $P_{0,i}$, last month’s volume $V_i$, and estimated volatility $\sigma_i$. The scaling constant $W^* = (40)(10^6)(0.02)$ corresponds to the trading activity for the hypothetical benchmark stock with price $40$ per share, trading volume of one million shares per day, and volatility of 2%. The term $X_i/(0.01)V_i$ is the size of the trade relative to average volume, scaled so that the size has a value of one for a trade of one percent of average daily volume. The variables are scaled so that $\lambda/2$ estimates in basis points the market impact costs of a trade of one percent of average daily volume.

To adjust standard errors for positive contemporaneous correlation in returns, the 441,865 observations are pooled by week over the 2001-2005 period into 4,389 clusters across 17 industry categories using the pooling option on Stata.

Recall that the three models make very different predictions concerning $\alpha_1$. The model of market microstructure invariance predicts $\alpha_1 = 1/3$. The model of invariant bet frequency predicts $\alpha_1 = 0$. The model of invariant bet size predicts $\alpha_1 = 1/2$.

The results of the non-linear regression are reported in table 7. The estimates for the parameters $\alpha_1$ are strongly supportive of the model of market microstructure invariance over the alternatives. The estimate for $\alpha_1$ is $\hat{\alpha}_1 = 0.31$ with standard error of 0.025. This point estimate is almost equal to the value of $1/3$ predicted by the model of market microstructure.
invariance. Furthermore, the standard error is sufficiently small that predictions of the other two models, \( \alpha_1 = 0.50 \) and \( \alpha_1 = 0 \), are soundly rejected. A Stata F-test for the hypothesis \( \alpha_1 = 1/3 \) can not be rejected, while similar F-tests soundly reject both alternatives.

The estimated coefficient for \( \alpha_1 \) remains close to \( 1/3 \) for samples of NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells. F-tests fail to reject the model of market microstructure invariance for three of the four samples and only reject the model with a borderline p-value of 0.0077 for NYSE Buys. Alternative models are strongly rejected.

The estimate for \( s \) is \( \hat{s} = 0.62 \) with standard error of 0.054, implying that transition managers incur as a realized cost only 62% of the quoted half-spread. As expected, the parameter \( s \) is positive and smaller than one. This is consistent with the interpretation that traders often obtain price improvement by submitting limit orders. Calibrated values are consistent with intuition in Goettler, Parlour, and Rajan (2005), who examine implications of the endogeneity of trading strategies in a dynamic limit order market.

The disaggregated results for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells also suggest that buying is more expensive than selling. For NYSE and NASDAQ, both estimated market impact costs are larger for buy orders than for sell orders by margins that are economically meaningful if not statistically significant. Portfolio transition managers have to pay a larger fraction of the quoted spread. This is consistent with the idea that the market believes that buy orders contain more information than sell orders. See Obizhaeva (2009a) for further discussion of this idea. It is also consistent with the possibility that closing benchmark prices are biased towards the bid side of the market.

**Impact and Spread Estimates in Nested Non-Linear Regression.** Instead of using the quoted spread as a proxy for the effective bid-ask spread, we now estimate both market impact and bid-ask spread from the implementation shortfall data. The four parameters \( \bar{\lambda}, \bar{\kappa}, \alpha_1, \alpha_2 \) are obtained from the following non-linear regression:

\[
\frac{\mathbb{I}_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot \frac{(0.02)}{\sigma_i} = \frac{1}{2} \cdot \bar{\lambda} \cdot \left[ \frac{W_i}{W^*} \right]^{\alpha_1} \cdot \frac{X_i}{(0.01)V_i} + \frac{1}{2} \cdot \bar{\kappa} \cdot \left[ \frac{X_{omt,i} + X_{exc,i}}{X_i} \right] \cdot \left[ \frac{W_i}{W^*} \right]^{\alpha_2} + \tilde{\epsilon}.
\]

This regression equation is similar to equation (42), with the only difference that the quoted bid-ask spread is replaced by the nested predictions of the invariance models in equation (38). The variables are scaled so that \( \bar{\lambda}/2 \) estimates in basis points the market impact costs of a trade of one percent of average daily volume in the benchmark stock, and \( \bar{\kappa}/2 \) estimates in basis points the effective spread cost for the benchmark stock.

The three models make very different predictions concerning \( \alpha_1 \) and \( \alpha_2 \). The model of market microstructure invariance predicts \( \alpha_1 = 1/3 \) and \( \alpha_2 = -1/3 \). The model of invariant bet frequency predicts \( \alpha_1 = 0 \) and \( \alpha_2 = 0 \). The model of invariant bet size predicts \( \alpha_1 = 1/2 \) and \( \alpha_2 = -1/2 \).

The results of the non-linear regression are reported in table 8. The estimates for the parameters \( \alpha_1 \) and \( \alpha_2 \) are strongly supportive of the model of market microstructure invariance over the alternatives. The estimate for \( \alpha_1 \) is \( \hat{\alpha}_1 = 0.33 \) with standard error of 0.024. This point estimate is almost exactly equal to the value of \( 1/3 \) predicted by the model of market microstructure invariance. Furthermore, the standard error is sufficiently small that predictions of the other two models, \( \alpha_1 = 0.50 \) and \( \alpha_1 = 0 \), are soundly rejected.
The estimate for $\alpha_2$ is $\hat{\alpha}_2 = -0.39$ with standard error of 0.025. This estimate is somewhat more negative than the value $\alpha_2 = -1/3$ predicted by the model of market microstructure invariance, by a margin of slightly more than two standard errors. The result suggests that effective bid-ask spreads decrease faster than the model predicts as trading activity increases. This is consistent with the back-of-the-envelope calculation from table 1 suggesting that as trading activity increases, quoted bid-ask spreads decline faster than the microstructure invariance predicts.

A Stata F-test for the joint hypothesis $\alpha_1 = 1/3$, $\alpha_2 = -1/3$ is rejected with a borderline p-value of 0.0742. Similar F-tests soundly reject the other two models with p-values less than 0.0001.

The estimate for half-price-impact is $\hat{\lambda}/2 = 2.85$ basis points with standard error of 0.245 ($t = 11.60$), and the estimate for half-spread is $\hat{\kappa}/2 = 6.30$ basis points with standard error of 1.131 ($t = 6.31$). These estimates imply that a hypothetical trade in the benchmark stock equal to one percent of daily volume incurs a market impact cost of 2.85 basis points and a spread cost of 6.30 basis points. The total cost of 9.15 basis points represents 3.66 cents per share for a $40$ stock, or $366$ for the hypothetical 10,000 share benchmark block.

The estimate for the bid-ask spread $\kappa$ is double the point estimate for the half-spread $\hat{\kappa}/2$, i.e. 12.60 basis points. This estimate is somewhat higher than the median spread of 8.09 basis points reported in table 1 for volume group 7, to which the hypothetical benchmark stock would belong. It is, however, similar to its mean value of 12.14 basis points. Similarly, the estimate for $\lambda$ is double the estimate of 2.85 basis points for $\lambda/2$, i.e., it is 5.70 basis points. This implies that a trade of 10,000 shares, one percent of average daily volume in the benchmark stock, incurs an impact cost of 2.28 cents per share.

When the four parameters are estimated separately for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, the results are also supportive of the model of market microstructure invariance. In three of the four regressions (with the exception of NYSE Buys), the estimated coefficient for $\alpha_1$ is close to the predicted value of $1/3$, but $\alpha_2$ is more negative than predicted. In these three cases, F-tests either fail to reject or narrowly reject the predictions of market microstructure invariance that $\alpha_1 = 1/3$, $\alpha_2 = -1/3$, with p-values of 0.1057, 0.9114, and 0.0443.

**A More General Specification.** Table 9 reports the results of a non-linear regression with a more general specification than table 8. Three separate market impact parameters and three separate spread parameters are estimated for open market trades, external crosses, and internal crosses. In addition, the exponents on the three components of market activity (volume, price, volatility) are allowed to differ. The regression estimated is

$$
\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot (0.02) = \frac{1}{2} \cdot \frac{\bar{\lambda}_{omt,i}X_{omt,i} + \bar{\lambda}_{ec,i}X_{ec,i} + \bar{\lambda}_{ic,i}X_{ic,i}}{(0.01)V_i} \cdot \left[ \frac{W_i}{W^*} \right]^{1/3} \cdot \frac{\sigma_i \cdot P_{0,i}^{\beta_1} \cdot V_i^{\beta_3}}{(0.02)(40)(10^6)}
$$

$$
+ \frac{1}{2} \cdot \frac{\bar{\kappa}_{omt,i}X_{omt,i} + \bar{\kappa}_{ec,i}X_{ec,i} + \bar{\kappa}_{ic,i}X_{ic,i}}{X_i} \cdot \left[ \frac{W_i}{W^*} \right]^{1/3} \cdot \frac{\sigma_i \cdot P_{0,i}^{\beta_1} \cdot V_i^{\beta_3}}{(0.02)(40)(10^6)} + \tilde{e}. \tag{44}
$$

35
Because the exponents on the $W$-terms are set to be $1/3$ and $-1/3$, the model of market microstructure invariance predicts

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0.$$ \hspace{1cm} (45)

The model of invariant bet frequency predicts

$$\beta_1 = \beta_2 = \beta_3 = -1/3, \quad \beta_4 = \beta_5 = \beta_6 = 1/3.$$ \hspace{1cm} (46)

The model of invariant bet size predicts

$$\beta_1 = \beta_2 = \beta_3 = 1/6, \quad \beta_4 = \beta_5 = \beta_6 = -1/6.$$ \hspace{1cm} (47)

The first column of the table presents the results for all buys and sells. The remaining four columns present results for separate regressions for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells.

F-tests of the above restrictions for the model of invariant bet frequency ($F = 78.25$) and the model of invariant bet size ($F = 13.71$) are rejected very strongly ($p << 0.0001$). An F-test of the restrictions of our model in equation (45) is rejected less strongly, with $F = 4.55$, $p = 0.0001$. From the table, it appears that one reason for this rejection is that bid-ask spreads decrease faster than predicted as trading volume increases. The estimate of $\beta_6$ is $-0.09$ with standard error of $0.025$ ($t = 3.46$). The rapid decrease in spreads as trading volume increases is consistent with the results from table 1. The bid-ask spread, however, does not decrease as fast as predicted when stock price increases since $\beta_5$ is estimated as $0.18$ with standard error of $0.062$ ($t = 1.83$). Our tests do not show any significant deviation of bid-ask spread from predicted values with respect to volatility. Another reason for the rejection is that the estimates of $\beta_1$ and $\beta_2$ are quite negative. The estimate of $\beta_1$ is $-0.31$ with standard error of $0.191$ ($t = -1.61$), and the estimate of $\beta_2$ is $-0.22$ with standard error of $0.191$ ($t = -2.26$). The estimate of $\beta_3$ is close to zero. These results say that market impact behaves as predicted by the model of market microstructure invariance when the number of shares traded increases, but market impact decreases relative to what is predicted when volatility and stock price increase.

The rejection of the model of market microstructure invariance seems to be related to the fact that the exponents for volatility and price behave differently from the coefficients for share volume; the coefficients for volatility behave similarly to the coefficients for price. This suggests that the rejection might depend in a subtle manner on tick effects. When volatility is high and stock price is high, the tick size is small relative to a typical day’s trading range.

Despite increasing the number of estimated parameters from four to twelve, the adjusted $R^2$ in the aggregate regression increases only from 0.0123 to 0.0129.

The estimates for the three half spread parameters are $\hat{k}_{omt}/2 = 6.56$, $\hat{k}_{ec}/2 = 6.26$, and $\hat{k}_{ic}/2 = 0.25$. These results support the assumption that there is no spread associated with internal crosses, and the spread associated with external crosses is the same as the spread associated with open market trades.

The point estimates for market impact parameters are $\hat{\lambda}_{omt}/2 = 4.49$, $\hat{\lambda}_{ec}/2 = 2.17$, and $\hat{\lambda}_{ic}/2 = 2.41$. These results support the assumption that internal crosses do have market impact. The results, however, suggest that the impact for open market trades may be greater
than the impact for internal and external crosses. While interpreting these results, we have to keep in mind, however, that the separate estimates of transaction costs parameters for different trading venues may suffer from selection bias since transition managers optimally chooses trading venues to minimize the total costs.

**Comparison of Three Models.** Table 10 presents estimates for equation (43) with the parameters $\alpha_1$ and $\alpha_2$ restricted to be as predicted in the model of market microstructure invariance, the model of invariant bet frequency, and the model of invariant bet size, respectively. For each of the three models, only two parameters are estimated: half-price-impact $\bar{\lambda}/2$ and half spread $\bar{\kappa}/2$.

For the model of market microstructure invariance, the reduction from four parameters to two parameters reduces the adjusted $R^2$ from 0.0123 to 0.0122, consistent with very mild rejection of the model reported in table 8. Comparing table 8 with table 10, the parameter estimates for half-market impact $\bar{\lambda}/2$ increases from 2.85 basis points to 2.89 basis points, and the parameter estimate for half bid-ask spread $\bar{\kappa}/2$ increases from 6.30 basis points to 7.90 basis points.

For the model of invariant bet frequency, the reduction from four parameters to two parameters reduces the $R^2$ greatly, from 0.0123 to 0.0075, consistent with very strong rejection of the model. Furthermore, the point estimate for half market impact drops enormously, from $\bar{\lambda}/2 = 2.85$ to $\bar{\lambda}/2 = 0.3788$. This is offset by a large increase in the estimated half spread, from $\bar{\kappa}/2 = 6.31$ to $\bar{\kappa}/2 = 15.29$. The model of invariant bet frequency is intuitively appealing since it suggests that the market impact of trading a given percentage of average daily volume is constant as a fraction of daily returns standard deviation, regardless of the level of trading activity in the stock. The strong rejection of this model, combined with the large changes in estimated coefficients, suggests that this model leads to the misleading empirical result that market impact is less important than it really is, and bid-ask spread is more important than it really is. Therefore, one of the justifications for the model of market microstructure invariance is that it allows for the importance of market impact to be estimated more accurately from a better specified model.

For the model of invariant bet size, the reduction from four parameters to two parameters reduces the adjusted $R^2$ from 0.0123 to 0.0110, consistent with a strong rejection of this model. The point estimates for half market impact and half bid-ask spread change in the opposite direction, with the point estimate for market impact increasing from 2.85 to 3.92 and the point estimate for half spread decreasing from 6.31 to 3.46.

In all three specifications, separate regressions for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells continue to suggest that buying is more expensive than selling or that benchmark prices are biased towards the bid side of the market.

For each subsample, the model of market microstructure invariance has the largest maximum log-likelihood function among three models under consideration. This evidence implies that, according to various likelihood-based comparison methods, our model delivers predictions that are more consistent with portfolio transition data than predictions of alternative models. Almost regardless of priors, a Bayesian statistician will calculate the posterior probability of our proposed model as being close to one and the posterior probabilities of alternative models as being close to zero.
Economic Interpretation. The model of market microstructure invariance provides a much better proxy for the spread costs than the quoted spread. Indeed, when the bid-ask costs are fixed to be a constant fraction of the quoted spread, we obtain $R^2 = 0.0112$, as reported in table 7. At the same time, when we model bid-ask costs as predicted by the invariance model, we obtain a much higher $R^2 = 0.0125$, as shown in table 8.

Dummy Variables as Robustness Check. Figure 5 presents the results of three linear regressions, one for each of the three proposed models. The regression represents a modification of equation (43) in two ways. First, similarly to table 10, for each of the three models, the values of $\alpha_1$ and $\alpha_2$ are fixed at the levels predicted by the models. Second, a dummy variable for each of the ten volume groups is associated with a half-market impact parameter and a half spread parameter for each group. The result is a regression with twenty coefficients, two coefficients for each volume bin, with one coefficient $\bar{\lambda}_j$ for half market impact and one coefficient $\bar{\kappa}_j$ for half spread. The regression equation can be written

$$\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot \frac{(0.02)}{\sigma_i} = \left( \sum_{j=1}^{10} \frac{W_i}{W^*} \cdot \frac{1}{2} \cdot \bar{\lambda}_j \right) \cdot \frac{X_i}{(0.01)V_i} \cdot \left( \sum_{j=1}^{10} \frac{W_i}{W^*} \cdot \frac{1}{2} \cdot \bar{\kappa}_j \right) \cdot \left( X_{omt,i} + X_{ec,i} \right) \cdot \frac{X_i}{W^*} \cdot \left[ \frac{W_i}{W^*} \right]^{\alpha_2} + \tilde{\epsilon}. \tag{48}$$

In the figure, for each of the three models, there is a graph of the estimates of the ten dummy variables for half market impact $\bar{\lambda}/2$ and a graph of the estimates of the ten dummy variables for half-spread $\bar{\kappa}/2$. Each graph also shows the 95% confidence intervals around the point estimates, as well as a horizontal line showing the point estimate from table 10. If the model is well specified, then the ten dummy variables should be the same and should equal the point estimates from table 10.

For the model of market microstructure invariance, all of the point estimates lie either within the 95% confidence bands or slightly outside the 95% confidence bands, consistent with the mild rejection of the model discussed above. For the smallest volume group, the estimate for the half-spread has a very small confidence band which anchors the point estimate close to the two-parameter model. For the smallest volume group, the estimate for half-market impact also has a relatively small confidence band which anchors it close to the two-parameter model as well. For the two largest groups, the half-spread estimates are somewhat larger than the unconstrained estimate and the half-market impact estimates are somewhat smaller. The data seem to be saying that for the very largest stocks, there is a somewhat bigger spread and somewhat less market impact than implied by the model of market microstructure invariance. For trade groups from 2 to 6, the data are saying the opposite, i.e., that the half spread should be smaller and the half market impact larger than in the two parameter model. These patterns suggest that the transition manager might be using basket trades and then splitting the total costs among individual stocks in a manner consistent with the model of invariant bet frequency, i.e., assigning smaller than needed costs to large stocks and larger than needed costs to small stocks. Alternatively, given a short horizon of portfolio transitions, execution of orders in less active stocks can be “too quick” relative to their natural trading game time, thus resulting in more substantial trading costs.
Similar graphs for the dummy variables in the model of invariant bet frequency are presented in the middle of the figure. This model predicts that effective bid-ask spreads do not decline as trading activity increases. It is clear from the figure, however, that the estimated effective bid-ask spreads for the smallest volume group are far greater than the estimated bid-ask spreads for the other nine groups. This places the effective spread for the smallest group far above the point estimate from the two parameter model and very far outside the 95% confidence bands. For the market impact parameters, the model generates a great deal of power from the smallest volume group because the trade sizes are large relative to volume for this group. The point estimate of half-market impact for the smallest volume group is therefore very close to the point estimate from the two parameter model. But this forces the dummy variables for half-market impact for the nine large volume groups to lie far above the point estimate from the two parameter model. If the smallest volume group were eliminated from consideration, it appears from the figure that the model of invariant bet frequency would perform almost as well as the model of market microstructure invariance. It is possible that the transition manager trades baskets of stocks and then marks the prices according to the model of invariant bet frequency. Perhaps basket trades tend to occur in the largest stocks; the model might look good for spurious reasons especially among the largest stocks.

Graphs of the dummy variables for the model of invariant bet size are presented on the right-hand side of the figure. The model generates very precise estimates of spreads for the smallest size group. For the larger size groups, the predicted spreads are much larger than the point estimates from the two parameter model. For half market impact, the dummy variables decrease almost monotonically, indicating that the rapid increase in market impact implied by the model value of $\alpha_1 = 1/2$ is greater than what is consistent with the data.

7 Calibration and Its Implications

Our estimates allow us to calibrate the probability distribution of the trading game invariant $\tilde{I}$ and the values of the parameters $\zeta$, $C_L$ and $C_K$. Based on this calibration, our microstructure invariance implies simple formulas for bet arrival rate, distribution of bet sizes, and expected trading costs. The result is an intuitive story of how stock prices result from the interaction of bets placed by institutional investors.

Note that many formulas in this section depend on our identifying assumptions about the values of volume deflator $\zeta$ and portfolio transition order size multiplier $\theta$. These parameters are not estimated in the paper; their calibration is an important question for future research. As a baseline case, we assume that a specialist intermediates all bets without involvement of other intermediaries ($\zeta = 2$) and portfolio transition orders are similar in size to other bets ($\theta = 1$).

7.1 Calibration of Parameter Values

Calibration of $\tilde{I}$. After plugging the assumption $\alpha = 1/3$ of market microstructure invariance into equation (33), then substituting the expression for $|Q^*|/V^*$ from equation (8)
into equation (33), we can obtain the following:
\[
\ln \left[ \frac{X_i}{V_i} \right] + \frac{2}{3} \ln \left[ \frac{W_i}{W^*} \right] \approx \ln (|\tilde{I}|) - \frac{1}{3} \ln E\{|\tilde{I}|\} - \frac{1}{3} \ln \frac{\zeta}{2} - \ln \theta - \frac{2}{3} \ln W^*. \tag{49}
\]

In equation (49), the left-hand side adjusts the mean of portfolio transition order size as a fraction of average daily volume \(\ln (X_i/V_i)\) for the stock's level of trading activity to obtain an invariant distribution. The right-hand side is the log of the invariant distribution, adjusted by the identifying parameters \(\zeta\) and \(\theta\) and trading activity in the benchmark stock.

Our tests in Section 5.1 indicate that the distribution of the trading game invariant \(\tilde{I}\) can be approximated as the product of two random variables: (1) a buy-sell indicator variable \(\tilde{\delta}\) assuming values of +1 or −1 with equal probability, (2) and a standardized normal random variable converted to a log-normal with mean \(\mu_I\) and variance \(\sigma_I^2\) via exponentiation:
\[
\tilde{I} \approx \tilde{\delta} \cdot e^{\mu_I + \sigma_I \tilde{Z}}. \tag{50}
\]

The empirical results in section 5.1 show that the left-hand side of (49) can be approximated as a normal distribution with mean −5.69 and variance 2.50. Using formulas for the moments of a log-normal, equating means and variances of the right- and left-hand sides of (49) yields

\[
E\{\ln \left[ \frac{X_i}{V_i} \right] + \frac{2}{3} \ln \left[ \frac{W_i}{W^*} \right] \} = -\frac{1}{3} \ln \frac{\zeta}{2} - \ln \theta - \frac{2}{3} \ln W^* - \frac{1}{6} \sigma_I^2 + \frac{2}{3} \mu_I = -5.69, \tag{51}
\]

\[
\text{Var}\{\ln \left[ \frac{X_i}{V_i} \right] + \frac{2}{3} \ln \left[ \frac{W_i}{W^*} \right] \} = \sigma_I^2 = 2.50. \tag{52}
\]

Equations (51) and (52) imply the following values for \(\mu_I\) and \(\sigma_I^2\):

\[
\mu_I = 5.6824 + \frac{1}{2} \ln (\zeta/2) + \frac{3}{2} \ln \theta, \tag{53}
\]

\[
\sigma_I^2 = 2.50. \tag{54}
\]

Note that from equation (53), the value of \(\mu_I\) must be adjusted for the non-bet volume multiplier \(\zeta\) and the portfolio transition order multiplier \(\theta\). In a baseline case (\(\zeta = 2, \theta = 1\)), we have
\[
\tilde{I} \approx \tilde{\delta} \cdot e^{5.6824 + \sqrt{2.50} \tilde{Z}}. \tag{55}
\]

**Calibration of the Expected Market Impact Cost Parameter \(C_L\).** Applying the definition of \(C_L\) in equation (9) to the benchmark stock, then substituting into equation (39) yields
\[
C_L = \frac{1}{2} \bar{\lambda} \cdot 10^{-4} \cdot \frac{P^*}{0.01 \cdot V^*} \cdot E\{(\tilde{Q}^*)^2\}. \tag{56}
\]

Combining our estimate \(\frac{1}{2} \bar{\lambda} = 2.89\) with equations (1), (7), (50), (53), and (54), we obtain the following estimate of the expected market impact cost of a bet:
\[
C_L = $1,960 \cdot \theta^2. \tag{57}
\]

\(^1\)The number -5.69 in equation (51) and the number 5.6824 in equation (53) are related by the following arithmetics: 5.6824 = −5.69 · 3/2 + ln(0.02 · 40 · 10^6) + 1/4 · 2.50.
The assumption of market impact invariance implies that this cost does not depend on the level of trading activity $W$. Furthermore, it does not depend on the parameter $\zeta$. It does depend, however, on our identifying assumption about the factor $\theta$ used to deflate the distribution of portfolio transition orders to match bets. If transition orders are similar to typical bets, implying $\theta = 1$, then the cost of a bet is equal to $\$1,960$. The larger are portfolio transition orders relative to other bets, the smaller is the expected impact cost of a bet.

**Calibration of the Expected Bid-Ask Spread Cost of a Bet $C_K$.** Applying the definition of $C_K$ in equation (11) for the benchmark stock to equation (40) yields the following formula for $C_K$:

$$C_K = \frac{1}{2} \tilde{\kappa} \cdot 10^{-4} \cdot P^* \cdot E\{|\tilde{Q}^*|\}. \quad (58)$$

Given our estimate $\frac{1}{2} \tilde{\kappa} = 7.90$, from equations (1), (7), (50), (53), and (54), we obtain the following estimate of the expected bid-ask spread cost of a bet:

$$C_K = \$373. \quad (59)$$

Like the estimate for the expected market impact cost of the bet, this cost depends neither on trading activity $W$ nor the volume deflator $\zeta$. Unlike the estimate of the expected market impact cost of a bet, it also does not depend on $\theta$.

**Calibration of $\psi$ and $\phi$.** The parameters $C_L$ and $C_K$ can be also expressed in terms of parameters $\psi$ and $\phi$.

The parameter $\psi$ relates the aggregate price impact of bets to the total variance of returns. Equations (21) and (57) imply that

$$\psi = 1.09 \cdot \left[\frac{\zeta}{2}\right]^{-1/2} \cdot \theta^{1/2}. \quad (60)$$

The calibrated values for $\psi$ depend on the identification assumptions about $\zeta$ and $\theta$. If we assume that $\zeta = 2$ and $\theta = 1$, then $\psi = 1.09$ implies that execution of bets is associated with more than 100 percent of volatility. This implies some transitory component of price impact equal to at least 9% of daily volatility, since some fraction of volatility may also result from public announcements. For other identifying assumptions about the values of the parameters of $\zeta$ and $\theta$, the decomposition of price impact into transitory and permanent components may be different. Studying this decomposition is an interesting issue for future research. Note, however, that a transitory component in market impact costs does not necessarily imply excess volatility and mean reversion in actual prices. A transitory component could arise, for example, from high frequency traders and other arbitragers using computer algorithms to “front-run” bets in an equilibrium where prices follow a martingale.

The parameter $\phi$ relates the expected market impact cost of a bet to the spread cost. Equations (25), (57), and (59) imply that

$$\phi = 0.19 \cdot \theta^{-2}. \quad (61)$$

This value implies that the expected market impact cost of a bet is about five items larger than the expected spread cost of a bet, under the assumption of $\theta = 1$. 

41
The implied values of $\psi$ and $\phi$ seem reasonable and consistent with typical views of market participants. For large institutional bets, market impact costs are much more important than bid-ask spread costs; for small retail bets, by contrast, the bid-ask spread cost is more important than the market impact cost. This point can be easily illustrated by comparing the expected execution costs of the average bet in (57) and (59) with execution costs for the median bet. The median bet is smaller than the average bet by a factor of $\exp(0.5 \cdot \sigma_I^2)$ or 0.28. This implies that the market impact cost of the median bet is only $13.20 \cdot \theta^2$, while its bid-ask spread cost is $107$. These intuitively reasonable approximations provide an additional consistency check for our invariance theory.

7.2 Implications For the Trading Process

Our empirical results suggest simple formulas for the expected number of bets per day and the distribution of bet size as a function of expected daily dollar trading volume $P \cdot V$ and expected daily return volatility $\sigma$. From equations (6), (50), (53), and (54), an estimate for the expected number of bets placed per day $\gamma$ is given by

$$\gamma = 85 \cdot \left[ \frac{\zeta}{2} \right]^{-1} \cdot \theta^{-1} \cdot \left[ \frac{V \cdot P \cdot \sigma}{(0.02)(40)(10^6)} \right]^{2/3}.$$  

From equations (8), (50), (53), and (54), an estimate of the distribution of bet size $\tilde{Q}$ as a fraction of average daily volume $V$, scaled in basis points by multiplying by $10^4$, is given by

$$\frac{\tilde{Q}}{V} \cdot 10^4 \approx 34 \cdot \theta \cdot \left[ \frac{V \cdot P \cdot \sigma}{(0.02)(40)(10^6)} \right]^{-2/3} \cdot e^{\sqrt{2.50} \cdot \tilde{Z}},$$

where $\tilde{Z}$ is the standard normal random variable.

For the benchmark stock, with price of $40 per share, trading volume of one million shares per day, and volatility of 2% per day, about $85 \cdot \left[ \frac{\zeta}{2} \right]^{-1} \cdot \theta^{-1}$ independent bets are placed per day. The implied average size of a bet is $471,838 \cdot \theta$ (or 11,796 \cdot \theta shares). The implied expected market impact cost of a bet is $1960 \cdot \theta^2$, and the implied expected spread cost is $373 \cdot \theta$. The median size of a bet is $135,184 \cdot \theta$ ($3,380 \cdot \theta$ shares), with implied market impact cost of $13.20 \cdot \theta^2$ and implied spread cost of $107 \cdot \theta$. Our formulas (62) and (63) show how to extrapolate our estimates to any other security given its share price, share volume, and volatility.

7.3 The Contribution of Large Bets to Volume and Volatility

A large fraction of trading volume comes from large bets, and a surprisingly larger fraction of returns variance comes from unusually large bets. This can easily be quantified using the properties of a log-normal distribution.

Since the number of shares in a bet $\tilde{Q}$ is proportional to the invariant $\tilde{I}$ from equation (5), we estimate both $\tilde{Q}$ and $\tilde{I}$ to be log-normals with log-variance $\sigma_I^2 = 2.50$. Let $\eta(z)$ and $N(z)$ denote the PDF and CDF of a standardized normal distribution, respectively. Define $F(z,k)$ by $F(\bar{z},m) = \int_{-\infty}^{\infty} \exp(m \sigma_I z) \cdot \eta(z) \cdot dz$. It is easy to show that $F(\bar{z},m) = \exp(k^2 \sigma_I^2) \cdot N(\bar{z} - m \sigma_I)$. It is straightforward to show that the fraction of the $m$th moment
of order size arising from bets greater than \( \bar{z} \) standard deviations above the mean is given by

\[
\frac{F(\bar{z}, m)}{F(-\infty, m)} = 1 - N(\bar{z} - m\sigma_I),
\]

(64)

where we estimate \( \sigma_I = \sqrt{2.50} = 1.58 \).

Bets larger than \( \bar{z} \) standard deviations above the mean generate a fraction of total trading volume given by \( 1 - N(\bar{z} - \sigma_I) \), obtained by plugging \( m = 1 \) into equation (64). This implies that 50% of trading volume is generated by the largest 5.71% of bets and the 94.29% of trading volume is generated by bets larger than the 50th percentile.

Assuming bets generate permanent price impact proportional to their sizes, the contribution of a bet to price volatility is proportional to the squared size of the bet. Thus, bets larger than \( \bar{z} \) standard deviations above the log-mean (median) bet size contribute a fraction of total volatility given by \( 1 - N(\bar{z} - 2\sigma_I) \), obtained by plugging \( m = 2 \) into equation (64). Since we estimate \( 2\sigma_I = 3.16 \), we infer that 50% of returns variance is generated by the 0.08% of bets larger than 3.16 standard deviations above the log-mean bet size, and 99.92% of returns variance is generated by bets greater in size than the 50th percentile. Since the benchmark stock has about 85 bets per day, about 50% of the variance of returns in the benchmark stock is generated by large bets arriving on average once every 14 trading days. Orders 4 standard deviations larger than the mean—expected to arrive about once every 355 trading days—generate about 20% of returns variance.

The fact that returns volatility is dominated by large bets suggests investigating in future research the extent to which large traders smooth out trades over time to avoid front-running and the extent to which rapid execution of large bets generates fat tails in the distribution of trading volume and returns.

### 7.4 Implications For Transaction Costs

This paper also provides a simple, practical formula for calculation of expected transaction costs as a function of observable dollar trading volume and volatility. Using portfolio transition data, we estimate the level of transaction costs for a benchmark stock. In particular, table 10 shows that, if the exponent parameters are set to the values implied by the model of market microstructure invariance \( \alpha_1 = 1/3 \) and \( \alpha_2 = -1/3 \), then the estimated values of half market impact and half bid-ask spread for the benchmark stock are given by \( \bar{\lambda}^*/2 = 2.89 \) basis points and \( \bar{\kappa}^*/2 = 7.91 \) basis points, respectively.

The model of market microstructure invariance describes how transaction costs vary across stocks with different levels of trading activity. A simple formula for expected costs \( C(X) \) shows how to extrapolate the estimated transaction costs for the benchmark stock to any other security. The expected trading costs for an order of \( X \) shares can be written

\[
C(X) = 2.89 \cdot \left( \frac{W}{(0.02)(40)(10^6)} \right)^{1/3} \cdot \frac{\sigma}{0.02} \cdot \frac{X}{(0.01) \cdot V} + 7.91 \cdot \left( \frac{W}{(0.02)(40)(10^6)} \right)^{-1/3} \cdot \frac{\sigma}{0.02},
\]

(65)

where trading activity \( W \) is the product of average daily dollar volume \( P \cdot V \) and daily returns volatility \( \sigma \).
Instead of using our concept of trading activity \( W = P \cdot V \cdot \sigma \), equation (65) can be expressed equivalently in terms of dollar volume \( P \cdot V \) and volatility \( \sigma \):

\[
C(X) = 2.89 \cdot \left( \frac{P \cdot V}{(40)(10^6)} \right)^{1/3} \cdot \left( \frac{\sigma}{0.02} \right)^{4/3} \cdot \frac{X}{(0.01) \cdot V} + 7.91 \cdot \left( \frac{P \cdot V}{(40)(10^6)} \right)^{-1/3} \cdot \left( \frac{\sigma}{0.02} \right)^{2/3}.
\]  

(66)

Note that in equation (66), the exponent of dollar volume in the market impact term is 1/3 while the exponent of volatility is 4/3. The effects of given percentage variations in volatility on market impact costs are 4 times greater than the effect of the same given percentage variation in dollar volume. Note also that in equation (66), the exponent of volatility in the market impact term is 4/3, while the exponent of volatility in the bid-ask spread term is only 2/3. Suppose that in a time of market stress, volatility increases while dollar volume remains constant. Equation (66) therefore implies that if spread costs of a bet of given dollar size double as a result of increased volatility, then market impact costs increase by a factor of four. This is consistent with conventional trader wisdom that at times when increased market volatility is associated with increased bid-ask spreads, the costs of executing large trades increase much more than the costs of executing small trades.

We conjecture that equations (65) or (66) may estimate transaction costs not only for the U.S. stocks but also for other securities. The concept of microstructure invariance is consistent with the intuition that markets for different securities are likely to share the same underlying principles. Given the estimates \( \lambda^* / 2 \) and \( \kappa^* / 2 \) for the benchmark stock, extrapolating equations (65) and (66) may allow us to estimate expected transaction costs for futures contracts, fixed income securities, foreign exchange, and commodities. Equations (65) and (66) may apply internationally, after adjusting our extrapolation method for differences in exchange rates. Whether the market microstructure invariance hypothesis is indeed of such a general nature is an interesting issue for future research.

Our paper provides answers to some important questions concerning management of transaction costs. Using the formula for expected costs and their components, asset managers can better forecast what expenses they will incur during implementation of their investment strategies as well as what amount of funds can be allocated to a strategy before it becomes unprofitable. Understanding cross-sectional variation in transaction costs has other practical implications. For example, when comparing execution quality across brokers specializing in stocks with different trading activity, performance metrics should take account of non-linearities documented in our paper.

When executing basket trades, it may not be appropriate to assign the same number of basis points of transactions to each stock in a basket. Instead, transaction cost attribution to individual securities in a basket should take account of the dollar trade size, volatility, and average dollar volume of the stock, in the context of an appropriate model. Finally, our analysis has implications for trading strategies. If it is reasonable to restrict trading of the benchmark stock to say 1% of average daily volume, then a smaller percentage would be appropriate for more active stocks and a larger percentage would be appropriate for less active stocks.
7.5 Market Velocity and Two Ways to Measure Liquidity

Intuitively, the bet arrival rate $\gamma$, the calibration for which is provided in equation (62), measures the “velocity” of the market. It measures the speed with which bets are placed or money changes hands as asset prices change. It is reasonable to conjecture that market liquidity is an increasing function of market velocity.

In the context of measuring trading costs, the term “liquidity” can have different meanings, including the cost of converting an asset to cash and the cost of transferring a risk. We will show that both of these measures of liquidity are related to velocity $\gamma$. Let $1/L_\theta$ denote a liquidity index measuring the expected cost of converting an asset to cash as a fraction of the dollar value of the asset (e.g., in basis points). Let $1/L_\sigma$ denote a liquidity index measuring the expected cost of transferring a risk of the sizes that are exchanged in the market (e.g., similarly to a Sharpe ratio, in units of risk). We think of trading costs as the reciprocal of liquidity, so increasing liquidity $L_\theta$ and $L_\sigma$ implies lower trading costs.

We define $1/L_\theta$ as the dollar-volume-weighted average transactions cost expressed in basis points. For the benchmark stock, from expected impact cost $\$1,960 \cdot \theta^2$, expected spread cost $\$373 \cdot \theta$, and expected dollar bet size $\$471,838 \cdot \theta$, we estimate $1/L_\theta^*$:

$$1/L_\theta^* = \frac{E\{\tilde{Q}^* \cdot P^* \cdot C(\tilde{Q}^*)\}}{E\{\tilde{Q}^* \cdot P^*\}} = \frac{1960 \cdot \theta + 373}{471,838} \text{ bp.} \quad (67)$$

If portfolio transition orders are similar in size to typical other bets ($\theta = 1$), we have $1/L_\theta^* \approx 50$ bp. For an arbitrary stock, the calculation of $L_\theta$ for the benchmark stock extrapolates similarly to equations (65) and (66):

$$1/L_\theta = \left[\frac{1960 \cdot \theta + 373}{471,838}\right] \cdot \left(\frac{W}{W^*}\right)^{-1/3} \cdot \frac{\sigma}{\sigma^*} = \left[\frac{1960 \cdot \theta + 373}{471,838}\right] \cdot \left(\frac{P \cdot V}{40 \cdot 10^6}\right)^{-1/3} \cdot \left(\frac{\sigma}{0.02}\right)^{2/3}. \quad (68)$$

Note that $L_\theta$ is proportional to the cube root of dollar volume per unit of returns variance $[PV/\sigma^2]^{1/3}$. “Dollars per unit of variance” is a simple and intuitive—if ad hoc—way to proxy for liquidity. Market microstructure invariance implies that taking the cube root generates a logical index for measuring liquidity based on the cost of converting an asset to cash.

Our second liquidity measure $1/L_\sigma$ measures the cost of exchanging risks of the sizes that market participants exchange in the market. This can be calculated by expressing $1/L_\theta$ in standard deviation units, obtained by dividing $1/L_\theta$ in equation (68) by $\sigma$,

$$1/L_\sigma = \left[\frac{1960 \cdot \theta + 373}{471,838 \cdot 0.02}\right] \cdot \left(\frac{W}{W^*}\right)^{-1/3} = \left[\frac{1960 \cdot \theta + 373}{471,838 \cdot 0.02}\right] \cdot \left(\frac{P \cdot V \cdot \sigma}{40 \cdot 10^6 \cdot 0.02}\right)^{-1/3}. \quad (69)$$

Note that $L_\sigma$ is proportional to the cube root of trading activity $[P \cdot V \cdot \sigma]^{1/3}$. Market microstructure invariance implies that taking the cube root generates a logical index for measuring liquidity based on the cost of transferring risks.

Both measures $L_\theta$ and $L_\sigma$ can be expressed in terms of market velocity $\gamma$. Plugging equation (62) into equations (68) and (69), we find

$$1/L_\theta = 1/L_\sigma \cdot \sigma = \left[\frac{1960 \cdot \theta + 373}{471,838}\right] \cdot \left(\frac{\zeta}{2}\right)^{-1/2} \cdot \left(\frac{\gamma}{85}\right)^{-1/2} \cdot \frac{\sigma}{0.02}. \quad (70)$$
For a baseline case ($\zeta = 2$ and $\theta = 1$), we obtain

$$1/L_\psi = 1/L_\sigma \cdot \sigma = 0.25 \cdot \left( \frac{\gamma}{85} \right)^{-1/2} \cdot \sigma. \quad (71)$$

For the benchmark stock, $1/L_\psi^*$ is 50 basis points and $1/L_\sigma^*$ is 0.25. Note that $L_\psi$ is proportional to the square root of market velocity per unit of returns variance $[\gamma/\sigma^2]^{1/2}$. The value of $L_\psi$ is obtained by deflating the velocity or speed with which money changes ($\gamma$) hands by the speed with which returns unfold ($\sigma^2$), then taking a square root. $L_\sigma$ is proportional to the square root of market velocity.

We conjecture that market velocity is likely related to the speed with which traders incorporate information into prices. Consider a trader who generates a signal which predicts accurately a proportion $\pi$ of the standard deviation of returns unfolding over a calendar period of $T$ days. How large does the holding period $T$ have to be to cover the round-trip costs of buying and selling a typical bet? The answer is that the required $T$ must solve $\pi \cdot \sigma \cdot T^{1/2} = 2 \cdot 0.25 \cdot \sigma \cdot (\gamma/85)^{-1/2}$, where round-trip costs are estimated using equation (71). This implies that a holding period inversely proportional to market velocity $\gamma$ given by $T = (2 \cdot 0.25)^2 \cdot \pi^{-2} \cdot (\gamma/85)^{-1}$. Constructing a self-contained theoretical model which illustrates this point is an interesting subject for future research which takes us beyond the scope of this paper. Such a model should have the property that the time horizon over which traders hold speculative positions is inversely proportional to velocity $\gamma$.

The measures of liquidity, $L_\psi$ and $L_\sigma$, are useful for different purposes. For example, consider what happens if volatility increases, holding dollar trading volume constant. Does liquidity increase or decrease? The value of $L_\psi$ decreases, implying that increased volatility increases the costs of converting assets to cash. The value of $L_\sigma$ increases, implying that increased volatility encourages speculation by making it less costly to transfer risks of given dollar standard deviation. These different measures of liquidity explain why some traders might believe increased volatility is bad for markets, while others believe it is good.

8 Conclusion

This paper proposes three models that differ in their assumptions about what features of trading games remain invariant as games themselves vary across securities with different levels of trading activity. Our preferred model of market microstructure invariance is based on the intuition that deep parameters of the trading game itself are invariant, but the length of the trading games varies across stocks because the time clock in different markets ticks at different rates. All three models make dramatically different predictions concerning how market impact and bid-ask spreads vary cross-sectionally across stocks.

Data on portfolio transitions are used to test the models in two ways. First, under the identifying assumption that portfolio transitions are proportional in size to independent bets made by market participants, the size of portfolio transition orders is used to test implications of the models concerning how order sizes vary across stock with different levels of trading activity. Second, their implications for market impact and spreads are tested using estimates derived from implementation shortfall.
The empirical results are supportive of the model of market microstructure invariance, but with some caveats. The model assumes that if trading activity increases by one percent, trade size as a fraction of daily volume falls by 2/3 of one percent. The trade size regressions provide strong support for this assumption. The coefficient estimate of -0.63 is reasonably close to the predicted value of -2/3. The predictions of the model of market microstructure invariance for market impact and bid-ask spread are also supported by data. The empirical prediction that a one percent increase in trading activity increases the market impact (in units of daily standard deviation) by 1/3 of one percent is almost exactly the point estimate of 0.33 from non-linear regressions based on implementation shortfall. The empirical predictions that a one percent increase in trading activity decreases the bid-ask spread (in units of standard deviation) by 1/3 of one percent is matched by data reasonably closely. The point estimate from using trading activity to estimate quoted spreads -0.36, while the point estimate for implicit spread cost in the non-linear implementation shortfall regression is is -0.39.

There are a number of issues which need further investigation. First, the statistical power behind implementation shortfall results come mostly from the 30 percent of stocks in the lowest dollar volume group. For the top 70 percent of stocks by dollar volume, it may be difficult to distinguish the model of market microstructure invariance from the model of invariant bet frequency. Second, in the implementation shortfall regressions, the bid-ask spreads decrease with increased activity somewhat faster than the model of market microstructure invariance predicts. Third, our measure of trading activity can be thought of as the product of share volume and price volatility in dollars per share. Although the model predicts that these two components of trading activity should behave similarly, both the implementation shortfall regressions and the trade size regressions suggest that they behave differently. Trading volume (measured in shares) seems to be more consistent with the model of market microstructure invariance than dollar price volatility. It is possible that these issues have something to do with the interaction between tick size effects, the minimum round lot size of 100 shares, nominal stock prices, and volatility.

Interesting issues for further research include testing the three proposed models on different databases.

• The model’s predictions concerning spreads can be tested using quoted spreads from TAQ data. Although it is difficult to measure the level of market depth from TAQ data using, for example, the approach of Lee and Ready (1991), the model’s cross-sectional implications concerning market impact might be testable using this approach. The predictions concerning noise trading quantities can be tested using changes in holdings of mutual funds or other reporting institutional traders.

• If we make the assumption that the flow of information per trading game is constant across stocks and the flow of news articles is proportional to the flow of information, then the hypothesis of market microstructure invariance implies a particular relationship between the number of news articles for individual stocks and the measure of trading activity. If we make the assumption that the flow of analyst forecasts is proportional to the flow of information, then the hypothesis of market microstructure invariance also implies a particular relationship between the number of analyst forecasts for individual stocks and the measure of trading activity. These predictions can be tested as well.
• It is also possible that the model tested on stock data in this paper can be generalized to other markets. For example, market impact and spreads in bond markets, currency markets, or futures markets may be consistent with the regressions estimated for stocks in this paper.

References


Table 1: Descriptive Statistics.

**Panel A: Properties of Securities**

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Med(V) (m $)</td>
<td>19.99</td>
<td>1.22</td>
<td>5.14</td>
<td>9.97</td>
<td>15.92</td>
<td>23.92</td>
<td>31.45</td>
<td>42.11</td>
<td>60.16</td>
<td>101.51</td>
<td>212.55</td>
</tr>
<tr>
<td>Med(σ)</td>
<td>1.89</td>
<td>2.04</td>
<td>2.00</td>
<td>1.92</td>
<td>1.95</td>
<td>1.88</td>
<td>1.85</td>
<td>1.79</td>
<td>1.78</td>
<td>1.76</td>
<td>1.76</td>
</tr>
<tr>
<td>Med(Sprd) (bps)</td>
<td>11.54</td>
<td>38.16</td>
<td>18.34</td>
<td>13.53</td>
<td>11.81</td>
<td>10.12</td>
<td>9.34</td>
<td>8.09</td>
<td>7.16</td>
<td>5.92</td>
<td>4.83</td>
</tr>
<tr>
<td>Mean(Sprd) (bps)</td>
<td>23.67</td>
<td>64.05</td>
<td>31.27</td>
<td>21.83</td>
<td>18.40</td>
<td>15.65</td>
<td>13.86</td>
<td>12.14</td>
<td>11.00</td>
<td>9.02</td>
<td>7.46</td>
</tr>
</tbody>
</table>

**Panel B: Properties of Orders**

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg(X/V) (%)</td>
<td>3.90</td>
<td>15.64</td>
<td>4.58</td>
<td>2.63</td>
<td>1.82</td>
<td>1.36</td>
<td>1.18</td>
<td>1.07</td>
<td>0.88</td>
<td>0.69</td>
<td>0.49</td>
</tr>
<tr>
<td>Med(X/V) (%)</td>
<td>0.56</td>
<td>3.48</td>
<td>1.39</td>
<td>0.80</td>
<td>0.54</td>
<td>0.40</td>
<td>0.35</td>
<td>0.30</td>
<td>0.25</td>
<td>0.20</td>
<td>0.14</td>
</tr>
<tr>
<td>Avg(X/Cap) (bps)</td>
<td>1.68</td>
<td>3.58</td>
<td>2.71</td>
<td>2.08</td>
<td>1.60</td>
<td>1.25</td>
<td>1.06</td>
<td>0.90</td>
<td>0.72</td>
<td>0.56</td>
<td>0.40</td>
</tr>
<tr>
<td>Med(X/Cap) (bps)</td>
<td>0.36</td>
<td>1.00</td>
<td>0.81</td>
<td>0.59</td>
<td>0.42</td>
<td>0.32</td>
<td>0.28</td>
<td>0.23</td>
<td>0.19</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>Avg OMT Share</td>
<td>0.31</td>
<td>0.38</td>
<td>0.33</td>
<td>0.32</td>
<td>0.32</td>
<td>0.31</td>
<td>0.31</td>
<td>0.30</td>
<td>0.29</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td>Avg EC Share</td>
<td>0.40</td>
<td>0.42</td>
<td>0.42</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.40</td>
<td>0.40</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
</tr>
<tr>
<td>Avg IC Share</td>
<td>0.29</td>
<td>0.20</td>
<td>0.25</td>
<td>0.26</td>
<td>0.27</td>
<td>0.28</td>
<td>0.29</td>
<td>0.30</td>
<td>0.32</td>
<td>0.35</td>
<td>0.40</td>
</tr>
<tr>
<td>&lt; 100 shares</td>
<td>0.03</td>
<td>0.05</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>[10pt] ∑ X/V (%)</td>
<td>0.29</td>
<td>0.41</td>
<td>0.40</td>
<td>0.34</td>
<td>0.30</td>
<td>0.27</td>
<td>0.27</td>
<td>0.26</td>
<td>0.24</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td># Obs</td>
<td>441,865</td>
<td>65,081</td>
<td>68,545</td>
<td>41,559</td>
<td>49,532</td>
<td>28,621</td>
<td>30,087</td>
<td>30,710</td>
<td>35,733</td>
<td>42,331</td>
<td>49,666</td>
</tr>
<tr>
<td># Stks</td>
<td>2,439</td>
<td>992</td>
<td>472</td>
<td>217</td>
<td>178</td>
<td>104</td>
<td>125</td>
<td>90</td>
<td>102</td>
<td>81</td>
<td>78</td>
</tr>
</tbody>
</table>

Table reports the characteristics of securities and transition orders in the sample. Panel A shows the median of average daily dollar volume \( V \) (in millions of $), the median of the daily returns volatility \( σ \) in percents, and the median and the mean of the percentage spread \( Sprd \) in basis points. Panel B shows the average and median order size (in percents of \( V \) and in basis points of market capitalization), the average fraction of transition order executed in open market (Avg OMT Share), external and internal crossing networks (Avg EC and IC Shares), the fraction of orders with less than 100 shares traded, the average contribution of transitions to volume each month, as well as the total number of observations and the number of stocks (for the last month). Results are presented for stocks with different dollar trading volume. Group 1 (Group 10) contains orders in stocks with lowest (highest) dollar trading volume. Each month, the observations are split into 10 bins according to stocks’ dollar trading volume in pre-transition month. The thresholds correspond to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar trading volume for common stocks listed on the NYSE. The sample ranges from January 2001 to December 2005.
Table 2: OLS Estimates of Order Size.

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
<td>Sell</td>
<td>Buy</td>
</tr>
<tr>
<td>\ln \left[ \bar{q} \right]</td>
<td>-5.67***</td>
<td>-5.68***</td>
<td>-5.63***</td>
<td>-5.75***</td>
</tr>
<tr>
<td>\alpha_0</td>
<td>-0.63***</td>
<td>-0.63***</td>
<td>-0.60***</td>
<td>-0.71***</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: \( \alpha_0 = -2/3 \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>17.01</td>
<td>13.74</td>
<td>72.00</td>
<td>6.53</td>
<td>18.56</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0107</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Frequency: \( \alpha_0 = 0 \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>5664.91</td>
<td>3740.45</td>
<td>5667.60</td>
<td>1440.32</td>
<td>2427.51</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Size: \( \alpha_0 = -1 \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>1920.13</td>
<td>1306.11</td>
<td>2537.08</td>
<td>229.30</td>
<td>966.99</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\[
\frac{Q^*/V^*}{V_i} = 34.68 \quad 34.08 \quad 35.96 \quad 31.85 \quad 35.27 \\
\frac{d/g/n}{d/1/4389 \quad g/2/1/4018 \quad n/2/1/4198 \quad d/2/1/2855 \quad g/2/1/2977} \\
\frac{#Obs}{441,865 \quad 135,006 \quad 152,701 \quad 69,774 \quad 84,384} \\
\frac{Adj. R^2}{0.3188 \quad 0.2588 \quad 0.2643 \quad 0.4364 \quad 0.3648} \\
\frac{R^2}{0.3188 \quad 0.2588 \quad 0.2643 \quad 0.4364 \quad 0.3648}
\]

Table presents the estimates \( \ln \bar{q} \) and \( \alpha_0 \) for the regression:

\[
\ln \left[ \frac{X_i}{V_i} \right] = \ln \left[ \bar{q} \right] + \alpha_0 \cdot \ln \left[ \frac{W_i}{W^*} \right] + \epsilon.
\]

Each observation corresponds to order \( i \). The left-hand side variable is the logarithm of the order size \( X_i \) as a fraction of expected daily volume \( V_i \). The trading activity \( W_i \) is the product of expected daily volatility \( \sigma_i \), benchmark price \( P_{0,i} \), and expected daily volume \( V_i \), measured as the last month’s average daily volume. The scaling constant \( W^* = (0.02)(40)(10^6) \) corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. \( \bar{q} \) is the measure of order size such that the median order size \( Q^*/V^* \) for a benchmark stock in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with \( d \) parameters, \( g \) restrictions, and \( n \) clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.
Table 3: Quantile Estimates of Order Size.

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>p1</th>
<th>p5</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>p95</th>
<th>p99</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\ln [\bar{q}])</td>
<td>-9.38***</td>
<td>-8.31***</td>
<td>-6.72***</td>
<td>-5.65***</td>
<td>-4.59***</td>
<td>-3.06***</td>
<td>-2.06***</td>
</tr>
<tr>
<td>(\alpha_0)</td>
<td>-0.69***</td>
<td>-0.67***</td>
<td>-0.63***</td>
<td>-0.63***</td>
<td>-0.62***</td>
<td>-0.64***</td>
<td>-0.62***</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: \(\alpha_0 = -2/3\)

| F-test | 26       | 1        | 447      | 495      | 699      | 67       | 75       |
| p-val  | 0.0000   | 0.2572   | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   |

Model of Invariant Bet Frequency: \(\alpha_0 = 0\)

| F-test | 22.824   | 43.129   | 110.000  | 140.000  | 100.000  | 47.189   | 16.251   |
| p-val  | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   |

Model of Invariant Bet Size: \(\alpha_0 = -1\)

| p-val  | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   | 0.0000   |

| \(Q^*/V^*\) | 0.84   | 2.46   | 12.07   | 35.18   | 101.53  | 468.88  | 1274.54  |
| Pseudo \(R^2\) | 0.1682 | 0.1558 | 0.1663 | 0.1750 | 0.1806 | 0.1937 | 0.2186 |
| #Obs     | 441,865 | 441,865 | 441,865 | 441,865 | 441,865 | 441,865 | 441,865 |

Table presents the estimates \(\ln \bar{q}\) and \(\alpha_0\) for the quantile regression:

\[
\ln \left[ \frac{X_i}{V_i} \right] = \ln [\bar{q}] + \alpha_0 \cdot \ln \left[ \frac{W_i}{W^*} \right] + \tilde{\epsilon}.
\]

Each observation corresponds to order \(i\). The left-hand side variable is the logarithm of the order size \(X_i\) as a fraction of expected daily volume \(V_i\). The trading activity \(W_i\) is the product of expected daily volatility \(\sigma_i\), benchmark price \(P_{0,i}\), and expected daily volume \(V_i\), measured as the last month’s average daily volume. The scaling constant \(W^* = (0.02)(40)(10^6)\) corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. \(\bar{q}\) is the measure of order size such that the median order size \(Q^*/V^*\) for a benchmark stock in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with \(d\) parameters, \(g\) restrictions, and \(n\) clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
<td>Sell</td>
<td>Buy</td>
</tr>
<tr>
<td>$\ln [q]$</td>
<td>-5.65***</td>
<td>-5.66***</td>
<td>-5.58***</td>
<td>-5.80***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.018)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.25***</td>
<td>0.30***</td>
<td>0.36**</td>
<td>0.17*</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.041)</td>
<td>(0.036)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.16***</td>
<td>0.11***</td>
<td>0.21***</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.018)</td>
<td>(0.014)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>$b_3$</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03**</td>
<td>-0.07***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.009)</td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: $b_1 = b_2 = b_3 = 0$

| F-test | 47.57 | 19.92 | 79.29 | 8.77 | 20.97 |
| p-val  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Model of Invariant Bet Frequency: $b_1 = b_2 = b_3 = 2/3$

| F-test | 2044.60 | 1286.85 | 1939.03 | 567.85 | 808.05 |
| p-val  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Model of Invariant Bet Size: $b_1 = b_2 = b_3 = -1/3$

| F-test | 747.74 | 465.71 | 947.01 | 78.19 | 325.63 |
| p-val  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $Q^*/V^*$ | 35.31 | 34.98 | 32.99 | 30.35 | 35.95 |
| #Obs   | 441,865 | 135,006 | 152,701 | 69,774 | 84,384 |
| Adj. $R^2$ | 0.3211 | 0.2614 | 0.2682 | 0.4382 | 0.3674 |
| $R^2$  | 0.3213 | 0.2616 | 0.2684 | 0.4384 | 0.3676 |

Each observation corresponds to order $i$. The left-hand side variable is the logarithm of the order size $X_i$ as a fraction of expected daily volume $V_i$. The trading activity $W_i$ is the product of expected daily volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected daily volume $V_i$, measured as the last month's average daily volume. The scaling constant $W^* = (0.02)(40)(10^6)$ corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. $q$ is the measure of order size such that the median order size $Q^*/V^*$ for a benchmark stock in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. *** , ** , * denotes significance at 1%, 5% and 10% levels, respectively.
Table 5: OLS Estimates of Order Size: Comparison of Three Models.

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
<td>Sell</td>
<td>Buy</td>
</tr>
<tr>
<td><strong>Model of Market Microstructure Invariance: ( \alpha_0 = -\frac{2}{3} )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln [\bar{q}] )</td>
<td>-5.69***</td>
<td>-5.70***</td>
<td>-5.67***</td>
<td>-5.70***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.022)</td>
<td>(0.019)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>( Q/V^* )</td>
<td>33.75</td>
<td>33.35</td>
<td>34.61</td>
<td>33.60</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.3177</td>
<td>0.2577</td>
<td>0.2608</td>
<td>0.4343</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.3178</td>
<td>0.2579</td>
<td>0.2609</td>
<td>0.4345</td>
</tr>
<tr>
<td>( \log(L) )</td>
<td>-828,757</td>
<td>-255,637</td>
<td>-287,149</td>
<td>-127,591</td>
</tr>
<tr>
<td><strong>Model of Invariant Bet Frequency: ( \alpha_0 = 0 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln [\bar{q}] )</td>
<td>-5.17***</td>
<td>-5.33***</td>
<td>-5.29***</td>
<td>-4.95***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.025)</td>
<td>(0.021)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>( Q/V^* )</td>
<td>56.85</td>
<td>48.44</td>
<td>50.42</td>
<td>70.83</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>-0.0002</td>
<td>-0.0002</td>
<td>-0.0002</td>
<td>-0.0004</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \log(L) )</td>
<td>-913,255</td>
<td>-275,771</td>
<td>-310,235</td>
<td>-147,479</td>
</tr>
<tr>
<td><strong>Model of Invariant Bet Size: ( \alpha_0 = -1 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ln [\bar{q}] )</td>
<td>-5.95***</td>
<td>-5.89***</td>
<td>-5.86***</td>
<td>-6.07***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.022)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>( Q/V^* )</td>
<td>26.05</td>
<td>27.67</td>
<td>28.51</td>
<td>23.11</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.2105</td>
<td>0.1683</td>
<td>0.1458</td>
<td>0.3669</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.2107</td>
<td>0.1685</td>
<td>0.1460</td>
<td>0.3671</td>
</tr>
<tr>
<td>( \log(L) )</td>
<td>-860,973</td>
<td>-263,318</td>
<td>-298,186</td>
<td>-131,520</td>
</tr>
<tr>
<td>#Obs</td>
<td>441,865</td>
<td>135,006</td>
<td>152,701</td>
<td>69,774</td>
</tr>
</tbody>
</table>

Table presents the estimates \( \ln [\bar{q}] \) for the regression:

\[
\ln \left( \frac{X_i}{V_i} \right) = \ln [\bar{q}] + \alpha_0 \cdot \ln \left( \frac{W_i}{W^*} \right) + \tilde{\epsilon},
\]

with \( \alpha_0 \) restricted to be as predicted in proposed models. Each observation corresponds to order \( i \). The left-hand side variable is the logarithm of the order size \( X_i \) as a fraction of expected daily volume \( V_i \). The trading activity \( W_i \) is the product of expected daily volatility \( \sigma_i \), benchmark price \( P_{0,i} \), and expected daily volume \( V_i \), measured as the last month’s average daily volume. The scaling constant \( W^* = (0.02)(40)(10^6) \) corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. \( \bar{q} \) is the measure of order size such that the median order size \( Q/V^* \) for a benchmark stock in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The log-likelihoods and \( R^2 \) are shown for each model. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.
Table 6: OLS Estimates of Log of Quoted Spread.

<table>
<thead>
<tr>
<th></th>
<th>NYSE All</th>
<th>NYSE Buy</th>
<th>NYSE Sell</th>
<th>NASDAQ Buy</th>
<th>NASDAQ Sell</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \left[ \bar{k}/(40 \cdot 0.02) \right]$</td>
<td>-3.07***</td>
<td>-3.09***</td>
<td>-3.08***</td>
<td>-3.04***</td>
<td>-3.04***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.36***</td>
<td>-0.31***</td>
<td>-0.31***</td>
<td>-0.40***</td>
<td>-0.40***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: $\alpha_0 = -1/3$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>81.55</td>
<td>0.0000</td>
<td>53.36</td>
<td>0.0000</td>
<td>40.50</td>
<td>0.0000</td>
<td>221.98</td>
<td>0.0000</td>
<td>273.88</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Frequency: $\alpha_0 = 0$

<table>
<thead>
<tr>
<th></th>
<th>F-test</th>
<th>p-val</th>
<th>F-test</th>
<th>p-val</th>
<th>F-test</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19,682.13</td>
<td>0.0000</td>
<td>10,594.71</td>
<td>0.0000</td>
<td>9,128.40</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Size: $\alpha_0 = -1/2$

<table>
<thead>
<tr>
<th></th>
<th>F-test</th>
<th>p-val</th>
<th>F-test</th>
<th>p-val</th>
<th>F-test</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3,203.68</td>
<td>0.0000</td>
<td>3,896.59</td>
<td>0.0000</td>
<td>3,285.33</td>
<td>0.0000</td>
</tr>
</tbody>
</table>


$d/g/n$             | 2/1/4388 | 2/1/4015 | 2/1/4197 | 2/1/2851 | 2/1/2968 |

Adj. $R^2$          | 0.4670  | 0.3545   | 0.3822   | 0.5586   | 0.5729  |

#Obs               | 437,210 | 134,147  | 151,549  | 68,745   | 82,769  |

Table presents the estimates $\ln \bar{q}$ and $\alpha_3$ for the regression:

$$\ln \left[ \frac{\kappa_i}{P_{0,i} \sigma_i} \right] = \ln \left[ \frac{\bar{k}}{40 \cdot 0.02} \right] + \alpha_3 \cdot \ln \left[ \frac{W_i}{W^*} \right] + \tilde{\epsilon}.$$  

Each observation corresponds to order $i$. The left-hand side variable is the logarithm of the quoted bid-ask spread $\kappa_i / P_{0,i}$ as a fraction of expected return volatility $\sigma_i$. The trading activity $W_i$ is the product of expected daily volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected daily volume $V_i$, measured as the last month’s average daily volume. The scaling constant $W^* = (0.02)(40)(10^6)$ corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. $\kappa_i / P_{0,i}$ is the percentage quoted spread. $\bar{k}$ is the measure of order size such that the median percentage spread for a benchmark stock is calculated as $\bar{k}/40$. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.
Table 7: Impact Estimates in Nested Non-Linear Regression with Quoted Spread.

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
<td>Sell</td>
<td>Buy</td>
</tr>
<tr>
<td>$\frac{1}{2}\bar{\lambda}$</td>
<td>2.97***</td>
<td>3.16***</td>
<td>2.31***</td>
<td>4.62***</td>
</tr>
<tr>
<td></td>
<td>(0.255)</td>
<td>(0.591)</td>
<td>(0.380)</td>
<td>(0.777)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.31***</td>
<td>0.22***</td>
<td>0.32***</td>
<td>0.32***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.041)</td>
<td>(0.054)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>$s$</td>
<td>0.62***</td>
<td>0.91***</td>
<td>0.54***</td>
<td>0.60***</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.143)</td>
<td>(0.117)</td>
<td>(0.110)</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: $\alpha_1 = 1/3$

<table>
<thead>
<tr>
<th>F-test</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>0.3621</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Frequency: $\alpha_1 = 0$

<table>
<thead>
<tr>
<th>F-test</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>151.97</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Size: $\alpha_1 = 1/2$

<table>
<thead>
<tr>
<th>F-test</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.72</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d/g/n$</th>
<th>#Obs</th>
<th>$R^2$</th>
<th>Adj. $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/1/4388</td>
<td>438,885</td>
<td>0.0112</td>
<td>0.0110</td>
</tr>
<tr>
<td>3/1/4015</td>
<td>134,560</td>
<td>0.0114</td>
<td>0.0111</td>
</tr>
<tr>
<td>3/1/4197</td>
<td>151,958</td>
<td>0.0063</td>
<td>0.0060</td>
</tr>
<tr>
<td>3/1/2851</td>
<td>69,131</td>
<td>0.0183</td>
<td>0.0179</td>
</tr>
<tr>
<td>3/1/2969</td>
<td>83,236</td>
<td>0.0175</td>
<td>0.0175</td>
</tr>
</tbody>
</table>

Table presents the estimates for $\bar{\lambda}$, $\alpha_1$, and $s$ in the regression:

$$I_{BS,i}(P_{ex,i} - P_{0,i})/P_{0,i} \cdot 10^4 \cdot (0.02) \cdot \frac{(W_i)^{\alpha_1}}{\sigma_i} \cdot \frac{X_i}{(0.01)V_i} + s \cdot \frac{\kappa_i}{P_{0,i}} \cdot (0.02) \cdot (X_{omt,i} + X_{ec,i}) + \epsilon.$$

Each observation corresponds to order $i$. The left-hand side variable is the implementation shortfall in basis points, where $I_{BS,i}$ is a buy/sell indicator, $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity $W_i$ is the product of expected volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected volume $V_i$. The scaling constant $W^* = (0.02)(40)(10^6)$ is the trading activity for the benchmark stock with volatility of 0.02, price $40$ per share, and trading volume of one million shares per day. $X_i$ is the number of shares in the order with $X_{omt,i}$ and $X_{ec,i}$ shares executed in open market and external crossing networks, respectively. $\kappa_i/P_{0,i}$ is the quoted spread. $\bar{\lambda}/2$ is the market impact costs of executing a trade of one percent of daily volume in a benchmark stock, in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.
Table 8: Impact and Spread Estimates in Nested Non-Linear Regression.

<table>
<thead>
<tr>
<th></th>
<th>NYSE</th>
<th></th>
<th>NASDAQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
<td>Sell</td>
<td>Buy</td>
</tr>
<tr>
<td>$\frac{1}{2}\lambda$</td>
<td>2.85***</td>
<td>2.50***</td>
<td>2.33***</td>
<td>4.20***</td>
</tr>
<tr>
<td></td>
<td>(0.245)</td>
<td>(0.515)</td>
<td>(0.365)</td>
<td>(0.753)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.33***</td>
<td>0.18***</td>
<td>0.33***</td>
<td>0.33***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.045)</td>
<td>(0.054)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>$\frac{1}{2}\kappa$</td>
<td>6.30***</td>
<td>14.94***</td>
<td>2.82*</td>
<td>8.38*</td>
</tr>
<tr>
<td></td>
<td>(1.131)</td>
<td>(2.529)</td>
<td>(1.394)</td>
<td>(3.328)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.39***</td>
<td>-0.19***</td>
<td>-0.46***</td>
<td>-0.36***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.045)</td>
<td>(0.061)</td>
<td>(0.061)</td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: $\alpha_1 = 1/3, \alpha_2 = -1/3$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>2.62</td>
<td>8.51</td>
<td>2.25</td>
<td>0.09</td>
<td>3.12</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0731</td>
<td>0.0002</td>
<td>0.1057</td>
<td>0.9114</td>
<td>0.0443</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Frequency: $\alpha_1 = 0, \alpha_2 = 0$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>176.14</td>
<td>14.79</td>
<td>47.03</td>
<td>33.11</td>
<td>71.06</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Model of Invariant Bet Size: $\alpha_1 = 1/2, \alpha_2 = -1/2$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test</td>
<td>30.30</td>
<td>39.63</td>
<td>5.23</td>
<td>7.21</td>
<td>5.92</td>
</tr>
<tr>
<td>p-val</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0054</td>
<td>0.0007</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

$\frac{d}{g/n}$ | 4/2/4389 | 4/2/4018 | 4/2/4198 | 4/2/2855 | 4/2/2977 |

$\#$Obs        | 441,865 | 135,006 | 152,701 | 69,774   | 84,384   |

$R^2$          | 0.0126  | 0.0136  | 0.0067   | 0.0211   | 0.0195   |

Adj. $R^2$     | 0.0123  | 0.0134  | 0.0064   | 0.0208   | 0.0192   |

Table presents the estimates for $\tilde{\lambda}, \tilde{\kappa}, \alpha_1$, and $\alpha_2$ in the regression:

$$\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot \frac{0.02}{\sigma_i} = \frac{1}{2}\tilde{\lambda} \left[ \frac{W_i}{W^*} \right]^{\alpha_1} \frac{X_i}{(0.01) V_i} + \frac{1}{2}\tilde{\kappa} \left[ \frac{X_{omt,i} + X_{ec,i}}{X_i} \right] \left[ \frac{W_i}{W^*} \right]^{\alpha_2} + \tilde{\epsilon}.$$  

Each observation corresponds to order $i$. The left-hand side variable is the implementation shortfall in basis points, where $I_{BS,i}$ is a buy/sell indicator, $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity $W_i$ is the product of expected volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected volume $V_i$. The scaling constant $W^* = (0.02)(40)(10^6)$ is the trading activity for the benchmark stock with volatility of 2% per day, price $40$ per share, and trading volume of one million shares per day. $X_i$ is the number of shares in the order with $X_{omt,i}$ and $X_{ec,i}$ shares executed in open market and external crossing networks, respectively. $\lambda/2$ is the market impact costs of executing a trade of one percent of daily volume in a benchmark stock, and $\kappa/2$ is the effective spread cost. Both $\lambda/2$ and $\kappa/2$ are in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, * denotes significance at 1%, 5% and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Buy</th>
<th>Sell</th>
<th>NYSE</th>
<th>Buy</th>
<th>Sell</th>
<th>NASDAQ</th>
<th>Buy</th>
<th>Sell</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{2}\lambda_{omt}$</td>
<td>4.49***</td>
<td>2.41***</td>
<td>4.35***</td>
<td>5.39***</td>
<td>4.33***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.601)</td>
<td>(0.433)</td>
<td>(0.798)</td>
<td>(1.449)</td>
<td>(1.058)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}\lambda_{ec}$</td>
<td>2.17***</td>
<td>3.02***</td>
<td>1.77**</td>
<td>3.64***</td>
<td>1.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.386)</td>
<td>(0.471)</td>
<td>(0.542)</td>
<td>(0.968)</td>
<td>(0.705)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}\lambda_{ic}$</td>
<td>2.40***</td>
<td>2.20***</td>
<td>1.77***</td>
<td>2.07*</td>
<td>1.51**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.334)</td>
<td>(0.622)</td>
<td>(0.420)</td>
<td>(0.891)</td>
<td>(0.504)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.31</td>
<td>-0.86***</td>
<td>-0.37</td>
<td>-0.10</td>
<td>-1.05***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.141)</td>
<td>(0.327)</td>
<td>(0.347)</td>
<td>(0.284)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.22*</td>
<td>-0.01</td>
<td>-0.43*</td>
<td>-0.17</td>
<td>-0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.141)</td>
<td>(0.327)</td>
<td>(0.347)</td>
<td>(0.284)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.04</td>
<td>-0.19***</td>
<td>0.13</td>
<td>0.01</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.031)</td>
<td>(0.077)</td>
<td>(0.053)</td>
<td>(0.037)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}\kappa_{omt}$</td>
<td>6.56***</td>
<td>18.55***</td>
<td>3.05*</td>
<td>14.43***</td>
<td>4.69*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.227)</td>
<td>(3.539)</td>
<td>(1.279)</td>
<td>(3.967)</td>
<td>(1.819)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}\kappa_{ec}$</td>
<td>6.26***</td>
<td>8.99***</td>
<td>4.98**</td>
<td>11.13**</td>
<td>5.08**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.124)</td>
<td>(2.355)</td>
<td>(1.520)</td>
<td>(3.690)</td>
<td>(1.841)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{2}\kappa_{ic}$</td>
<td>0.26</td>
<td>5.31</td>
<td>-4.38**</td>
<td>7.78</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.846)</td>
<td>(3.987)</td>
<td>(1.429)</td>
<td>(7.699)</td>
<td>(1.317)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.10</td>
<td>-0.06</td>
<td>0.60*</td>
<td>-0.29</td>
<td>0.99***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
<td>(0.249)</td>
<td>(0.260)</td>
<td>(0.297)</td>
<td>(0.258)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.18**</td>
<td>-0.22</td>
<td>0.06</td>
<td>0.26</td>
<td>0.36***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.172)</td>
<td>(0.123)</td>
<td>(0.139)</td>
<td>(0.104)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.09***</td>
<td>0.26***</td>
<td>-0.12*</td>
<td>0.05</td>
<td>-0.11*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.055)</td>
<td>(0.054)</td>
<td>(0.050)</td>
<td>(0.049)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model of Market Microstructure Invariance: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$.

F-test 4.59 20.96 2.80 1.38 11.26
p-val 0.0001 0.0000 0.0102 0.2168 0.0000

Model of Invariant Bet Frequency: $\beta_1 = \beta_2 = \beta_3 = -1/3, \beta_4 = \beta_5 = \beta_6 = 1/3$.

F-test 78.26 12.06 26.46 10.71 23.27
p-val 0.0000 0.0000 0.0000 0.0000 0.0000

Model of Invariant Bet Size: $\beta_1 = \beta_2 = \beta_3 = 1/6, \beta_4 = \beta_5 = \beta_6 = -1/6$.

F-test 13.77 44.25 5.94 6.85 28.79
p-val 0.0000 0.0000 0.0000 0.0000 0.0000

<table>
<thead>
<tr>
<th>d/g/n</th>
<th>12/6/4389</th>
<th>12/6/4018</th>
<th>12/6/4198</th>
<th>12/6/2855</th>
<th>12/6/2977</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Obs</td>
<td>441,865</td>
<td>135,006</td>
<td>152,701</td>
<td>69,774</td>
<td>84,384</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.0129</td>
<td>0.0147</td>
<td>0.0076</td>
<td>0.0222</td>
<td>0.0214</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0131</td>
<td>0.0150</td>
<td>0.0079</td>
<td>0.0225</td>
<td>0.0217</td>
</tr>
</tbody>
</table>

58
Table presents the estimates for $\bar{\lambda}_{omt}, \bar{\lambda}_{ec}, \bar{\lambda}_{ic}, \bar{\kappa}_{omt}, \bar{\kappa}_{ec}, \bar{\kappa}_{ic}, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ in the regression:

$$\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot \frac{(0.02)}{\sigma_i}$$

$$= \frac{1}{2} \cdot \frac{\bar{\lambda}_{omt,i}X_{omt,i} + \bar{\lambda}_{ec,i}X_{ec,i} + \bar{\lambda}_{ic,i}X_{ic,i}}{(0.01)V_i} \cdot \left[ \frac{W_i}{W^*} \right]^{1/3} \cdot \frac{\sigma_i^\beta_1 \cdot P_{0,i}^\beta_2 \cdot V_i^\beta_3}{(0.02)(40)(10^6)}$$

$$+ \frac{1}{2} \cdot \frac{\bar{\kappa}_{omt,i}X_{omt,i} + \bar{\kappa}_{ec,i}X_{ec,i} + \bar{\kappa}_{ic,i}X_{ic,i}}{X_i} \cdot \left[ \frac{W_i}{W^*} \right]^{-1/3} \cdot \frac{\sigma_i^\beta_4 \cdot P_{0,i}^\beta_5 \cdot V_i^\beta_6}{(0.02)(40)(10^6)} + \bar{\epsilon}.$$  

Each observation corresponds to order $i$. The left-hand side variable is the implementation shortfall in basis points, where $I_{BS,i}$ is a buy/sell indicator, $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity $W_i$ is the product of expected volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected volume $V_i$. The scaling constant $W^* = (0.02)(40)(10^6)$ is the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. $X_i$ is the number of shares in the order with $X_{omt,i}$, $X_{ec,i}$ and $X_{ic,i}$ shares executed in open market, external crossing networks, and internal crossing networks, respectively. $\bar{\lambda}_{omt}/2, \bar{\lambda}_{ec}/2, \bar{\lambda}_{ic}/2$ are the market impact costs of executing a trade of one percent of daily volume in a benchmark stock using open market trades, external crosses, and internal crosses, respectively. $\bar{\kappa}_{omt}/2, \bar{\kappa}_{ec}/2, \bar{\kappa}_{ic}/2$ are the effective spread costs for corresponding trading venues. Both market impact and spread costs are in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. *** , ** , * denotes significance at 1% , 5% and 10% levels, respectively.
Table 10: Market Impact and Spread Estimates: Comparison of Three Models.

<table>
<thead>
<tr>
<th>Model of Market Microstructure Invariance: $\alpha_1 = 1/3, \alpha_2 = -1/3$</th>
<th>NYSE</th>
<th>NASDAQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>2.8898***</td>
<td>3.4199***</td>
</tr>
<tr>
<td>(0.1948)</td>
<td>(0.4421)</td>
<td>(0.3231)</td>
</tr>
<tr>
<td>$\hat{\kappa}$</td>
<td>7.9036***</td>
<td>10.9695***</td>
</tr>
<tr>
<td>(0.6889)</td>
<td>(1.3724)</td>
<td>(1.1065)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.0122</td>
<td>0.0129</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0125</td>
<td>0.0131</td>
</tr>
<tr>
<td>$log(L)$</td>
<td>1061,808</td>
<td>324,590</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model of Invariant Bet Frequency: $\alpha_1 = 0, \alpha_2 = 0$</th>
<th>NYSE</th>
<th>NASDAQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>0.3788***</td>
<td>1.3291***</td>
</tr>
<tr>
<td>(0.0884)</td>
<td>(0.1130)</td>
<td>(0.1041)</td>
</tr>
<tr>
<td>$\hat{\kappa}$</td>
<td>15.2686***</td>
<td>19.1143***</td>
</tr>
<tr>
<td>(1.5658)</td>
<td>(2.5247)</td>
<td>(2.3157)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.0075</td>
<td>0.0126</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0077</td>
<td>0.0128</td>
</tr>
<tr>
<td>$log(L)$</td>
<td>1060,752</td>
<td>324,568</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model of Invariant Bet Size: $\alpha_1 = 1/2, \alpha_2 = -1/2$</th>
<th>NYSE</th>
<th>NASDAQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Buy</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>3.9202***</td>
<td>3.9578***</td>
</tr>
<tr>
<td>(0.3036)</td>
<td>(0.6274)</td>
<td>(0.4339)</td>
</tr>
<tr>
<td>$\hat{\kappa}$</td>
<td>3.4648***</td>
<td>5.7598***</td>
</tr>
<tr>
<td>(0.2917)</td>
<td>(0.6487)</td>
<td>(0.4105)</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.0110</td>
<td>0.0108</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0112</td>
<td>0.0110</td>
</tr>
<tr>
<td>$log(L)$</td>
<td>1061,533</td>
<td>324,446</td>
</tr>
<tr>
<td>#Obs</td>
<td>441,865</td>
<td>135,006</td>
</tr>
</tbody>
</table>
Table presents the estimates $\bar{\lambda}, \bar{\kappa}$ for the regression:

$$\frac{I_{BS;i}(P_{ex;i} - P_{0;i})}{P_{0;i}} \cdot 10^4 \cdot \frac{(0.02)}{\sigma_i} = \frac{1}{2} \frac{W_i}{W^*} \cdot \frac{X_i}{(0.01)V_i} + \frac{1}{2} \cdot \frac{\bar{\kappa}}{X_i} \cdot \frac{(X_{omt;i} + X_{ec;i})}{W^*} \cdot \frac{W_i}{W^*} + \epsilon,$$

with $\alpha_1$ and $\alpha_2$ restricted to be as predicted in proposed models. Each observation corresponds to order $i$. The left-hand side variable is the implementation shortfall in basis points, where $I_{BS;i}$ is a buy/sell indicator, $P_{ex;i}$ and $P_{0;i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity $W_i$ is the product of expected volatility $\sigma_i$, benchmark price $P_{0;i}$, and expected volume $V_i$. The scaling constant $W^* = (0.02)(40)(10^6)$ is the trading activity for the benchmark stock with volatility of 2% per day, price $40$ per share, and trading volume of one million shares per day. $X_i$ is the number of shares in the order with $X_{omt;i}$ and $X_{ec;i}$ shares executed in open market and external crossing networks, respectively. $\bar{\lambda}/2$ is the market impact costs of executing a trade of one percent of daily volume in a benchmark stock, and $\bar{\kappa}/2$ is the effective spread cost. Both $\bar{\lambda}/2$ and $\bar{\kappa}/2$ are in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. F-statistics and p-values are reported for three models with $d$ parameters, $g$ restrictions, and $n$ clusters in the regression. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively. The log-likelihoods and $R^2$ are shown for each model. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively.
The figure illustrates the assumptions of our models. The trading game for the benchmark stock (top chart) is compared to the trading game for a stock with trading volume eight times larger than that of the benchmark stock (bottom chart). The benchmark stock has four bets expected per calendar day; buy orders are marked in green color, and sell orders are marked in red color. Prices and volatilities are assumed to be the same for both stocks. Thus, the length of the bars is proportional to dollar bet size. The model of market microstructure invariance assumes that the time clock operates four times faster for the more active stock. This generates four times as many bets per day. To prevent speeding up the time clock from increasing volatility in the model of “market microstructure invariance,” the stock has been delevered by a factor of two to keep volatility constant. This delevering multiplies dollar bet size by a factor of two. The model of “invariant bet frequency” assumes that the difference in volume comes from the difference in bet sizes. The model of “invariant bet size” assumes that the difference in volume comes from the difference in bet frequencies.
Figure 2: Invariant Order Size Distribution.

The figure shows a distribution of transition order sizes, adjusted for differences in the trading activity, for stocks sorted into 10 volume groups and 5 volatility groups (only volume groups 1, 4, 7, 9, 10 and volatility groups 1, 3, 5 are reported). The adjustment is done according to the model of market microstructure invariance, i.e.

\[ \ln \left( \frac{\tilde{X}}{V} \cdot \left( \frac{W_i}{W^*} \right)^{2/3} \right), \]

where \( X_i \) is an order size in shares, \( V_i \) is the average daily volume in shares, and \( W_i \) is the measure of trading activity. The ten volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. The five volatility groups are based on thresholds corresponding to 20th, 40th, 60th, and 80th percentiles for common NYSE-listed stocks. Volatility group 1 (group 5) has stocks with the lowest (highest) volatility. Each subplot also shows the number of observations (\( N \)), the mean (\( m \)), the variance (\( v \)), the skewness (\( s \)), and the kurtosis (\( k \)) for depicted distribution. The normal distribution with the common mean of -5.69 and variance of 2.50 is imposed on each subplot. The common mean and variance are calculated as the mean and variance of adjusted order sizes for the entire sample. The sample ranges from January 2001 to December 2005.
Figure 3: Invariant Order Size Distribution.


Panel B: Logarithm of Ranks against Quantiles of Empirical Distribution.

Panel A shows quantile-quantile plots of empirical distributions of transition orders re-scaled according to the invariance theory, \( \ln \left( \frac{\bar{X}}{V} \cdot \left( \frac{W_i}{W^*} \right)^{2/3} \right) \), and a normal distribution for stocks sorted into 10 volume groups (only volume groups 1, 4, 7, 9, 10 are reported). Panel B depicts the logarithm of ranks based on the re-scaled order sizes. The ten volume groups are based on average dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Volume group 1 (group 10) has stocks with the lowest (highest) trading volume. Each subplot shows the number of observations \( (N) \), the mean \( (m) \), the variance \( (v) \), the skewness \( (s) \), and the kurtosis \( (k) \) of a depicted distribution. There are >400,000 data points. The sample ranges from January 2001 to December 2005.
Figure 4: Dummy Variables as a Robustness Check in OLS Regression for Order Size.

Figure shows the dummy variables \( \ln \bar{q}_j \) with the 95%-confidence intervals for the ten volume groups and three proposed models from regression

\[
\ln \left( \frac{X_i}{V_i} \right) = \sum_{j=1}^{10} \mathbb{I}_{j,i} \cdot \ln \left( \bar{q}_j \right) + \alpha_0 \cdot \ln \left( \frac{W_i}{W^*} \right) + \tilde{\epsilon},
\]

where \( j \)th dummy variable corresponds to the average logarithm of the order size \( X_i \) as a fraction of expected daily volume \( V_i \) for \( j \)th volume group. Each observation corresponds to order \( i \). \( \mathbb{I}_{j,i} \) is an indicator equal to one if order \( i \) is executed in a stock from volume group \( j \). In the model of market microstructure invariance, \( \alpha_0 = -2/3 \); in the model of invariant bet frequency, \( \alpha_0 = 0 \); in the model of invariant bet size, \( \alpha_0 = -1 \). The trading activity \( W_i \) is the product of expected daily volatility \( \sigma_i \), benchmark price \( P_{0,i} \), and expected daily volume \( V_i \), measured as the last month’s average daily volume. The scaling constant \( W^* = (0.02)(40)(10^6) \) corresponds to the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. The ten volume groups are based on the pre-transition dollar trading volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume. The standard errors are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. On each subplot, the point estimate of \( \ln \bar{q} \) from Table 5 is superimposed for each model.
Figure 5: Market Impact and Spread: Dummy Variables as Robustness Check.

Figure shows the estimates of half market impact $\lambda_j/2$ (top) and half spread $\bar{\kappa}_j/2$ (bottom) with the 95%-confidence intervals for the ten volume groups and three proposed models from the regression:

$$\frac{I_{BS,i}(P_{ex,i} - P_{0,i})}{P_{0,i}} \cdot 10^4 \cdot \frac{0.02}{\sigma_i} = \left( \sum_{j=1}^{10} I_{j,i} \cdot \frac{1}{2} \bar{\lambda}_j \right) \cdot \left[ \frac{W_i}{W^*} \right]^{\alpha_1} \cdot \frac{X_i}{(0.01)V_i} + \left( \sum_{j=1}^{10} I_{j,i} \cdot \frac{1}{2} \bar{\kappa}_j \right) \cdot \frac{(X_{omt,i} + X_{ec,i})}{X_i} \cdot \left[ \frac{W_i}{W^*} \right]^{\alpha_2} + \tilde{\epsilon},$$

where $\bar{\lambda}_j/2$ and $\bar{\kappa}_j/2$ correspond to the market impact costs of a trade of one percent of daily volume and the spread cost for $j$th volume group; both in basis points. Each observation corresponds to order $i$. $I_{j,i}$ is an indicator equal to one if order $i$ is executed in a stock from volume group $j$. In the model of market microstructure invariance, $\alpha_1 = 1/3$, $\alpha_2 = -1/3$. In the model of invariant bet frequency, $\alpha_1 = 0$, $\alpha_2 = 0$. In the model of invariant bet size, $\alpha_1 = 1/2$, $\alpha_2 = -1/2$. The left-hand side variable is the implementation shortfall in basis points, where $I_{BS,i}$ is a buy/sell indicator, $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i$ adjusts for heteroscedasticity. The trading activity $W_i$ is the product of expected volatility $\sigma_i$, benchmark price $P_{0,i}$, and expected volume $V_i$. The scaling constant $W^* = (0.02)(40)(10^6)$ is the trading activity for the benchmark stock with volatility of 2% per day, price $40 per share, and trading volume of one million shares per day. $X_i$ is the number of shares in the order with $X_{omt,i}$ and $X_{ec,i}$ shares executed in open market and external crossing networks, respectively. The ten volume groups are based on the pre-transition dollar volume with thresholds corresponding to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) volume. The standard errors are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. On each subplot, the point estimates of $\bar{\lambda}_j/2$ and $\bar{\kappa}_j/2$ from Table 10 are superimposed as horizontal lines for each model.