Market Microstructure Invariance: Empirical Hypotheses*

Albert S. Kyle
University of Maryland
akyle@rhsmith.umd.edu

Anna A. Obizhaeva
New Economic School
aobizhaeva@nes.ru

March 1, 2016†

Abstract

Using the intuition that financial markets transfer risks in business time, "market microstructure invariance" is defined as the hypotheses that the distributions of risk transfers ("bets") and transaction costs are constant across assets when measured per unit of business time. The invariance hypotheses imply that bet size and transaction costs have specific, empirically testable relationships to observable dollar volume and volatility. Portfolio transitions can be viewed as natural experiments for measuring transaction costs, and individual orders can be treated as proxies for bets. Empirical tests based on a dataset of 400,000+ portfolio transition orders support the invariance hypotheses. The constants calibrated from structural estimation imply specific predictions for the arrival rate of bets ("market velocity"), the distribution of bet sizes, and transaction costs.

Keywords: market microstructure, liquidity, bid-ask spread, market impact, transaction costs, order size, invariance, structural estimation.

*We are grateful to Elena Asparouhova, Peter Bossaerts, Xavier Gabaix, Lawrence Glosten, Larry Harris, Pankaj Jain, Mark Loewenstein, Natalie Popovic, Sergey N. Smirnov, Georgios Skoulakis, Vish Viswanathan, and Wenyuan Xu for helpful comments. Obizhaeva is also grateful to the Paul Woolley Center at the London School of Economics for its hospitality as well as Ross McLellan, Simon Myrgren, Sébastien Page, and especially Mark Kritzman for their help. Kyle has worked as a consultant for various companies, exchanges, and government agencies. He is a non-executive director of a U.S.-based asset management company.

†This paper supersedes a previous manuscript, "Market Microstructure Invariance: Theory and Empirical Tests" (October 17, 2014), which also contains a theoretical model which is now described in Kyle and Obizhaeva (2016*b*). The previous paper was a revised version of an earlier manuscript (June 7, 2013) which combined and superseded two earlier papers: the theoretical paper "Market Microstructure Invariants: Theory and Implications of Calibration" (December 12, 2011) and the empirical paper "Market Microstructure Invariants: Empirical Evidence from Portfolio Transitions" (December 12, 2011). These two papers superseded an older combined manuscript "Market Microstructure Invariants" (May 8, 2011).

This paper proposes and tests two empirical hypotheses that we call "market microstructure invariance." When portfolio managers trade financial assets, they can be modeled as playing trading games in which risks are transferred. Market microstructure invariance begins with the intuition that these risk transfers, which we call "bets," take place in business time. The rate at which business time passes—"market velocity"—is the rate at which new bets arrive into the market. For actively traded assets, business time passes quickly; for inactively-traded assets, business time passes slowly. Market microstructure characteristics—such as bet size, market impact, and bid-spreads—vary across assets and across time. Market microstructure invariance hypothesizes that these microstructure characteristics become constants—"microstructure invariants"—when viewed in business time.

Section 1 formulates two *invariance principles* as empirical hypotheses, conjectured to apply for all assets and across time.

- *Invariance of Bets*: The distribution of the dollar risk transferred by a bet is the same when the dollar risk is measured in units of business time.

- *Invariance of Transactions Costs*: The expected dollar transaction cost of executing a bet is the same function of the size of the bet when the bet's size is measured as the dollar risk it transfers in units of business time.

When business time is converted to calendar time, these invariance hypotheses imply specific empirical restrictions relating market microstructure characteristics to volume and volatility.

The implications of the first invariance hypothesis can be described using the concept of "trading activity," defined as the product of dollar volume and returns volatility. Invariance implies that the number of bets per calendar day is proportional to the two-thirds power of trading activity. Average bet size, expressed as a fraction of trading volume, is inversely proportional to the two-thirds power of trading activity; otherwise, the shape of the distribution of bet size is the same across assets and time.

The implications of the second invariance hypothesis can be described using a measure of illiquidity defined as the cube root of the ratio of returns variance to dollar volume. Invariance implies that the percentage bid-ask spread is proportional to this measure of illiquidity. Percentage transaction costs are proportional to the product of this asset-specific illiquidity measure and some invariant function of bet size, scaled by volume in business time to convert bet size into invariant dollar risk transfer. Invariance does not restrict the shape of this function; it can be consistent with either linear or square-root models of price impact.

Section 2 shows how invariance can be used to impose testable restrictions on transaction-cost models described in the theoretical market microstructure. For example, the model of Kyle (1985) implies that market depth is proportional to the standard deviation of order imbalances. Order imbalances are not directly observable in transactions data. By imposing restrictions on the size and number of bets—which determine the composition of the order flow—invariance shows how to infer the standard deviation of order imbalances from volume and volatility and thereby make correct empirical predictions.

Section 3 describes the portfolio transitions data used to test invariance relationships concerning bet sizes and transaction costs. The dataset consists of more than 400,000 portfolio transition orders executed over the period 2001 through 2005 by a leading vendor of portfolio transition services. In portfolio transitions, institutional fund sponsors hire a third party to execute the orders necessary to transfer funds from legacy portfolio managers to new managers in order to replace fund managers, change asset allocations, or accommodate cash inflows and outflows. Portfolio transitions provide a good natural experiment for identifying bets and measuring transaction costs.

Section 4 examines whether bet sizes are consistent with the invariance-of-bets hypothesis under the identifying assumption that portfolio transition orders are proportional to bets. When scaled as suggested by invariance, the distributions of portfolio transition orders are indeed similar across volume and volatility groups. Regression analysis also confirms this finding.

Moreover, this distribution is well-described by a log-normal with estimated log-variance of 2.53 (figure 1). The bimodal distribution of signed order size (obtained by multiplying the size of sell orders by -1) has much more kurtosis than the normal distribution often assumed for analytical convenience in the theoretical literature. The fat tails of the estimated log-normal distribution suggest that very large bets represent a significant fraction of trading volume and an even more significant fraction of returns variance. Kyle and Obizhaeva (2016*a*) investigate the idea that execution of large stock-market bets may trigger stock market crashes.

Section 5 uses implementation shortfall to examine whether transaction costs are consistent with the invariance hypotheses. Even though our statistical tests usually reject the invariance hypothesis, the results are economically close to those implied by invariance. Consistent with invariance, transaction-cost functions can be closely approximated by the product of an asset-specific illiquidity measure (proportional to the cube root of the ratio of returns variance to dollar volume) and an invariant function of bet size (figure 4). Invariance itself does not impose a particular form on the transaction-cost function. Empirically, both a linear model and a square-root model explain transaction costs well. A square-root model explains transaction costs for orders in the 90th to 99th percentiles better than a linear model; a linear model explains transaction costs for the largest 1% of orders slightly better than the square-root model. Quoted bid-ask spreads are also consistent with the predictions of invariance.

Section 6 calibrates several deep parameters and shows how to extrapolate them to obtain estimates for the distribution of bet size, the number of bets, and transaction-cost functions. Given values of a tiny number of proportionality constants, the invariance relationships allow *microscopic* features of the market for a financial asset, such as number of bets and their size, to be inferred from *macroscopic* market characteristics, such as dollar volume and returns volatility.

The potential benefits of invariance hypotheses for empirical market microstructure are enormous. In the area of transaction-cost measurement, for example, controlled experiments are costly and natural experiments, such as portfolio transitions, are rare; even well-specified tests of transaction-cost models tend to have low statistical power. Market microstructure invariance defines parsimonious structural relationships leading to precise predictions about how various microstructure characteristics, including transaction costs, vary across time and assets with different dollar volume and returns volatility. These predictions can be tested with structural estimates of a handful of parameters, pooling data from many different assets.

Due to market frictions, we do not expect the empirical invariance hypotheses to hold *exactly* across all assets and all times. The predictions of invariance may hold most closely when tick size is small, market makers are competitive, and transaction fees and taxes are minimal. If not, the invariance hypotheses provide a benchmark from which the importance of these frictions can be compared across markets.

This paper focuses on market microstructure invariance as two empirical hypotheses. Kyle and Obizhaeva (2016*b*) develop an equilibrium structural model in which these hypotheses are endogenous implications of a dynamic equilibrium model of informed trading. The model derives invariance relationships under the assumption that the effort required to generate one discrete bet does not vary across assets and time.

The idea of using invariance principles in finance and economics, at least implicitly, is not new. The theory of Modigliani and Miller (1958) is an example of an invariance principle. The idea of measuring trading in financial markets in business time or transaction time is not new either. The time-change literature has a long history, beginning with Mandelbrot and Taylor (1967), who link business time to transactions, and Clark (1973), who links business time to volume. Allais (1956) and Allais (1966) are other early examples of models with time deformation. More recent papers include Hasbrouck (1999), Ané and Geman (2000), Dufour and Engle (2000), Plerou et al. (2000), and Derman (2002). Some of these papers are based on the idea that returns volatility is constant in transaction time. This is different from the invariance hypothesis that the dollar risks transferred by bets have the same probability distribution in bet time.

1 Market Microstructure Invariance as Empirical Hypotheses

Market microstructure characteristics such as order size, order arrival rate, price impact, and bid-ask spread vary across assets and across time. We define "market microstructure invariance" as the empirical hypotheses that these variations almost disappear when these characteristics are examined at an asset-specific "business-time" scale which measures the rate at which risk transfers take place.

Although the discussion below is mostly based on cross-sectional implications of invariance for equity markets for individual stocks, we believe that invariance hypotheses generalize to markets for commodities, bonds, currencies, and aggregate indices such as exchange-traded funds and stock index futures contracts. We also believe that invariance hypotheses generalize to time series. We will thus use subscripts j for assets and t for time periods in what follows.

For simplicity, we assume that a bet transfers only idiosyncratic risk about a single asset, not market risk. Modeling both idiosyncratic and market risks simultaneously takes us beyond the scope of this paper.

Bets and Business Time. In the market for an individual asset, institutional asset managers buy and sell shares to implement *bets*. Our concept of a "bet" is new. We think of a bet as a decision to acquire a long-term position of a specific size, distributed approximately *independently* from other such decisions. Intermediaries with short-term trading strategies—market makers, high frequency traders, and other arbitragers—clear markets by taking the other side of bets placed by long-term traders.

Bets can be difficult for researchers to observe. Bets are neither orders nor trades nor prints; bets are portfolio decisions which implement trading ideas; they are similar to meta-orders. Consider an asset manager who places one bet by purchasing 100,000 shares of IBM stock. The bet might be implemented by placing orders over several days, and each of the orders might be shredded into many small trades. To implement a bet, the trader might place 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 show up in the Trade and Quote (TAQ) database as multiple prints. Since the various individual orders, trades, and prints are positively correlated because they implement a common bet, it would not be appropriate to think of them as independent increments in the intended order flow; they are pieces of bets, not bets. To recover the size of the original bet, all trades which implement the bet must be added together. Thus, individual bets are almost impossible to reconstruct from publicly disseminated records of time-stamped prices and quantities such as those contained in TAQ data.

Bets result from new ideas, which can be shared. If an analyst's 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 issues a buy recommendation to ten different customers and each of the customers quickly places executable orders to buy 10,000 shares, it might be appropriate to think of the ten orders as one bet for 100,000 shares. Since the ten purchases are all based on the same information, the ten individual orders lack statistical independence. Conceptually, it is this independence property of bets that allows us to link their arrival rate to the speed of business time.

To fix ideas, assume that bets arrive randomly. Let γ_{jt} denote the expected arrival rate of bets in asset j at time t; γ_{jt} is measured in bets per calendar day. Suppose that a bet arrives at time t. Let $\tilde{Q}_{jt}$ denote a random variable whose probability distribution represents the signed size of this bet; $\tilde{Q}_{jt}$ is measured in shares (positive for buys, negative for sells) with $E\{\tilde{Q}_{jt}\} = 0$. The expected bet arrival rate γ_{jt} measures market velocity, the rate at which business time passes for a particular asset.[1]

The variables $\tilde{Q}_{jt}$ and γ_{jt} are usually difficult to observe. The invariance hypotheses help to link these variables to volume and volatility, which are easier to observe. To set up this link, it is first useful to make two identifying assumptions. Strictly speaking, these assumptions are not necessary for developing the intuition for invariance. Instead, they help to define issues for future empirical work.

First, let V_{jt} denote trading volume, measured in shares per day. It consists of bet volume reflecting the arrival of bets and intermediation volume reflecting trades of intermediaries. Assume that, on average, each unit of bet volume results in ζ_{jt} units of total volume, implying one unit of bet volume leads to $\zeta_{jt} - 1$ units of intermediation volume. If all trades are bets and there are no intermediaries, then $\zeta_{jt} = 1$, since each unit of trading volume matches a buy-bet with a sell-bet. If a monopolistic specialist intermediates all bets without involvement of other intermediaries, then $\zeta_{jt} = 2$. If each bet is intermediated by different market makers, each of whom lays off inventory by trading with other market makers, then $\zeta_{jt} = 3$. If positions are passed around among multiple intermediaries, then $\zeta_{jt} \geq 4$.

Define expected "bet volume" $\bar{V}_{jt}$ as the share volume from bets, $\bar{V}_{jt} := \gamma_{jt} \cdot E\{|\tilde{Q}_{jt}|\}$. In terms of bet volume $\bar{V}_{jt}$ and the volume multiplier ζ_{jt}, trading volume is equal to $V_{jt} = \zeta_{jt}/2 \cdot \bar{V}_{jt}$, where dividing by two implies that a buy-bet matched to a sell-bet is counted as one unit of volume, not two units. Bet volume $\bar{V}_{jt}$ and trading volume V_{jt} therefore satisfy the relationship

$$\bar{V}_{jt} := \gamma_{jt} \cdot E\{|\tilde{Q}_{jt}|\} = \frac{2}{\zeta_{jt}} \cdot V_{jt}. \tag{1}$$

While bet volume $\bar{V}_{jt}$ is not directly observed, the second equality in equation (1) shows how it can be inferred from trading volume V_{jt} if the *volume multiplier* ζ_{jt} is known. In what follows, we make the identifying assumption, consistent with Occam's razor, that ζ_{jt} is constant across assets and time; thus, for some constant ζ, we assume $\zeta_{jt} = \zeta$ for all j and t.

Second, define returns volatility σ_{jt} as the percentage standard deviation of an asset's daily returns. Some price fluctuations result from the market impact of bets while

[1] Over long periods of time, the inventories of intermediaries are unlikely to grow in an unbounded manner; this requires bets to have small negative autocorrelation. Also, both the bet arrival rate and the distribution of bet size change over longer periods of time as the level of trading activity in an asset increases or decreases.

others result from release of information directly without trading, such as overnight news announcements. Let ψ_{jt}^2 denote the fraction of returns variance σ_{jt}^2 resulting from bet-related order imbalances. Define "bet volatility" $\bar{\sigma}_{jt}$ as the standard deviation of returns resulting from the market impact of bets, not news announcements. Bet volatility $\bar{\sigma}_{jt}$ and returns volatility σ_{jt} satisfy

$$\bar{\sigma}_{jt} = \psi_{jt} \cdot \sigma_{jt}. \tag{2}$$

While bet volatility is not directly observed, it can be inferred from returns volatility σ_{jt} using the *volatility multiplier* ψ_{jt}. Let P_{jt} denote the price of the asset; then dollar bet volatility is $P_{jt} \cdot \bar{\sigma}_{jt} = \psi_{jt} \cdot P_{jt} \cdot \sigma_{jt}$. To simplify the empirical analysis below, we make the identifying assumption that ψ_{jt} is the same across assets and time; thus, for some constant ψ, we assume $\psi_{jt} = \psi$ for all j and t.

To illustrate, if $\zeta = 2$ and $\psi = 0.80$, then for all assets and time periods, bets are intermediated by a monopolist market maker, and bets generate 64 percent of returns variance, whereas the remaining 36 percent of returns variance comes from news announcements.

The assumptions that ζ_{jt} and ψ_{jt} are constants is important for testing the predictions of market microstructure invariance empirically. These assumptions can be tested empirically. If ζ_{jt} and ψ_{jt} are correlated with V_{jt} and σ_{jt}, empirical estimates of parameters predicted by invariance may be biased. An interesting alternative approach, which takes us beyond the scope of this paper, is to examine these correlations empirically and then to make necessary adjustments in tests of our invariance hypotheses.

The assumptions that ζ_{jt} and ψ_{jt} are constants is not important for understanding market microstructure invariance theoretically. To understand invariance theoretically, it suffices to assume $\zeta = 2$ and $\psi = 1$, in which case $\bar{V}_{jt} = V_{jt}$ and $\bar{\sigma}_{jt} = \sigma_{jt}$, so that the distinction between variables with and without *bars* can be ignored.

Invariance of Bets. We call our first invariance hypothesis "invariance of bets." Since business time is linked to the expected arrival of bets γ_{jt}, returns volatility in one unit of business time $1/\gamma_{jt}$ is equal to $\bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$. A bet of dollar size $P_{jt} \cdot \tilde{Q}_{jt}$ generates a standard deviation of dollar mark-to-market gains or losses equal to $P_{jt} \cdot |\tilde{Q}_{jt}| \cdot \bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$ in one unit of business time. The signed standard deviation $P_{jt} \cdot \tilde{Q}_{jt} \cdot \bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$ measures both the direction and the dollar size of the risk transfer resulting from the bet.

The size of the bet can be measured as the dollar amount of risk it transfers per unit of business time, which we denote $\tilde{I}_{jt}$ and define by

$$\tilde{I}_{jt} := P_{jt} \cdot \tilde{Q}_{jt} \cdot \frac{\bar{\sigma}_{jt}}{\gamma_{jt}^{1/2}}. \tag{3}$$

Since the primary function of financial markets is to transfer risks, it is economically more meaningful to measure the size of bets in terms of the dollar risks they transfer rather than the dollar value or number of shares transacted. Indeed, transactions can be large in terms of shares traded but small in terms of dollar amounts transacted, as in markets for low-priced stocks. Transactions can also be large in dollar terms but small in terms of risks transferred, as in the market for U.S. Treasury bills with low returns volatility. Furthermore, as discussed next, we believe that empirical regularities in bet size and transaction costs become more apparent when returns volatility is examined in business time units.

The variable $\tilde{I}_{jt}$ in equation (3) is a good candidate for measuring the economic content of bets because it is immune to splits and changes in leverage. Indeed, a stock split which changes the number of shares does not change the dollar size of a bet $P_{jt} \cdot \tilde{Q}_{jt}$, returns volatility $\bar{\sigma}_{jt}$, or the number of bets γ_{jt} in equation (3). For example, a two-for-one stock split should theoretically double the share volume of bets, but reduce by one-half the dollar value of each share, without affecting dollar size of bets or returns volatility.

Also, a change in leverage does not change dollar volatility $P_{jt} \cdot \bar{\sigma}_{jt}$, contract size of a bet $\tilde{Q}_{jt}$, or the number of bets γ_{jt} in equation (3). For example, if a company levers 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, thus keeping dollar volatility constant. 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.

Hypothesis: Invariance of Bets. *The distribution of the dollar risk transferred by a bet in units of business time is the same across asset j and time t, in the sense that there exists a random variable $\tilde{I}$ such that for any j and t,*

$$\tilde{I}_{jt} \overset{d}{=} \tilde{I}, \tag{4}$$

i.e., the distribution of risk transfers $\tilde{I}_{jt}$ is a market microstructure invariant.

This hypothesis implies that the distribution of bet sizes is such that for any asset j and time t, bets in the same percentile transfer risks of the same size in business time. It does not say that volatility in business time is constant.

Consider the following numerical example. Suppose that a 99th percentile bet in stock A is for \$10 million (e.g., 100,000 shares at \$100 per share) while a 99th percentile bet in stock B is for \$1 million (e.g., 100,000 shares at \$10 per share). The dollar sizes of these bets differ by a factor of 10. Since both bets occupy the same percentile in the bet-size distribution for their respective stocks, the invariance of bets implies that the realized value of $\tilde{I}_{jt}$ is the same in both cases. Even though stock A may be more actively traded than stock B, its returns volatility per unit of business time must be lower by a factor of 10 for the invariance hypothesis to hold. For example, if the stocks have the same volatility in calendar time, say two percent per day, then the arrival rate of bets for stock A must be $100 = 10^2$ times greater than stock B.

We next derive the implications of this invariance hypothesis for observable calendar-time measures of volume and volatility. By analogy with the definition of $\tilde{I}_{jt}$, define "trading activity" W_{jt} as the product of expected dollar trading volume $P_{jt} \cdot V_{jt}$ and calendar returns volatility σ_{jt}:

$$W_{jt} := \sigma_{jt} \cdot P_{jt} \cdot V_{jt}. \tag{5}$$

Trading activity measures the aggregate dollar risk transferred by all bets during one calendar day. Similarly, define "bet activity" $\bar{W}_{jt}$ as the product of dollar bet volume $P_{jt} \cdot \bar{V}_{jt}$ and bet volatility $\bar{\sigma}_{jt}$, i.e., $\bar{W}_{jt} := \bar{\sigma}_{jt} \cdot P_{jt} \cdot \bar{V}_{jt}$.[2] Given values of the volume multiplier ζ_{jt} and the volatility multiplier ψ_{jt}, more-easily-observed trading activity W_{jt} can be converted into less-easily-observed bet activity $\bar{W}_{jt}$ using the relationship $\bar{W}_{jt} = W_{jt} \cdot 2\psi_{jt}/\zeta_{jt}$.

Bet activity $\bar{W}_{jt}$ can be expressed as the product of an invariant constant and a power of unobservable market velocity γ_{jt}:

$$\bar{W}_{jt} = \bar{\sigma}_{jt} \cdot P_{jt} \cdot \gamma_{jt} \cdot E\{|\tilde{Q}_{jt}|\} = \gamma_{jt}^{3/2} \cdot E\{|\tilde{I}_{jt}|\} = \gamma_{jt}^{3/2} \cdot E\{|\tilde{I}|\}. \tag{6}$$

In equations (6), the first equality follows from the definition of $\bar{W}_{jt}$ and equation (1), the second equality follows from equation (3), and the third equality follows from the invariance of bets (4). Invariance of bets therefore makes it possible to infer market velocity γ_{jt} from the level of bet activity $\bar{W}_{jt}$, up to some dollar proportionality constant $E\{|\tilde{I}|\}$, which—according the invariance of bets—does not vary across assets j or times t.

Define $\iota := (E\{|\tilde{I}|\})^{-1/3}$; since $\tilde{I}$ has an invariant probability distribution, ι is a constant. Equations (6) makes it possible to express the unobservable bet arrival rate γ_{jt} and the expected size of bets $E\{|\tilde{Q}_{jt}|\}$ in terms of the observable variables P_{jt}, V_{jt}, and σ_{jt} (and W_{jt}):

$$\gamma_{jt} = \bar{W}_{jt}^{2/3} \cdot \iota^2, \qquad E\{|\tilde{Q}_{jt}|\} = \bar{W}_{jt}^{1/3} \cdot \frac{1}{P_{jt} \cdot \bar{\sigma}_{jt}} \cdot \iota^{-2}. \tag{7}$$

The shape of the entire distribution of bet size $\tilde{Q}_{jt}$ can be obtained by plugging γ_{jt} from equation (7) into equation (3). Traders often measure the size of orders as a fraction of average daily volume. Similarly, expressing bet size $\tilde{Q}_{jt}$ as a fraction of expected bet volume $\bar{V}_{jt}$ yields the following prediction for the distribution of bet sizes:

$$\frac{\tilde{Q}_{jt}}{\bar{V}_{jt}} \stackrel{d}{=} \bar{W}_{jt}^{-2/3} \cdot \tilde{I} \cdot \iota. \tag{8}$$

Equations (7) and (8) summarize the empirical implications of invariance for the distribution of bet size $\tilde{Q}_{jt}$ and the arrival rate of bets γ_{jt}. We test these implications in section 4.

[2] In principle, we could distinguish between $\bar{P}_{jt}$ and P_{jt} based on adjustments for transaction fees, fee rebates, taxes, and tick size effects. To keep matters simple, we ignore these issues and effectively assume $\bar{P}_{jt} = P_{jt}$.

Equation (7) describes the implied composition of order flow. The specific exponents 2/3 and 1/3 are very important. Equation (7) implies that if $\bar{W}_{jt}$ increases by one percent, then the arrival rate of bets γ_{jt} increases by 2/3 of one percent and the distribution of bet size $\tilde{Q}_{jt}$ shifts upwards by 1/3 of one percent. The exponents 1/3 and 2/3 have simple intuition. For example, suppose the expected arrival rate of bets γ_{jt} speeds up by a factor of 4, but volatility in calendar time $\bar{\sigma}_{jt}$ does not change. Then volatility per unit of business time $\bar{\sigma} \cdot \gamma_{jt}^{-1/2}$ decreases by a factor of 2. The invariance hypothesis (4) therefore requires bet size $\tilde{Q}_{jt}$ to increase by a factor of 2 to keep the distribution of $\tilde{I}_{jt}$ invariant. The resulting increase in volume by a factor of $8 = 4^{3/2}$ can be decomposed into an increase in the number of bets by a factor of $8^{2/3} = 4$ and an increase in the size of bets by a factor of $8^{1/3} = 2$.

As bet activity increases, the number of bets increases twice as fast as their size. This specific relationship between the number and size of bets lies at the very heart of invariance. By suggesting a particular composition of order flow, invariance links observable volume to unobservable order imbalances, which in turn has further implications for transaction costs. Indeed, our next hypothesis about transaction costs relies on the order flow having this specific composition.

Invariance of Transaction Costs. We call our second invariance hypothesis "invariance of transaction costs." The risk transferred per unit of business time by a bet of $\tilde{Q}_{jt}$ shares is measured by $\tilde{I}_{jt} = P_{jt} \cdot \tilde{Q}_{jt} \cdot \bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$. Let $C_{B,jt}(\tilde{I}_{jt})$ denote the expected dollar cost of executing this bet.

Hypothesis: Invariance of Transaction Costs*. The dollar expected transaction cost of executing a bet is the same function of the size of the bet when its size is measured as the dollar risk it transfers in units of business time, in the sense that there exists a function $C_B(I)$ such that for any j and t,*

$$C_{B,jt}(I) = C_B(I), \tag{9}$$

i.e., the dollar transaction-cost function $C_{B,jt}(I)$ is a market microstructure invariant.

This hypothesis implies that the dollar cost of executing bets of the same percentile is the same across time and assets. Define the unconditional expected dollar cost $\bar{C}_{B,jt} := E\{C_{B,jt}(\tilde{I}_{jt})\}$. Invariance of transaction costs implies $\bar{C}_{B,jt} := E\{C_B(\tilde{I}_{jt})\}$, and invariance of bets further implies $\bar{C}_{B,jt} = \bar{C}_B$ for some constant $\bar{C}_B$. The hypothesis says that the dollar costs, not the percentage costs, are the same across time or across assets for orders in the same percentiles.

As in the previous example, suppose that a 99th percentile bet in stock A is for \$10 million while a 99th percentile bet in stock B is for \$1 million (e.g., 100,000 shares at \$10 per share). While the dollar sizes of these bets are different, their corresponding measures of dollar risk transfers are the same. Even though the bet in stock A

has 10 times the dollar value of the bet in stock B, invariance of transaction costs implies that the expected cost of executing both bets must the same in dollars because both bets transfer the same amount of risk per stock-specific unit of business time. Traders typically measure transaction costs in basis points, not dollars. Invariance of transaction costs implies that the percentage transaction cost for stock B must be 10 times greater than for stock A.

As discussed next, invariance of transaction costs places strong, empirical testable restrictions on transaction-cost models.

Consider the implications for the percentage costs of executing bets. Let $C_{jt}(Q)$ denote the asset-specific expected cost of executing a bet of Q shares, expressed as a fraction of the notional value of the bet $|P_{jt} \cdot Q|$:

$$C_{jt}(Q) := \frac{C_{B,jt}(I)}{|P_{jt} \cdot Q|}, \qquad \text{where} \qquad I \equiv P_{jt} \cdot Q \cdot \frac{\bar{\sigma}_{jt}}{\gamma_{jt}^{1/2}}. \tag{10}$$

The notations Q and I are two ways to refer to the same bet. The quantity I rescales the bet from share units into dollar units so that bets becomes comparable across assets and time. Using equation (3), this percentage cost function can be expressed as the product of two factors:

$$C_{jt}(Q) = \frac{\bar{C}_{B,jt}}{E\{|P_{jt} \cdot \tilde{Q}_{jt}|\}} \cdot \frac{C_{B,jt}(I)/\bar{C}_{B,jt}}{|I|/E\{|\tilde{I}_{jt}|\}}. \tag{11}$$

We next discuss these two factors separately in more detail.

The first factor on the right side of equation (11), denoted $1/L_{jt}$, is the *asset-specific* liquidity measure defined by

$$\frac{1}{L_{jt}} := \frac{\bar{C}_{B,jt}}{E\{|P_{jt} \cdot \tilde{Q}_{jt}|\}}. \tag{12}$$

It measures the dollar-volume-weighted expected percentage cost of executing a bet. For an asset manager who places many bets in the same asset, this measure intuitively expresses the expected transaction cost as a fraction of the dollar value traded. For example, if an asset manager executes \$10 million at a cost of 20 basis points, \$5 million at a cost of 10 basis points, and \$100 million at a cost of 80 basis points, then the implied approximation for this measure of illiquidity $1/L_{jt}$ is equal to 72 basis points ($= (10 \cdot 20 + 5 \cdot 10 + 100 \cdot 80)/115$).

The second factor on the right side of equation (11), denoted $f(I)$, is the *invariant* average cost function defined as

$$f(I) := \frac{C_{B,jt}(I)/\bar{C}_{B,jt}}{|I|/E\{|\tilde{I}_{jt}|\}} = \frac{C_B(I)/\bar{C}_B}{|I|/E\{|\tilde{I}|\}}. \tag{13}$$

The second equality follows directly from the invariance hypotheses. Thus, the two invariance hypotheses imply that the function $f(I)$, which does not require subscripts

j and t, describes the shape of transaction-cost functions in a manner that does not vary across assets or across time.

Intuitively, the function $f(I)$ is the ratio of $C_B(I)$ to $|I|$ when both are expressed as multiples of the means of $C_B(\tilde{I})$ and $|\tilde{I}|$, respectively. For example, if I_1 denotes a bet that is equal to an average unsigned bet of size $E\{|\tilde{I}|\}$ and its dollar cost is 1.20 times higher than the average dollar cost $\bar{C}_B$, then $f(I_1) = 1.20$. If I_2 denotes a bet that is 5 times greater than an average unsigned bet of size $E\{|\tilde{I}|\}$ and its dollar cost is 10 times greater than the average dollar cost $\bar{C}_B$, then $f(I_2) = 10/5 = 2$.

To summarize, we obtain the following important decomposition of transaction-cost functions:

Theorem: Decomposition of Transaction-Cost Functions. *The percentage transaction cost $C_{jt}(Q)$ of executing a bet of Q shares in asset j at time t is equal to the product of the* asset-specific *illiquidity measure* $1/L_{jt}$ *and an* invariant *transaction-cost function $f(I)$,*

$$C_{jt}(Q) = \frac{1}{L_{jt}} \cdot f(I), \qquad \textit{where} \qquad I \equiv P_{jt} \cdot Q \cdot \frac{\bar{\sigma}_{jt}}{\gamma_{jt}^{1/2}}. \tag{14}$$

The decomposition (14) represents the strong restriction which invariance hypotheses places on transaction-cost models. The percentage transaction-cost function $C_{jt}(Q)$ varies significantly across assets and time. When its argument Q is converted into an equivalent risk-transfer I and its value is scaled with the asset-specific illiquidity index $1/L_{jt}$, then this function turns into an invariant function $f(I)$. As we discuss next, $1/L_{jt}$ is proportional to returns volatility in business time.

An Illiquidity Measure. It can be shown that the asset-specific illiquidity measure $1/L_{jt}$, defined in equation (12), is proportional to bet-induced returns volatility in business time. Furthermore, it can be also conveniently expressed in terms of observable volume and volatility.

Theorem: Illiquidity Index. *The illiquidity index $1/L_{jt}$ for asset j and time t satisfies*

$$\frac{1}{L_{jt}} = \frac{\bar{C}_B}{E\{|\tilde{I}|\}} \cdot \frac{\bar{\sigma}_{jt}}{\gamma_{jt}^{1/2}} = \iota^2 \bar{C}_B \cdot \frac{\bar{\sigma}_{jt}}{\bar{W}_{jt}^{1/3}} = \left(\frac{\zeta \psi^2}{2}\right)^{1/3} \cdot \iota^2 \bar{C}_B \cdot \left(\frac{\sigma_{jt}^2}{P_{jt} \cdot V_{jt}}\right)^{1/3}. \tag{15}$$

The first equality says that $1/L_{jt}$ is proportional to returns volatility in business time $\bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$ with invariant proportionality factor $\bar{C}_B/E\{|\tilde{I}|\}$. The second equality expresses $1/L_{jt}$ in terms of bet activity $\bar{\sigma}_{jt}$ and bet volatility $\bar{W}_{jt}$. The third equality expresses $1/L_{jt}$ as the cube root of the asset-specific ratio of observable returns variance σ_{jt}^2 to observable dollar volume $P_{jt} \cdot V_{jt}$, multiplied by an invariant proportionality factor.

In equation (15), the first equality follows from the definition of I_{jt} in equation (3). The second equality can be proved using equation (7) for γ_{jt}. The third equality follows from the definition of $\bar{W}_{jt}$ in equations (6), from equations (1) and (2), and the identifying assumption $\zeta_{jt} = \zeta$ and $\psi_{jt} = \psi$.

The right side of equation (15) is the asset-specific *illiquidity index* implied by invariance. Since the volume multiplier ζ and the volatility multiplier ψ are assumed not to vary across assets, the quantity $L_{jt} \propto \left[P_{jt} \cdot V_{jt}/\sigma_{jt}^2\right]^{1/3}$ becomes a simple index of liquidity based on observable volume and volatility, with a proportionality constant that does not vary with j and t.

The idea that liquidity is related to dollar volume per unit of returns variance $P_{jt} \cdot V_{jt}/\sigma_{jt}^2$ is intuitive. Traders believe that transaction costs are high in markets with low dollar volume and high volatility.

The cube root is necessary to make the illiquidity measure behave properly when leverage changes. If a stock is levered up by a factor of two, then P_{jt} halves and σ_{jt}^2 increases by a factor of four. Without the cube root, the ratio of returns variance to dollar volume therefore increases by a factor of eight. If dollar risk transfers do not change, then the dollar transaction cost should not change either. This requires percentage transaction costs to double since the dollar size of bets $P_{jt} \cdot Q_{jt}$ halves while the share size Q_{jt} remains the same. Taking a cube root changes the factor of eight to two, so that percentage transaction costs double as required.

As discussed in the next section, the liquidity measure L_{jt} is an intuitive and practical alternative to other measures of liquidity, such as Amihud (2002) and Stambough and Pastor (2003). Its value can easily be calibrated from price and volume data provided by the Center for Research in Security Prices (CRSP).

The liquidity measure L_{jt} in equation (15) is also similar to the definition of "market temperature" $\chi = \bar{\sigma}_{jt} \cdot \gamma_{jt}^{1/2}$ in Derman (2002); substituting for γ_{jt} from equation (7), we obtain $\chi = \iota \cdot [P_{jt} \cdot \bar{V}_{jt}]^{1/3} \cdot [\bar{\sigma}_{jt}]^{4/3} \propto L_{jt} \cdot \sigma_{jt}^2$.

While $1/L_{jt}$ is defined as a measure of trading illiquidity, it may also be a good measure of funding illiquidity as well. A reasonable measure of funding liquidity is the reciprocal of a repo haircut that sufficiently protects a creditor from losses if the creditor sells the collateral due to default by the borrower. Such a haircut should be proportional to the volatility of the asset's return over the horizon during which defaulted collateral would be liquidated. As suggested by invariance, this horizon should be proportional to business time $1/\gamma_{jt}$, making volatility over the liquidation horizon proportional to $\bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$, which as we showed earlier is itself proportional to $1/L_{jt}$. Thus, invariance suggests that both trading liquidity and funding liquidity are proportional to L_{jt}.

Transaction-Cost Models. When bet size is measured as a fraction of bet volume $Q/\bar{V}_{jt}$, the cost function $C_{jt}(Q)$ can be also expressed conveniently in terms of bet activity $\bar{W}_{jt}$ and the invariant constants $\iota := (E\{|\tilde{I}|\})^{-1/3}$ and $\bar{C}_B$ as

$$C_{jt}(Q) = \bar{\sigma}_{jt} \cdot \bar{W}_{jt}^{-1/3} \cdot \iota^2 \bar{C}_B \cdot f\left(\frac{\bar{W}_{jt}^{2/3}}{\iota} \cdot \frac{Q}{\bar{V}_{jt}}\right). \tag{16}$$

This can be proved using equations (14), the definition of $1/L_{jt}$ in equation (12), the invariance of bets (7), the invariance of transaction costs (9), the definition of $\tilde{I}_{jt}$ in equation (3), and the definition of γ_{jt} in equation (7). This is the *transaction-costs model* implied by invariance. We test this specification empirically in section 5.

The two invariance hypotheses do not imply a specific functional form for function $f(I)$ in the transaction-cost model (16). In what follows, we focus on two specific functional forms as benchmarks: linear price-impact costs and square-root price-impact costs. Both are special cases of a more general power function specification for $f(I)$. The linear price-impact function is consistent with price-impact models based on adverse selection, such as Kyle (1985). The square-root price-impact function is consistent with empirical findings in the econophysics literature, such as Gabaix et al. (2006), although these results are based on single trades rather than bets. Some papers, including Almgren et al. (2005), find an exponent closer to 0.60 than to the square-root exponent 0.50. For both functional forms, we also include a proportional bid-ask spread cost component.

For the linear model, express $f(\tilde{I})$ as the sum of a bid-ask spread component and a linear price-impact cost component, $f(\tilde{I}) := (\iota^2 \bar{C}_B)^{-1} \cdot \kappa_0 + (\iota \bar{C}_B)^{-1} \cdot \kappa_I \cdot |\tilde{I}|$, where invariance implies that the bid-ask spread cost parameter κ_0, the market impact cost parameter κ_I, and the constants ι and $\bar{C}_B$ do not vary across assets. Since the specific coefficients $\iota^2 \bar{C}_B$ and $\iota \bar{C}_B$ in the specification for $f(\tilde{I})$ are chosen to cancel out in a nice way, equation (14) implies that the cost function $C_{jt}(Q)$ has the simple form

$$C_{jt}(Q) = \bar{\sigma}_{jt}\left(\kappa_0 \cdot \bar{W}_{jt}^{-1/3} + \kappa_I \cdot \bar{W}_{jt}^{1/3} \cdot \frac{|Q|}{\bar{V}_{jt}}\right). \tag{17}$$

When bet sizes are measured as a fraction of expected bet volume and transaction costs are measured in basis points and further scaled in units of bet volatility $\bar{\sigma}_{jt}$, equation (17) says that bid-ask spread costs are proportional $\bar{W}_{jt}^{-1/3}$ and market impact costs are proportional to $\bar{W}_{jt}^{1/3}$ for a given fraction of bet volume.

For the square-root model, express $f(\tilde{I})$ as the sum of a bid-ask spread component and a square-root function of $|\tilde{I}|$, obtaining $f(\tilde{I}) := (\iota^2 \bar{C}_B)^{-1} \kappa_0 + (\iota^{3/2} \bar{C}_B)^{-1} \kappa_I \cdot |\tilde{I}|^{1/2}$, where invariance implies that κ_0, κ_I, ι, and $\bar{C}_B$ do not vary across assets. The proportional cost function $C_{jt}(Q)$ from (14) is then given by

$$C_{jt}(Q) = \bar{\sigma}_{jt}\left(\kappa_0 \cdot \bar{W}_{jt}^{-1/3} + \kappa_I \cdot \left|\frac{Q}{\bar{V}_{jt}}\right|^{1/2}\right). \tag{18}$$

When transaction costs are measured in units of bet volatility $\bar{\sigma}_{jt}$, bid-ask spread costs remain proportional to $\bar{W}_{jt}^{-1/3}$, but the square-root model implies that the bet activity coefficient $\bar{W}_{jt}^{1/3}$ cancels out of the price-impact term. Indeed, the square root is the only function for which invariance leads to the empirical prediction that impact costs (measured in units of returns volatility) depend only on bet size as a fraction of bet volume $\tilde{Q}_{jt}/\bar{V}_{jt}$ and not on any other asset characteristics. If there are no bid-ask spread costs so that $\kappa_0 = 0$, then the square-root model implies the parsimonious transaction-cost function $C_{jt}(Q) = \bar{\sigma}_{jt} \cdot \kappa_I \cdot [|Q|/\bar{V}_{jt}]^{1/2}$.

Torre (1997) propose a square root model like specification (18) based on empirical regularities observed by Loeb (1983). Practitioners sometimes refer to it as "the Barra model." Grinold and Kahn (1995) use an inventory risk model to derive a square-root price-impact formula. Gabaix et al. (2006) formalize this approach under the assumptions that orders are executed as a constant fraction of volume and liquidity providers have mean-variance utility functions linear in expected wealth and its standard deviation (not variance).

To formalize our predictions about the bid-ask spread, let s_{jt} denote the dollar bid-ask spread. As shown earlier for intercepts in equation (17) or (18), the invariance hypotheses imply that the percentage spread s_{jt}/P_{jt} is proportional to $\bar{\sigma}_{jt} \cdot \bar{W}_{jt}^{-1/3}$:

$$\frac{s_{jt}}{P_{jt}} = \kappa_0 \cdot \bar{\sigma}_{jt} \cdot \bar{W}_{jt}^{-1/3}. \tag{19}$$

For example, holding volatility constant, increasing trading activity by a factor of 8 reduces the percentage bid-ask spread by a factor of 2. Equations (15) and (19) above imply that the percentage bid-ask spread is proportional to both the illiquidity measure $1/L_{jt}$ and to volatility in business time $\bar{\sigma}_{jt} \cdot \gamma_{jt}^{-1/2}$, with invariant constants of proportionality.

Numerical Example. The following numerical example illustrates the invariance hypotheses. Suppose a stock has daily volume of \$40 million and daily returns volatility of 2%. Suppose there are approximately 100 bets per day and the mean size of a bet is \$400,000. A daily volatility of 2% implies a standard deviation of mark-to-market dollar gains and losses equal to \$8,000 per calendar day (2% times \$400,000) for the mean bet. Since business time passes at the rate bets arrive into the market, 100 bets per day implies about 1 bet every 4 minutes; the business clock therefore ticks once every 4 minutes. Over a 4-minute period, the standard deviation of returns is 20 basis points ($200/\sqrt{100}$). Thus, the mean bet has a standard deviation of risk transfer of \$800 per unit of business time.

Invariance implies that the specific number \$800 is constant across stocks and across time. For example, if the arrival rate of bets increases by a factor of 4, then the business clock ticks 4 times faster or once every minute, and the standard deviation of returns per tick on that clock is reduced from 20 basis points to 10 basis points

$(200/\sqrt{400})$, keeping daily volatility constant. For the standard deviation of mark-to-market dollar gains and losses on the mean bet to remain constant at \$800, invariance implies that the dollar size of the mean bet must increase by a factor of 2 from \$400,000 to \$800,000. Thus, holding volatility constant, the bet arrival rate increases by a factor of 4 and the size of bets increases by a factor of 2. This implies that daily volume increases by a factor of 8 from \$40 million to \$320 million. Holding volatility constant, the bet arrival rate increases by a factor proportional the 2/3 power of the factor by which dollar volume changes, and the size of bets increases by the 1/3 power of the factor by which dollar volume changes.

Invariance of transaction costs further says that the dollar cost of executing a bet of a given size percentile is the same across different stocks. The dollar costs of executing the mean bets is therefore the same constant across assets and across time; suppose it is equal to \$2,000. Then, doubling the dollar size of the mean bet from \$400,000 in the first stock to \$800,000 in the second stock decreases the cost of \$2,000, measured in basis points, by a factor of 2 from 50 basis points to 25 basis points. The percentage transaction cost of executing a bet is therefore inversely proportional to the 1/3 power of the factor by which dollar volume changes, holding volatility constant.

Similar arguments are valid for bets in different percentiles of the bet-size distribution. Suppose, for example, that the standard deviation of mark-to-market gains or losses on a 99th percentile bet is \$10,000 per tick of business time. Since the standard deviation of returns per tick of business time is equal to 20 basis points in the first stock and 10 basis points in the second stock, the size of such a 99th percentile bet is equal to \$5 million in the first stock and \$10 million in the second stock; the 99th percentile bets differ in dollar size by a factor of 2. Suppose the dollar costs of executing all 99th percentile bets is equal to \$50,000. This corresponds to a percentage cost of 100 basis points in the first stock and 50 basis points in the second stock. The percentage cost for the second stock is lower by a factor of 2; this difference is inversely proportional to the 1/3 power of the factor by which dollar volume changes, holding volatility constant. Similarly, the percentage bid-ask spread of the second stock will be lower by a factor of 2 than the percentage bid-ask spread of the first stock.

Discussion. Our invariance hypotheses have essential properties which potentially allow them to be extended to more general settings.

First, invariance relationships are consistent with irrelevance of the units in which time is measured. The values of I, $C_{B,jt}(I)$, $f(I)$, and $1/L_{jt}$—and therefore the economic content of the predictions of invariance—remain the same regardless of whether researchers measure γ_{jt}, $\bar{V}_{jt}$, $\bar{\sigma}_{jt}$, and $\bar{W}_{jt}$ using daily weekly, monthly, or annual units of time. This is unlike some other models, such as ARCH and GARCH.

Second, invariance relationships are based on the implicit assumption that bets are executed at an endogenously determined *natural* speed that trades off the benefits of faster execution against higher transaction costs. Invariance does not rule out the possibility that unusually fast execution of a bet would lead to execution costs higher than the costs implied by the functions $C_{B,jt}(I)$ and $f(I)$. For example, it is possible to consider more general invariant cost functions $C_{B,jt}(I, T/\gamma_{jt})$ and $f(I, T/\gamma_{jt})$ that depend not only of the size of bets but also on execution horizons T converted from units of calendar time into units of business time T/γ_{jt}.

Third, the values of invariants $\tilde{I}$ and $C_B(I)$ are measured in dollars. Although not considered in the current paper, invariance relationships can also be applied to an international context in which markets have different currencies or different real exchange rates; they can also be applied across periods of time when the price level is changing significantly. Invariance is consistent with the idea that these nominal values $\tilde{I}$ and $C_B(I)$ should be equal to the nominal cost of financial services calculated from the productivity-adjusted wages of finance professionals in the local currency of the given market during the given time period. Since wages are measured in dollars per day and productivity is measured in bets per day, the ratio of wages to productivity is measured in dollars per bet, exactly the same units as $\tilde{I}$ and $C_B(I)$. Like fundamental constants in physics, dividing the invariants $\tilde{I}$ and $C_B(I)$ by the ratio of wages to productivity makes them dimensionless.

Fourth, it is possible to develop an equilibrium market microstructure model that endogenously generates invariance hypotheses. Several modeling assumptions are essential. Trading volume results from bets placed by traders. Bets induce price volatility, and long-term price impact of bets is linear in its size. The effort required to generate one bet is the same across assets and time. There are no barriers of entry into securities trading. Then, if for some reason profit opportunities increase, more traders enter the market, the market becomes more liquid, prices become more accurate, profits per trader decrease, and traders scale up sizes of their bets in order to break even and to continue covering the costs of generating trading ideas. Trading volume increases both due to an increase in the number of bets and an increase in their sizes. The order flow has a specific "$2/3 - 1/3$" composition because bet sizes must be kept inversely proportional to returns volatility per unit of business time for the profits of traders to remain constant across assets and time. Kyle and Obizhaeva (2016*b*) discuss a model along these lines.

We model market microstructure using invariance in a manner similar to the way modern physicists model turbulence. Kolmogorov (1941*a*) derived his "two-thirds law" (or "five-thirds law") for the energy distribution in a turbulent fluid based on dimensional analysis and scaling.[3] Our analysis is also similar in spirit to inferring the size and number of molecules in a mole of gas from measurable large-scale physical quantities.

[3]We thank an anonymous referee and Sergey N. Smirnov for pointing out the connection to Kolmogorov's model of turbulence.

2 Microstructure Invariance in the Context of the Market Microstructure Literature

Market microstructure invariance builds a bridge from theoretical models of market microstructure to empirical tests of those models. Theoretical microstructure models usually suggest measures of liquidity based on the idea that order imbalances move prices. By scaling business time to be proportional to the rate at which bets arrive, market microstructure invariance imposes cross-sectional or time-series restrictions which make it easier to implement liquidity measures based on order imbalances.

Many theoretical models use game theory to model trading. These models typically make specific assumptions about the risk aversion of traders, 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 sides of trades. Some models emphasize adverse selection, such as Treynor (1995), Kyle (1985), Glosten and Milgrom (1985), and Back and Baruch (2004); some models emphasize inventory dynamics, such Grossman and Miller (1988) and Campbell and Kyle (1993); some models emphasize both, such as Grossman and Stiglitz (1980) and Wang (1993).

While these theoretical models are all based on the idea that order imbalances move prices (with particular parameters depending on specifics of each model), it is difficult to infer precise empirical implications from these models. Theoretical models usually provide neither a unified framework for mapping the theoretical concept of an order imbalance into its empirical measurements nor precise predictions concerning how price impact varies across different assets.

Instead, researchers have taken an approach based on *ad hoc* empirical intuition. For example, price changes can be regressed on imperfect empirical proxies for order imbalances—e.g., the difference between uptick and downtick volume, popularized by Lee and Ready (1991)—to obtain market impact coefficients, which can then be related to stock characteristics such as market capitalization, trading volume, and volatility. Breen, Hodrick and Korajczyk (2002) is an example of this approach. A voluminous empirical literature describes how the rate at which orders arrive in calendar time, the dollar size of orders, the market impact costs, and bid-ask spread costs vary across different assets. For example, Brennan and Subrahmanyam (1998) estimate order size as a function of various stock characteristics. Hasbrouck (2007) and Holden, Jacobsen and Subrahmanyam (2015) provide surveys of this empirical literature.

In contrast to the existing literature, microstructure invariance generates precise empirically testable predictions about how the size of bets, arrival rate of bets, market impact costs, and bid-ask spread costs vary across assets with different levels of trading activity. These predictions are consistent with intuition shared by many models. The unidentified parameters in theoretical models show up as invariant constants (e.g., $E\{|\tilde{I}|\}$ and $\bar{C_B}$), which can be calibrated from data.

In this sense, microstructure invariance is a modeling principle applicable to different models, not a model itself. It compliments theoretical models by making it easier to test them empirically.

Applying Invariance to the Model of Kyle (1985). We use the continuous-time model of Kyle (1985) as an example to discuss how invariance helps to map the theoretical predictions of a model of order imbalances to data on volume and volatility.[4]

In this model, the market depth formula $\lambda = \sigma_V/\sigma_U$ measures market depth (in units of dollars per share-squared) as the ratio of the standard deviation of asset price changes σ_V (measured in dollars per share per square root of time) to the standard deviation in order imbalances σ_U (measured in shares per square root of time). This formula asserts that price fluctuations result from the linear impact of order imbalances. The market depth formula itself does not depend on specific assumptions about interactions among market makers, informed traders, and noise traders. An empirical implementation of the market impact formula $\lambda = \sigma_V/\sigma_U$ should *not* be considered a test of the specific assumptions of the model of Kyle (1985), such as the existence of a monopolistic informed trader who trades smoothly and patiently in a context where less patient liquidity traders trade more aggressively and market makers set asset prices efficiently. Instead, empirical implementation of the formula $\lambda = \sigma_V/\sigma_U$ attempts the more general task of measuring a market impact coefficient λ based on the assumption that price fluctuations result from the linear impact of order-flow innovations, a property shared by many models.

Measuring the numerator σ_V is much more straightforward than measuring the denominator σ_U. The value of σ_V is easily inferred from the asset price and returns volatility. We have $\sigma_V = \sigma_{jt} \cdot P_{jt}$.

Measuring the denominator σ_U is difficult because the connection between observed trading volume and order imbalances is not straightforward. Intuitively, σ_U should be related to trading volume in some way. The continuous-time model provides no help concerning what this relationship is; in the Brownian motion model of Kyle (1985), trading volume is infinite. Without some other approach for measuring σ_U, the model is not testable. We can think of Brownian motion as an approximation to order imbalances resulting from discrete, random, zero-mean decisions by traders to change asset holdings. We call these decisions bets. Since bets are independently distributed, the standard deviation of order imbalances is given by $\sigma_U = \gamma_{jt}^{1/2} \cdot [E\{\tilde{Q}_{jt}^2\}]^{1/2}$. This approach is also consistent with the spirit of other models, such as Glosten and Milgrom (1985) and Back and Baruch (2004).

The formulas for the numerator σ_V and denominator σ_U imply that the price impact of a bet of X shares, expressed as a fraction of the value of a share P_{jt}, is

[4] To simplify this discussion, we omit the distinction between variables with and without a *bar* by assuming $\zeta = 2$, $\psi = 1$, and therefore $\bar{\sigma}_{jt} = \sigma_{jt}$ and $\bar{V}_{jt} = V_{jt}$.

given by

$$\frac{\lambda_{jt}}{P_{jt}^2} \cdot (X \cdot P_{jt}) = \frac{\sigma_V}{\sigma_U} \cdot \frac{X}{P_{jt}} = \sigma_{jt}\gamma_{jt}^{-1/2} \cdot \frac{X}{[E\{\tilde{Q}_{jt}^2\}]^{1/2}}. \tag{20}$$

This formula reflects simple intuition. A one-standard deviation bet $X = [E\{\tilde{Q}_{jt}^2\}]^{1/2}$ has a price impact $\sigma_{jt}\gamma_{jt}^{-1/2}$ equal to one standard deviation of returns volatility σ_{jt} measured over a time interval $1/\gamma_{jt}$ equal to the expected time interval between bet arrivals.

Empirical tests of this formula require assumptions about how $\tilde{Q}_{jt}$ and γ_{jt} vary with volume and volatility so that the standard deviation of order imbalances can be calculated. The invariance of bets provides the required assumptions. Using equations (7) and (8) to determine how γ_{jt} and moments of $\tilde{Q}_{jt}$ vary with observable volume and volatility, it follows that the price-impact cost of an order of dollar size $X \cdot P_{jt}$, as a fraction of the value traded, is

$$\frac{\lambda_{jt}}{P_{jt}^2} \cdot (X \cdot P_{jt}) = \frac{\sigma_{jt}}{P_{jt} \cdot \gamma_{jt}^{1/2} \cdot (E\tilde{Q}_{jt}^2)^{1/2}} \cdot (X \cdot P_{jt}) = \frac{[E\{|\tilde{I}|\}]^{2/3}}{[E\{\tilde{I}^2\}]^{1/2}} \cdot \frac{\sigma_{jt}}{P_{jt} \cdot V_{jt}} \cdot W_{jt}^{1/3} \cdot (X \cdot P_{jt}). \tag{21}$$

The percentage price impact is proportional to $W_{jt}^{1/3} \cdot \sigma_{jt}/(P_{jt} \cdot V_{jt})$, which itself is proportional to the squared illiquidity measure $1/L_{jt}^2$. Invariance of bets makes the proportionality factor $[E\{\tilde{I}^2\}]^{-1/2} \cdot [E\{|\tilde{I}|\}]^{2/3}$ invariant. Thus, implementation of the market impact formula (21) requires calibration of only one proportionality constant for all assets and all time periods. By applying the invariance-of-bets hypothesis to the model of Kyle (1985), we have obtained a linear version of invariance of transactions costs consistent with equation (17) with no bid-ask spread term.

Note that this constant does not depend on the units of time in which variables are measured, because $\tilde{I}$ is measured in units of dollars.

As an alternative to invariance, the formula $\lambda = \sigma_V/\sigma_U$ can be implemented empirically by imposing different assumptions concerning the connection between σ_U and trading volume. For example, we can assume that the expected arrival rate of bets is some unknown constant, the same for all assets and time periods; this will further imply that σ_U is proportional to volume V_{jt} and the illiquidity measure in equation (20) is proportional to $\sigma_{jt}/(P_{jt} \cdot V_{jt})$. Based on this assumption, we obtain

$$\frac{\lambda_{jt}}{P_{jt}^2} \cdot (X \cdot P_{jt}) := \frac{\sigma_{jt}}{P_{jt} \cdot \gamma_{jt}^{1/2} \cdot (E\tilde{Q}_{jt}^2)^{1/2}} \cdot (X \cdot P_{jt}) \propto \frac{\sigma_{jt}}{P_{jt} \cdot V_{jt}} \cdot (X \cdot P_{jt}). \tag{22}$$

This empirical implementation of the transaction-costs formula can be loosely speaking thought of as the illiquidity ratio in Amihud (2002). Indeed, Amihud's illiquidity ratio is the time-series average of the daily ratios of the absolute value of percentage returns to dollar volume. To the extent that dollar volume is relatively stable across time and returns are drawn from the same distribution, this illiquidity ratio is effectively proportional to $\sigma_{jt}/(P_{jt} \cdot V_{jt})$. Although this is a logically consistent way to connect theory with empirical implementation, it is unrealistic to assume that the most actively traded and least actively traded assets have the same number of bets per day; empirical intuition suggests that assets with high levels of trading activity have more bets per day than assets with low levels of trading activity. We are aware of no empirical studies which claim that the number of orders or bets in different assets is the same. Thus, the assumption that the standard deviation of order imbalances is proportional to volume seems to be unrealistic.

The same issue can be addressed by thinking about time units. Unlike our illiquidity measure $1/L_{jt} = \iota^2 \bar{C}_B \cdot [P_{jt} \cdot V_{jt}/\sigma_{jt}^2]^{-1/3}$, the Amihud ratio $\sigma_{jt}/(P_{jt} \cdot V_{jt})$ has time units. Indeed, σ_{jt}^2 and $P_{jt} \cdot V_{jt}$ in our illiquidity measure have the same time units, but σ_{jt} and $P_{jt} \cdot V_{jt}$ in the Amihud ratio do not; the Amihud ratio thus depends on the time horizon over which volume and volatility are measured. Since the left side of equation (22) does not have time units, then to keep the left side consistent with the right side, the proportionality constant in that equation must change when time units are changed. Furthermore, if the invariance-implied market impact formula (21) is indeed correct, then Amihud's market impact formula (22) theoretically implies a different proportionality constant for every stock. This problem can be "fixed"—i.e., the same proportionality coefficient can be obtained for every stock using Amihud's approach—if data for each stock is sampled at a different stock-specific frequency appropriate to the stock's level of trading activity. Invariance implies that the appropriate sampling frequency should be approximately proportional to $1/\gamma_{jt}$, which is proportional to $W_{jt}^{-2/3}$.

Illiquidity ratios calculated using data sampled at the same calendar time frequencies, as implemented in many empirical studies, implicitly rely on the unrealistic assumption that the standard deviation of order imbalances is proportional to trading volume. By contrast, our illiquidity measure $1/L_{jt}$ does not depend on time units, and therefore it does not matter over what time horizons its components are measured. Even if different horizons are used by researchers for different assets, its value will be the same.

3 Data

Portfolio Transitions Data. We test the empirical implications of market microstructure invariance using a proprietary dataset of portfolio transitions from a leading vendor of portfolio transition services.[5] During the evaluation period, this portfolio transition vendor supervised more than 30 percent of outsourced U.S. portfolio transitions. The sample includes 2,552 portfolio transitions executed over the period 2001-2005 for U.S. clients. A portfolio transition may involve orders for hun-

[5]The non-disclosure agreement does not allow revealing the name of the vendor or making the data describing individual customer trades public. Research validating the invariance hypotheses, including research using public data sources, is described at the end of the online Appendix.

dreds of individual stocks. Each order is a stock-transition pair potentially executed over multiple days using a combination of internal crosses, external crosses, and open-market transactions.

The portfolio transitions dataset contains fields identifying the portfolio transition; its starting and ending dates; the stock traded; the trade date; the number of shares traded; a buy or sell indicator; the average execution price; the pre-transition benchmark price (closing price the day before the transition trades began); commissions; SEC fees; and a trading venue indicator distinguishing among internal crossing networks, external crossing networks, open market transactions, and in-kind transfers.

When old legacy and new target portfolios overlap, positions are transferred from the legacy to the new portfolio as "in-kind" transfers. For example, if the legacy portfolio holds $10,000$ shares of IBM stock and the new portfolio holds $4,000$ shares of IBM, then $4,000$ shares are transferred in-kind and the balance of $6,000$ shares is sold. The in-kind transfers do not incur transaction costs and have no effect on our empirical analysis. The $6,000$ shares sold constitute one portfolio transition order, even if the $6,000$ shares are sold over multiple days.

We augment the portfolio transitions data with stock price, returns, and volume data from CRSP. 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 from January 2001 through December 2005 are included in the sample. ADRs, REITs, and closed-end funds are excluded. Also excluded are stocks with missing CRSP information necessary to construct variables used for empirical tests, transition orders in high-priced Berkshire Hathaway class A shares, and transition observations which appeared to contain typographical errors and obvious inaccuracies. Since it is unclear from the data whether adjustments for dividends and stock splits are made in a consistent manner across all transitions, observations with non-zero payouts during the first week following the starting date of portfolio transitions were excluded from statistical tests.

After exclusions, there are 439,765 observations (orders), including 201,401 buy orders and 238,364 sell orders.

CRSP Data: Prices, Volume, and Volatility. For each of the transition-stock observations ($i = 1, \ldots, 439765$), we collect data on the stock's pre-transition price, expected volume, and expected volatility.

The price, denoted P_i, is the closing price for the stock the evening before the first trade is made in any of the stocks in the portfolio transition.

A proxy for expected daily trading volume, denoted V_i (in shares), is the average daily trading volume for the stock in the previous full pre-transition calendar month.

The expected volatility of daily returns, denoted σ_i for order i, is calculated using past daily returns in two different ways.

First, for each stock j and each calendar month m, we estimate the monthly standard deviation of returns $\sigma_{j,m}$ as the square root of the sum of squared daily returns for the full calendar month m (without de-meaning or adjusting for autocorrelation). We define $\sigma_i = \sigma_{j,m}/N_m^{1/2}$, where j corresponds to the stock traded in order i, m is the previous full calendar month preceding order i, and N_m is the number of CRSP trading days in month m.

Second, to reduce effects from the positive skewness of the standard deviation estimates, we estimate for each stock j a third-order moving average process for the changes in $\ln[\sigma_{j,m}]$ for all months m over the entire period 2001-2005. Specifically, letting $\mathbb{L}$ denote the lag operator, we estimate $(1-\mathbb{L})\ln[\sigma_{j,m}] = \Theta_{j,0} + (1 - \Theta_{j,1}\mathbb{L} - \Theta_{j,2}\mathbb{L}^2 - \Theta_{j,3}\mathbb{L}^3)u_{j,m}$. Letting $y_{j,m}$ denote the estimate of $\ln[\sigma_{j,m}]$ and $\hat{V}_j$ the variance of the prediction error, we alternatively define the conditional forecast for the volatility of daily returns by $\sigma_i = \exp(y_{j,m} + \hat{V}_j/2)/N_m^{1/2}$, where m is the *current* full calendar month for order i.

These volatility estimates can be thought of as instrumental variables for true expected volatility. While below we report results using the second definition of σ_i based on the log-ARIMA model, these results remain quantitatively similar when we use the first definition of σ_i based on simple historical volatility during the preceding full calendar month.

Except to the extent that the ARIMA model uses in-sample data to estimate model parameters, we use the pre-transition variables known to the market before portfolio transition trades are executed in order to avoid any spurious effects from using contemporaneous variables.

Descriptive Statistics. Table 1 reports descriptive statistics for traded stocks in panel A and for individual transition orders in panel B. The first column reports statistics for all stocks in aggregate; the remaining ten columns report statistics for stocks in ten dollar-volume groups. Instead of dividing the stocks into ten deciles with the same number of stocks in each decile, 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 of dollar trading volume. Group 10 approximately corresponds to the universe of S&P 100 stocks. The top five groups approximately cover the universe of S&P 500 stocks. Narrower percentile bands 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 descriptive statistics for traded stocks. For the entire sample, the median daily volume is \$18.72 million, ranging from \$1.13 million for the lowest volume group to \$212.85 million for the highest volume group. The median volatility is 1.93 percent per day, ranging from 1.76 percent in the highest-volume decile to 2.16 in the lowest-volume decile. Since there is so much more cross-sectional variation in dollar volume than in volatility across stocks, the variation in trading activity across stocks is related mostly to variation in dollar volume. Trading activity differs by a factor of 150 between stocks in the lowest group and stocks in the highest group, and this variation creates statistical power helpful in determining how transaction costs and order sizes vary with trading activity.

The median quoted bid-ask spread, obtained from the transition dataset, is 12.04 basis points; its mean is 25.42 basis points. From lowest-volume group to highest-volume group, the median spread declines monotonically from 40.96 to 4.83 basis points, by a factor of 8.48. A back-of-the-envelope calculation based on invariance suggests that spreads should decrease approximately by a factor of $150^{1/3} \approx 5.31$ from lowest- to highest-volume group. The difference between 5.31 and 8.48 is partially explained by differences in returns volatility across the volume groups and warrants further investigation. The monotonic decline of almost one order of magnitude is potentially large enough to generate significant statistical power in estimates of a bid-ask spread component of transaction costs based on implementation shortfall.

Panel B of table 1 reports properties of portfolio transition order sizes. The average order size is 4.20% of average daily volume, declining monotonically across the ten volume groups from 16.23% in the smallest group to 0.49% in the largest group, by a factor of 33.12. The median order is 0.57% of average daily volume, also declining monotonically from 3.33% in the smallest group to 0.14% in the largest group, by a factor of 23.79. The invariance hypothesis implies that order sizes should decline by a factor of approximately $150^{2/3} \approx 28.23$, a value which matches the data closely. The medians are much smaller than the means, indicating that distributions of order sizes are skewed to the right. We show below that the distribution of order sizes closely fits a log-normal.

The average trading cost (estimated based on implementation shortfall, as explained below) is 16.79 basis points per order, ranging from 44.95 basis points in the lowest-volume group to 6.16 basis points in the highest-volume group. Invariance suggests that these costs should fall by a factor of $150^{1/3} \approx 5.31$, somewhat smaller than the actual decline. The cost estimates exclude commissions and SEC fees.[6]

One portfolio transition typically contains orders for dozens or hundreds of stocks. It typically takes several days to execute all of the orders. About 60% of orders are executed during the first day of a portfolio transition. Since transition managers often operate under a cash-in-advance constraint—using proceeds from selling stocks in a legacy portfolio to acquire stocks in a target portfolio—sell orders tend to be executed slightly faster than buy orders (1.72 days versus 1.85 days). In terms of dollar volume, about 41%, 23%, 15%, 7%, and 5% of dollar volume is executed on the first day through the fifth days respectively. The two longest transitions in the sample were executed over 18 and 19 business days. The time frame for a portfolio

[6]The SEC fee represents a cost of about 0.29 basis points, which does not vary much across volume groups. The average commission is 7.43 basis points, declining monotonically by a factor of 7.30 from 14.90 basis points for the lowest group to 2.68 basis points for the highest group. Since commissions may be negotiated for the entire transition, the allocation of commission costs to individual stocks is an accounting exercise with little economic meaning.

transition is usually set before its actual implementation begins.

4 Empirical Tests Based on Order Sizes

Market microstructure invariance predicts that the distribution of $\bar{W}_{jt}^{2/3} \cdot \tilde{Q}_{jt}/\bar{V}_{jt}$ does not vary across stocks or time (see equation (8)). We test these predictions using data on portfolio transition orders, making the identifying assumption that portfolio transition orders are proportional to bets.

Portfolio Transitions and Bets. Since bets are statistically independent intended orders, bets can be conceptually difficult for researchers to observe. Consider, for example, a trader who makes a decision on Monday to make one bet to buy 100,000 shares of stock, then implements the bet by purchasing 20,000 shares on Monday and 80,000 shares on Thursday. To an econometrician, this one bet for 100,000 shares may be difficult to distinguish from two bets for 20,000 shares and 80,000 shares respectively. In the context of a portfolio transition, identifying a bet is easier because the size of the order for 100,000 shares is known and recorded on Monday, even if the order is executed over several subsequent days.

Portfolio transition orders may not have a size distribution matching precisely the size distribution of typical bets. Transition orders may be smaller than bets if transitions tend to liquidate a portion of an asset manager's positions or larger than bets if transitions liquidate the sum of bets made by the asset manager in the past. When both target and legacy portfolios hold long positions in the same stock, the portfolio transition order may represent the difference between two bets.

Let X_i denote the unsigned number of shares transacted in portfolio transition order i, $i = 1, \ldots, 439765$. The quantity X_i sums shares traded over multiple days, excluding in-kind transfers.

We make the identifying assumption that, for some constant δ which does not vary across stocks with different characteristics such as volatility and trading activity, the distribution of scaled portfolio transition orders $\delta \cdot X_i$ is the same as the distribution of unsigned bets in the same stock at the same time, denoted $|\tilde{Q}|$. If $\delta = 1$, the distribution of portfolio transition orders matches the distribution of bets. If the scaling constant δ were correlated with volatility or trading activity, parameter estimates might be biased.

The Empirical Hypotheses of Invariance and Log-Normality for the Size Distribution of Portfolio Transition Orders. Let $W_i := V_i \cdot P_i \cdot \sigma_i$ and $\bar{W}_i := \bar{V}_i \cdot P_i \cdot \bar{\sigma}_i$ denote trading activity and bet activity, respectively, for the stock in transition order i. Under the identifying assumption that portfolio transition orders are proportional to bets, invariance of bets implies invariance of portfolio transition orders. Specifically, replacing $\tilde{Q}_{jt}$ with X_i in equation (8) implies that the distribution of $\bar{W}_i^{2/3} \cdot X_i/\bar{V}_i$ does not vary with stock characteristics such as volume, volatility, stock price, or market capitalization.

To facilitate intuitive interpretation of parameter estimates, we scale observations by a hypothetical *benchmark stock* with price P^* of \$40 per share, daily volume V^* of one million shares, and volatility σ^* of 2% per day, implying $W^* = 40 \cdot 10^6 \cdot 0.02$. This benchmark stock would belong to the bottom tercile of S&P 500 (volume group 7 in table 1).

Combining invariance of portfolio transition orders with equations (1) and (2) to convert the bet activity variables $\bar{W}_i$ and $\bar{V}_i$ into trading activity variables W_i and V_i and taking logs, invariance implies the empirically testable relationship

$$\ln\Big[\Big(\frac{W_i}{W^*}\Big)^{2/3} \cdot \frac{X_i}{V_i}\Big] = \ln[\bar{q}] + \tilde{\epsilon}_i. \tag{23}$$

Under the identifying assumptions that the volume multiplier ζ, the volatility multiplier ψ, and the deflator δ do not vary across observations, $\ln[\bar{q}]$ is an invariant constant $\ln[\bar{q}] = E\{\ln[|\tilde{Q}^*|/V^*]\} - \ln[\delta]$ and $\tilde{\epsilon}_i$ is a zero-mean error with the same invariant distribution as $\ln[|\tilde{I}|] - E\{\ln[|\tilde{I}|]\}$.[7] Adjustment by W^* in equation (23) scales each observation on the left side so that it has the same invariant distribution as the log of a hypothetical portfolio transition order in the benchmark stock, expressed as a fraction of its expected daily volume.

We will also examine the stronger log-normality hypothesis—not implied by microstructure invariance—that the distribution of unsigned order sizes adjusted for trading activity $[W_i/W^*]^{2/3} \cdot X_i/V_i$ has a log-normal distribution, i.e., $\tilde{\epsilon}_i$ in equation (23) has a normal distribution. The log-normality hypothesis implies that the right-side of equation (23) is characterized by two invariant constants, the mean $\ln[\bar{q}]$ and the variance of $\tilde{\epsilon}_i$.

Next, we implement several tests to examine this hypothesis.

The Graphical Relationship Between Order Sizes and Trading Activity. One way to examine the invariance hypothesis is to plot the log of order size as a fraction of average daily volume $\ln[X_i/V_i]$ against the log of scaled trading activity $\ln[W_i/W^*]$. Figure 3 presents a cloud of points for all 400,000+ portfolio transition orders. The line $\ln[X_i/V_i] = -5.71 - 2/3 \cdot \ln[W_i/W^*]$ is also shown for comparison. The slope of this line is fixed at $-2/3$, as implied by invariance; the intercept, estimated from an OLS regression, is the sample mean. On the horizontal axis, zero represents the log of trading activity in the benchmark stock; on the vertical axis, zero represents orders for 100% of expected daily volume. The shape of the "super-cloud" conforms well with the invariance hypothesis in that the slope of $-2/3$ is close to the shape of the plotted points and there is only little evidence of heteroscedasticity.

[7]More generally, $\ln[\bar{q}] := E\{\ln[|\tilde{Q}^*|/V^*]\} - 1/3 \cdot \ln[\zeta_i/\zeta^*] - 2/3 \cdot \ln[\psi_i/\psi^*] - \ln[\delta_i]$.

Log-Normal Order Size Distribution for Volume and Volatility Groups. When the portfolio transition orders are sorted into different groups based on characteristics such as dollar volume, volatility, stock price, and turnover, the joint hypotheses of invariance and log-normality imply that the means and variances of $\ln[(W_i/W^*)^{2/3} \cdot X_i/V_i]$ for each group should match the mean and variance of the pooled sample. The pooled sample mean of $\ln[(W_i/W^*)^{2/3} \cdot X_i/V_i]$ is -5.71; the pooled sample variance is 2.53. The pooled sample mean is an estimate of $\ln[\bar{q}]$; the pooled sample variance is an estimate of the variance of the error $\tilde{\epsilon}_i$.[8]

To examine this hypothesis visually, we plot the empirical distributions of the left side of equation (23), $\ln[(W_i/W^*)^{2/3} \cdot X_i/V_i]$, for selected volume and volatility groups. As before, we define ten dollar-volume groups with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of NYSE dollar volume. We define five volatility groups with thresholds corresponding to the 20th, 40th, 60th, and 80th percentiles of returns standard deviation for NYSE stocks. On each plot, we superimpose the bell-shaped density function $N(-5.71, 2.53)$ matching the mean and variance of the pooled sample.

Figure 1 shows plots of the empirical distributions of $\ln[(W_i/W^*)^{2/3} \cdot X_i/V_i]$ for volume groups 1, 4, 7, 9, and 10 and for volatility groups 1, 3, and 5. Consistent with the invariance hypothesis, these fifteen distributions of W-adjusted order sizes are all visually strikingly similar to the superimposed normal distribution. Results for the remaining 35 subgroups also look very similar and therefore are not presented in this paper. The visual similarity of the distributions is reflected in the similarity of their first four moments. For the 15 volume-volatility groups, the means range from -6.03 to -5.41, close to the mean of -5.71 for the pooled sample. The variances range from 2.23 to 2.90, also close to the variance of 2.53 for the pooled sample. The skewness ranges from -0.21 to 0.10, close to skewness of zero for the normal distribution. The kurtosis ranges from 2.73 to 3.38, also close to the kurtosis of 3 for a normal random variable. These results suggest that it is reasonable to assume that unsigned order sizes have a log-normal distribution. Scaling order sizes by $[W_i/W^*]^{2/3}$, as implied by the invariance hypothesis, adjusts the means of the distributions so that they visually appear to be 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. Microstructure invariance does not describe the data perfectly, but it makes a good benchmark from which the modest deviations seen in these plots can be investigated in future research.

Figure 2 further examines log-normality by focusing on the tails of the distributions of portfolio transition orders. For each of the five volume groups 1, 4, 7, 9, and 10, panel A shows quantile-quantile plots of the empirical distribution of

[8] There is not much difference in the distributions of buy and sell orders. For buy orders, the mean is -5.70 and the variance is 2.51; for sell orders, the mean is -5.71 and the variance is 2.55.

$\ln[(W_i/W^*)^{2/3} \cdot X_i/V_i]$ versus a normal distribution with the same mean and variance. The more similar these empirical distributions are to a normal distribution, the closer the plots should be to the 45-degree line. Panel B shows logs of ranks based on scaled order sizes. Under the hypothesis of log-normality, the right tail should be quadratic. A straight line in the right tail would imply a power law. Both panels show that the empirical distributions are similar to a normal distribution, except in the far right and left tails.

In panel A, the smallest orders in the left tails tend to be smaller than implied by a normal distribution. These observations are economically insignificant. Most of them represent one-share transactions in low-price stocks (perhaps the result of coding errors in the data). There are too few such orders to have a meaningful effect on our statistical results.

In panel A, the largest orders in the right tails are much more important economically. On each subplot, a handful of positive outliers (out of 400,000+ observations) do not appear to fit a normal distribution. The largest orders in low-volume stocks appear to be smaller than implied by a normal distribution, and the largest orders in high-volume stocks appear to be larger than implied by a normal distribution.

The finding that the largest orders in low-volume stocks are smaller than implied by a log-normal may be explained by reporting requirements. When an owner acquires more than 5% of the shares of a publicly traded company, the SEC requires information to be reported on Schedule 13D. To avoid reporting requirements, large institutional investors may intentionally acquire fewer shares when intended 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 about 2%, 3%, 4%, 4%, and 4% of shares outstanding, respectively, just below the 5% threshold. The largest order in high-volume stocks is for only about 1% of shares outstanding.

To summarize, we conclude that the distribution of portfolio transition order sizes appears to conform closely to—but not exactly to— the invariance hypothesis. Furthermore, the distribution of order sizes appears to be similar to—but not exactly equal to—a log-normal.

OLS Estimates of Order Size. The order size predictions from equation (23) can also be tested using a simple log-linear OLS regression

$$\ln\left[\frac{X_i}{V_i}\right] = \ln\left[\bar{q}\right] + \alpha_0 \cdot \ln\left[\frac{W_i}{W^*}\right] + \tilde{\epsilon}_i. \tag{24}$$

Invariance of bets implies $\alpha_0 = -2/3$.

To adjust standard errors of OLS estimates of α_0 for positive contemporaneous correlation in transition order sizes across different stocks, the $439,765$ observations are pooled by week over the 2001-2005 period into $4,389$ clusters across 17 industry categories. The double clustering by weeks and industries conservatively adjusts standard errors for large portfolio transitions that may involve hundreds of relatively large orders, executed during the course of a week and potentially concentrated in particular industries.[9]

Table 2 presents estimates for the OLS coefficients in equation (24). 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 parameters are estimated separately for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells.

For the entire sample, the estimate for α_0 is $\hat{\alpha}_0 = -0.62$ with standard error of 0.009. Economically, the point estimate for α_0 is close to the value $-2/3$ predicted by the invariance hypothesis, but the hypothesis $\alpha_0 = -2/3$ is strongly rejected $(F = 25.31, p < 0.0001)$ because the standard error is very small.

When the sample is broken down into NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, it is interesting to note that the estimated coefficients for buy orders, -0.63 for NYSE and -0.71 for NASDAQ, are closer to $-2/3$ than the coefficients for sell orders, -0.59 for both NYSE and NASDAQ. Since portfolio transitions tend to be applied to long-only portfolios, sell orders tend to represent liquidations of past bets. If the size distribution of sell orders depends on past values of volume and volatility—not current values—there is an errors-in-variables problem related to current trading activity being used as a noisy version of past trading activity. This will bias the absolute values of coefficient estimates downwards, consistent with the absolute values of the coefficient estimates for NYSE and NASDAQ sell orders being less than 2/3.

Quantile Estimates of Order Sizes. Table A.1 in the Appendix presents quantile regression results for equation (24) based on the 1st (smallest orders), 5th, 25th, 50th, 75th, 95th, and 99th percentiles (largest orders). The corresponding quantile estimates for α_0 are -0.65, -0.64, -0.61, -0.62, -0.61, -0.64, and -0.63, respectively. Although the hypothesis $\alpha_0 = -2/3$ is rejected due to small standard errors, all quantile estimates are economically close to the value of $-2/3$ predicted by the invariance hypothesis.

Model Calibration and Its Economic Interpretation. Under the invariance-of-bets and log-normality hypotheses, we can calibrate the distribution of bet sizes by imposing the restriction $\alpha_0 = -2/3$ on equation (24). Thus, only the constant term in the regression needs to be estimated.

[9] A potential econometric issue with the log-linear specification in equation (24) is that taking the log of order size as a fraction of average daily volume may create large negative outliers from tiny, economically meaningless orders, with an inordinately large influence on coefficient estimates. Since we have shown above that the shape of the distribution of scaled order sizes closely matches a log-normal, these tiny orders are expected to have only a negligible distorting effect on estimates.

The results of this calibration exercise are presented in table 3. The estimated constant term, -5.71, is the previously reported sample mean of $\ln[\bar{q}]$ in equation (23). The mean-square error, 2.53, is the previously reported sample variance of $\tilde{\epsilon}_i$ in equation (23).

The R^2 (with zero degrees of freedom) is 0.3149; the log of trading activity $\ln[W_i/W^*]$, with the coefficient $\alpha_0 = -2/3$ imposed by invariance, explains a significance percentage of the variation of order size as a fraction of volume X_i/V_i.

When the parameter α_0 is estimated rather than held fixed, changing α_0 from the predicted value of $\alpha_0 = -2/3$ to the estimated value of $\hat{\alpha_0} = -0.62$ increases the R^2 from 0.3149 (table 3) to 0.3167 (table 2), a modest increase of 0.0018. Although statistically significant, the addition of one degree of freedom does not add much explanatory power.

We relax the specification further by allowing the coefficients on the three components of trading activity—volatility σ_i, price P_i, and volume V_i—as well as monthly turnover rate ν_i to vary freely:

$$\ln\left[\frac{X_i}{V_i}\right] = \ln\left[\bar{q}\right] + \alpha_0 \cdot \ln\left[\frac{W_i}{W^*}\right] + b_1 \cdot \ln\left[\frac{\sigma_i}{0.02}\right] + b_2 \cdot \ln\left[\frac{P_i}{40}\right] + b_3 \cdot \ln\left[\frac{V_i}{10^6}\right] + b_4 \cdot \ln\left[\frac{\nu_i}{1/12}\right] + \tilde{\epsilon}_i. \tag{25}$$

This regression imposes on $\ln[W_i/W^*]$ the coefficient $\alpha_0 = -2/3$ predicted by invariance and then allows the coefficients b_1, b_2, b_3, b_4 on the three components of W_i and turnover rate to vary freely. The invariance hypothesis implies $b_1 = b_2 = b_3 = b_4 = 0$. Table 3 reports that increasing the degrees of freedom from one to four increases the R^2 of the regression from 0.3167 to 0.3229, an increase of 0.0062. Although statistically significant, the improvement in R^2 is again modest. Invariance explains much—but not quite all—of the variation in portfolio transition order size across stocks that can be explained by all four variables.

The point estimates for the coefficient on volatility of $\hat{b}_1 = 0.42$, the coefficient on price of $\hat{b}_2 = 0.24$, the coefficient on share volume of $\hat{b}_3 = 0.06$, the coefficient on turnover rate of $\hat{b}_4 = -0.18$ are all statistically significant, with standard errors of 0.040, 0.019, 0.010, and 0.015, respectively (see table A.2 in the Appendix). The coefficients on volatility and price are significantly positive, indicating that order size—as a fraction of average daily volume—does not decrease with increasing volatility and price as fast as predicted by the invariance hypothesis. The statistically significant positive coefficient on volume may be partially offset by a statistically significant negative coefficient on turnover rate.

Discussion. The documented log-normality of bet size is strikingly different from the typical assumptions of microstructure models, where innovations in order flow from noise traders are distributed as a normal, not a log-normal or power law. Although normal random variables are a convenient modeling device—they allow conditional expectations to be linear functions of underlying jointly normally distributed variables—their implications are qualitatively very different.

The estimated log-mean of -5.71 implies that a median portfolio transition order size is equal to 0.33% of expected daily volume for the benchmark stock, since $\exp(-5.71) \approx 0.0033$. The estimated log-variance of 2.53 implies that a one standard deviation increase in order size is a factor of 4.90 for all stocks, since $\exp(2.53^{1/2}) \approx 4.90$.

We next explain why the log-variance of 2.53 also implies that a large fraction of trading volume and an even larger fraction of returns variance come from large bets.

Let $\eta(z)$ and $N(z)$ denote the PDF and CDF, respectively, of a standardized normal distribution. Define $F(\bar{z}, p)$ by $F(\bar{z}, p) := \int_{z=\bar{z}}^{\infty} \exp(p \cdot \sqrt{2.53} \cdot z) \cdot \eta(z) \cdot dz$. It is easy to show that $F(\bar{z}, p) = \exp(p^2 \cdot 2.53/2) \cdot (1 - N(\bar{z} - p \cdot \sqrt{2.53}))$. This implies that the fraction of the pth moment of order size arising from bets greater than $\bar{z}$ standard deviations above the log-mean is given by $F(\bar{z}, p)/F(-\infty, p) = 1 - N(\bar{z} - p \cdot \sqrt{2.53})$.

Plugging $p = 1$, we find that bets larger than $\bar{z}$ standard deviations above the log-mean (median) generate a fraction of total trading volume given by $1 - N(\bar{z} - \sqrt{2.53})$. Bets larger than the 50th percentile generate 94.41% of trading volume ($\bar{z} = 0$). Bets larger than $\sqrt{2.53}$ standard deviations above the log-mean (median) bet size—i.e., the largest 5.39% of bets—generate 50% of trading volume ($\bar{z} = \sqrt{2.53}$).

Plugging $p = 2$, we find that bets larger than $\bar{z}$ standard deviations above the log-mean bet size contribute a fraction of total returns variance given by $1 - N(\bar{z} - 2 \cdot \sqrt{2.53})$ under the assumption that the contribution of bets to price variance is proportional to their squared size (implied by linear price impact). Bets greater than the 50th percentile generate 99.93% of returns variance ($\bar{z} = 0$). Bets larger than $2 \cdot \sqrt{2.53} = 3.18$ standard deviations above the log-mean—i.e., the largest 0.07% of bets—generate 50% of returns variance ($\bar{z} = 2 \cdot \sqrt{2.53}$).

Under the assumptions $\zeta/2 = \psi = \delta = 1$ (stronger than our identifying assumptions), the estimates of mean and variance imply that the benchmark stock has about 85 bets per day for each of the 252 trading days in a calendar year. These estimates then imply that the $1,155$ largest bets out of $21,420$ bets generate approximately half of the trading volume during one year, and the 15 largest bets generate approximately half of returns variance during one year.

Rare large bets may not only account for a significant percentage of returns variance but may also account for some of the stochastic time series variation in volatility. We conjecture that the pattern of short-term volatility associated with execution of rare large bets may depend on the speed with which such bets are executed. As discussed by Kyle and Obizhaeva (2016*a*), large market disturbances such as the stock market crashes of 1929 and 1987, the liquidation of Jerome Kerviel's rogue trades by Société Générale, and the flash crash of May 6, 2010, could have been induced by execution of gigantic bets.

Another implication of log-normality may be a greater kurtosis in the empirical distribution of price changes than a normal distribution would suggest. Given the estimated log-variance of 2.53, it can be shown that the excess kurtosis of one bet has the enormous value of about $\exp(4 \cdot 2.53)$, or approximately 22,000.

Invariance implies a different way of thinking about trading data from that in the time-change literature, which goes back to Mandelbrot and Taylor (1967) and Clark (1973). Mandelbrot and Taylor (1967) begin with the intuition that the distribution of price changes is a stable distribution, i.e., a distribution with the property that a linear combination of two independent random variables has the same shape, up to location and scale parameters. Since it has fatter tails than a normal distribution, it is confined to be a stable Pareto distribution. Following this line of research, the econophysics literature—such as Gopikrishnan et al. (1998), Plerou et al. (2000), and Gabaix et al. (2006)—estimates different power-laws for the probability distributions of different variables and searches for price-formation models consistent with those distributions. Whether order size follows a power law or a log-normal distribution is an interesting question for future research.

Clark (1973) suggests as an alternative hypothesis that the distribution of daily price changes is subordinated to a normal distribution with a time clock linked to a log-normally distributed trading volume. The log-normal distribution is neither stable nor infinitely divisible; the sum of random variables with independent log-normal distributions is not log-normal. Thus, if daily price changes can be described by Clark's hypothesis, neither half-day price changes nor weekly price changes will be described by the same hypothesis.

In some sense, our approach seems to be closer to Mandelbrot and Taylor (1967), who imagine orders of different sizes arriving in the market, with business time linked to their arrival rates rather than to trading volume.

Empirical regularities similar to those implied by invariance can be inferred from the previous literature. Bouchaud, Farmer and Lillo (2009) report, for example, that the number of TAQ prints per day is proportional to market capitalization raised to powers between 0.44 to 0.86. Under the assumption that volatility and turnover rates are stable across stocks as shown in table 1, the midpoint 0.65 of that interval is close to the value of 2/3 implied by invariance for the number of bets per day.

The log-normality of bet size may be related to the log-normality of assets under management for financial firms. Schwarzkopf and Farmer (2010) study the size of U.S. mutual funds and find that its distribution closely conforms to a log-normal with log-variance of about 2.50, similar to our estimates of log-variance for portfolio transition orders. Their annual estimates of log-variance are stable during the twelve years from 1994 to 2005, ranging from 2.43 to 2.59. For years 1991, 1992, and 1993, the log-variance estimates of 1.51, 1.98, and 2.09 are slightly lower, probably because many observations are missing from the CRSP U.S. mutual funds dataset for those years.

As discussed by Aitchison and Brown (1957), log-normal distributions can be found in many areas of natural science. For example, Kolmogorov (1941*b*) proves mathematically that the probability distribution of the sizes of particles under fragmentation converges over time to a log-normal.

5 Empirical Tests Based on Transaction Costs

To examine statistically whether transaction costs conform to the predictions of market microstructure invariance in equation (16), we use the concept of implementation shortfall developed by Perold (1988). Specifically, we estimate costs by comparing the average execution prices of portfolio transition orders with closing prices the evening before any portfolio transition orders begin to be executed. Our tests measure implicit transaction costs resulting from bid-ask spreads and market impact; they exclude explicit transaction costs such as commissions and fees.

Portfolio Transitions and Implementation Shortfall. In portfolio transitions, quantities to be traded are known precisely before trading begins, these quantities are recorded accurately, and all intended quantities are executed. In other trading situations, quantities intended to be traded may not be recorded accurately, and orders may be canceled or quantities may be revised in response to price movements after trading begins. When orders are canceled after prices move in an unfavorable direction or when order size is increased after prices move in a favorable direction, implementation shortfall may dramatically underestimate actual transaction costs. Portfolio transitions data are not subject to these concerns.

Portfolio transition trades are unlikely to be based on short-lived private information about specific stocks because decisions to undertake portfolio transitions and their timing likely result from regularly scheduled meetings of investment committees and boards of plan sponsors, not from fast-breaking private information in the hands of fund managers. Transaction-cost estimates are therefore unlikely to be biased upward as a result of short-lived private information being incorporated into prices while orders are being executed.

These properties of portfolio transitions are not often shared by other data. Consider a dataset built up from trades by a mutual fund, a hedge fund, or a proprietary trading desk at an investment bank. In such samples, the intentions of traders may not be recorded in the dataset. For example, a dataset 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 actual intention to buy another 200,000 shares after the first 100,000 shares are bought. Furthermore, trading intentions may not coincide with realized trades because the trader changes his mind as market conditions change. Indeed, traders often condition their trading strategies on prices by using limit orders or canceling orders, thus hard-wiring into their strategies a selection bias problem for using such data to estimate transaction costs. The dependence of actually traded quantities on prices usually 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 data are particularly well suited for using implementation shortfall to measure transaction costs because portfolio transitions data avoid these sources of statistical bias. These selection-bias issues are discussed in more detail by

Obizhaeva (2012).

The Empirical Hypotheses of Invariance and a Power Function Specification for Transaction Costs. For each unsigned transition order X_i, let $\mathbb{I}_{BS,i}$ denote a buy-sell indicator variable which is equal to $+1$ for buy orders and -1 for sell orders. For transition order i, let C_i denote the *expected* transaction cost as a fraction of the value transacted. Let S_i denote the *actual* implementation shortfall, defined by $S_i = \mathbb{I}_{BS,i} \cdot (P_{ex,i} - P_i)/P_i$, where $P_{ex,i}$ is the average execution price of order i and P_i is a pre-transition benchmark price defined above. Implementation shortfall is positive when orders are unusually costly and negative when orders are unusually cheap.

Invariance imposes the restriction that the unobserved transaction cost C_i has the form given in equation (16):

$$C_i = \bar{\sigma}_i \cdot \bar{W}_i^{-1/3} \cdot \left(\iota^2 \cdot \bar{C}_B \cdot f\Big(\frac{\bar{W}_i^{2/3}}{\iota} \cdot \frac{\mathbb{I}_{BS,i} \cdot X_i}{\bar{V}_i}\Big)\right). \tag{26}$$

Under the stronger hypothesis that the cost function has a power specification for market impact costs, invariance implies the generalization of equations (17) and (18)

$$C_i = \bar{\sigma}_i \cdot \bar{W}_i^{-1/3} \cdot \left(\kappa_0 + \kappa_I \cdot \left(\bar{W}_i^{2/3} \cdot \frac{X_i}{\bar{V}_i}\right)^z\right), \tag{27}$$

where $z = 1$ for the linear specification and $z = 1/2$ for the square-root specification.

Next, we test whether the cost functions can be in fact represented as the product of $\bar{\sigma}_i \cdot \bar{W}_i^{-1/3}$ and an invariant function of $\bar{W}_i^{2/3} \cdot [X_i/\bar{V}_i]$.

The predictions invariance makes about transaction costs can be expressed in terms of a non-linear regression. To justify nonlinear regression estimation, we can think of implementation shortfall as representing the sum of two components: (1) the transaction costs incurred as a result of order execution and (2) the effect of other random price changes between the time the benchmark price is set and the time the trades are executed. If we make the identifying assumption that the implementation shortfall from the portfolio transition dataset is an unbiased estimate of the transaction cost, we can think of modeling the other random price changes as an error in a regression of implementation shortfall on transaction costs.

For example, suppose that while one portfolio transition order is being executed, there are 99 other bets being executed at the same time. The temporary and permanent price impact of executing the portfolio transition order shows up as a transaction cost, while the temporary and permanent price impact of the other 99 unobserved bets being executed shows up as other random price changes. Since the portfolio transition order is one of 100 bets being simultaneously executed, the R^2 of the regression is likely to be about 0.01.

To further develop a non-linear regression framework for testing invariance, we need to make several adjustments.

First, using equation (1) and equation (2), we replace the *bar* variables $\bar{\sigma}_i$, $\bar{V}_i$, and $\bar{W}_i$ with observable variables σ_i, V_i, and W_i and with potentially unobservable constants. We also incorporate the assumption that portfolio transition orders are proportional to bets.

Second, since we want the error in our regression of implementation shortfall on transaction costs be positive when the stock price is moving up and negative when the stock price is moving down, we multiply both the implementation shortfall S_i and the transaction-cost function $C(X_i)$ by the buy-sell indicator. The regression specification can be then written $\mathbb{I}_{BS,i} \cdot S_i = \mathbb{I}_{BS,i} \cdot C_i + \tilde{\epsilon}_i$. Note that $\mathbb{I}_{BS,i} \cdot S_i = (P_{ex,i} - P_i)/P_i$ since $\mathbb{I}^2_{BS,i} = 1$. This gives us a non-linear regression of the form

$$\mathbb{I}_{BS,i} \cdot S_i = \mathbb{I}_{BS,i} \cdot \left[\frac{\psi}{\psi^*}\right]^{2/3} \left[\frac{\zeta}{\zeta^*}\right]^{1/3} \left[\frac{\sigma_i}{\sigma^*}\right] \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot f(I_i \cdot \delta^{-1})/L^* + \tilde{\epsilon}_i, \tag{28}$$

where $I_i := \phi^{-1} \cdot (W_i/W^*)^{2/3} \cdot X_i/V_i$, with invariant constant ϕ obtained from equation (8) and illiquidity measure for the benchmark stock $1/L^* := \iota^2 \cdot \bar{C}_B \cdot \bar{\sigma}^* \cdot [\bar{W}^*]^{-1/3}$ obtained from equation (15).[10] Note that $f(I \cdot \delta^{-1})/L^*$ denotes the invariant cost function for the benchmark stock, expressed as a fraction of notional value, similar to equation (14).

Third, we make another technical adjustment. Since W_i, X_i, and V_i are observable, the quantity $\phi \cdot I_i$ is observable. The quantity I_i itself in equation (28), however, is not observable because the constant ϕ is defined in terms of potentially unobservable constants ι, δ, ψ, and ζ. To estimate the nonlinear regression equation (28), we substitute for $f(.)$ a different function $f^*(.)$ defined by $f^*(x) = (\psi/\psi^*)^{2/3} \cdot (\zeta/\zeta^*)^{1/3} \cdot f(\phi^{-1}\delta^{-1}x)$. Using $x = \phi \cdot I_i$, the right side of equation (28) becomes a simpler expression in terms of observable data, with various potentially unobserved constants incorporated into the definition of f^*, whose functional form is to be estimated from the data. Under the identifying assumptions $\psi = \psi^*$ and $\zeta = \zeta^*$, we have $f^*(\phi I_i) = f(I_i \cdot \delta^{-1})$, where $\phi \cdot I_i := (W_i/W^*)^{2/3} \cdot X_i/V_i$ is observable. The unobserved constants hidden in ϕ affect the economic interpretation of the scaling of the estimated functional form for $f^*(.)$, but they do not otherwise affect the estimation itself. If ζ_i, ψ_i, and δ_i are not constants, but instead are correlated with variables like W_i, this would raise problems of statistical bias in our parameter estimates.

Fourth, the variance of errors in the regression is likely to be proportional in size to the variance of returns and the execution horizon. On average, portfolio transition orders tend to be executed in about one day. To correct for heteroscedasticity resulting from differences in returns volatility, we divide both the right and left sides by returns volatility σ_i/σ^*, where $\sigma^* = 0.02$. Indeed, this adjustment makes the root mean squared error of the resulting regression approximately equal to 0.02.

Fifth, to control for the economically and statistically significant influence that general market movements have on implementation shortfall, we add the CRSP value-weighted market return $R_{mkt,i}$ on the first day of the transition to the right side of

[10]More specifically, $\phi := \delta^{-1}\iota\psi^{-2/3}(\zeta/2)^{-1/3}(W^*)^{-2/3}$.

the regression equation. To the extent that portfolio transition orders are sufficiently large to move the entire U.S. stock market, this adjustment will result in understated transaction costs by measuring only the idiosyncratic component of transaction costs. It is an interesting subject for future research to investigate how large trades in multiple stocks affect general market movements.

Upon making these two changes and using the definition of $f^*(.)$, regression equation (28) becomes

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \left[\frac{W_i}{W^*}\right]^{\alpha} \cdot f^*(\phi I_i)/L^* + \tilde{\epsilon}_i. \tag{29}$$

where invariance implies $\alpha = -1/3$. One of our tests is designed to examine this prediction.

Scaling by W^* makes the function $f^*(\phi I_i)/L^*$ in equation (29) measure the transaction cost for the benchmark stock in terms of the observable value ϕI_i. Although invariance itself does not specify a function form for $f^*(.)/L^*$, the regression places strong cross-sectional restrictions on the shape of the transaction cost function. In addition to the restriction $\alpha = -1/3$, it requires that the same function $f^*(\phi I_i)/L^*$ with $\phi I_i = (W_i/W^*)^{2/3} \cdot X_i/V_i$ for order i be used for all stocks. We test this prediction as well.

We do not undertake separate estimates of transaction-cost parameters for internal crosses, external crosses, and open market transactions. Such estimates would be difficult to interpret due to selection bias resulting from transition managers optimally choosing trading venues to minimize costs.

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

Dummy Variable Regression. In our first test, we fix $\alpha = -1/3$ in transaction-costs regression equation (29), estimate function $f^*(.)/L^*$ using dummy variables, and examine calibrated functions across ten volume groups. Invariance predicts those functions to be similar. The test does not put restrictions on the specific functional forms of $f^*(.)$.

We sort all $439,765$ orders into 100 order size bins of equal size based on the value of the invariant order size $\phi \cdot I_i = [W_i/W^*]^{2/3} \cdot [X_i/V_i]$. As before, we also place each order into one of ten volume groups based on average dollar trading volume in the underlying stock $P_i \cdot V_i$, with thresholds corresponding to the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of NYSE dollar volume. As shown in section 4, the distribution of ϕI_i is approximately invariant across volume groups; specifically, across all volume groups $k = 1, \ldots, 10$, each bin h has a similar number of observations and similar magnitudes for ϕI_i.

In the regression equation (29), we replace the function $f^*(\phi I_i)/L^*$ with $1,000$ dummy variables $D_i^*(k,h)$, $k = 1, \ldots, 10$ and $h = 1, \ldots, 100$, where $D_i^*(k,h) = 1$

if bet i belongs to volume group k based on dollar volume $P_i \cdot V_i$ and to order size bin h based on ϕI_i; otherwise $D_i^*(k,h) = 0$. We then estimate $1,000$ coefficients $f^*(k,h)/L^*$, $k = 1,\ldots,10$, $h = 1,\ldots,100$ for the dummy variables using a separate OLS regression for each of the volume groups, $k = 1,\ldots,10$,

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot \sum_{h=1}^{100} D_i^*(k,h) \cdot f^*(k,h)/L^* + \tilde{\epsilon}_i. \tag{30}$$

For each volume group k, the 100 dummy variable coefficients $f^*(k,h)/L^*$ (where $h = 1,\ldots,100$) track the shape of function $f^*(.)/L^*$, without imposing any particular restrictions on its functional form. Invariance predicts that the ten values of the coefficients $f^*(k,h)/L^*$, $k = 1,\ldots,10$ should be the same for each order size bin h, $h = 1,\ldots,100$. In other words, $f^*(k,h)/L^*$ is predicted not to depend on volume-group index k.

Figure 4 shows ten plots, one for each of the ten volume groups, with the 100 estimated coefficients for the dummy variables plotted as solid dots in each plot. On each plot, we also superimpose the 95% confidence intervals for 100 dummy variable coefficients estimated based on the pooled sample (dotted lines). The superimposed confidence bands help to assess the degree of similarity between cost functions estimated separately based on observations in each volume bin.

On each of the ten plots, the horizontal and vertical axes are scaled in the same way to facilitate comparison. On the horizontal axis, we plot the value for order-size bin h equal to the log of the average ϕI_i for observations in that size bin and corresponding volume group k.

On the right vertical axis, we plot the values of the dummy variable coefficients $f^*(k,h)/L^*$ quantifying for the benchmark stock the cost function as a fraction of notional value, scaled in basis points. To make deviations of cost patterns from invariance visually obvious, we have effectively scaled cost functions as suggested by invariance using regression (30): We multiply orders sizes X_i/V_i by $(W_i/W^*)^{2/3}$ and divide implementation shortfalls S_i by $L^*/L_i = (\sigma_i/\sigma^*) \cdot (W_i/W^*)^{-1/3}$. Here $1/L^* := \iota^2 \cdot \bar{C}_B \cdot \bar{\sigma}^* \cdot [\bar{W}^*]^{-1/3}$ is the illiquidity measure for the benchmark stock from equation (28). The invariance hypotheses imply that the 100 points plotted for each of the 10 volume groups will describe the same underlying cost function when the vertical axis is scaled according to invariance.

On the left vertical axis, we plot actual average transaction cost $f^*(k,h)/L^k$ as a fraction of notional value, scaled in basis points. For each volume group k, this scaling reverses invariance-based scaling by multiplying estimated coefficients $f^*(k,h)/L^*$ by L^*/L^k, where $1/L^k$ is the illiquidity measure for orders in volume group k given by $1/L^k := \iota^2 \cdot \bar{C}_B \cdot \bar{\sigma}^k_{med} \cdot (\bar{W^k}_{med})^{-1/3}$, with $\bar{\sigma}^k_{med}$ denoting median bet volatility and $\bar{W}^k_{med}$ denoting median bet activity for volume group k.

Without appropriate scaling, the data do not reveal their invariant properties. The actual costs on the left vertical axes vary significantly across volume groups. In the low volume group, costs range from −220 basis points to 366 basis points; in the high-volume group, costs range from −33 basis point to 55 basis points, 7 times less than in the low-volume group.

After applying invariance scaling, however, our plots appear to be visually consistent with the invariance hypotheses. For all ten subplots in figure 4, the estimated dummy variable coefficients on the right vertical axes are very similar across volume groups. They also line up along the superimposed confidence band.

Moving from low-volume groups to high-volume groups, these estimates also become visually more noisy. For low-volume group 1, dummy variable estimates lie within the confidence band, very tightly pining down the estimated shape for the function $f^*(.)/L^*$. For high-volume group 10, many dummy variable estimates lie outside of the confidence band, with 11 observations above the band and about 40 observations below the band. These patterns suggest that the statistical power of our tests concerning transaction costs comes mostly from low-volume groups.

Invariance suggests that orders might be executed over horizons inversely proportional to the speed of business time, implying very slow executions for large orders in stocks with low trading activity. Portfolio transitions are, however, usually implemented within a clearly defined tight calendar time frame, which has the effect of speeding up the natural execution horizon for stocks with low trading activity. When transition orders are executed over a fixed number of calendar days, the execution in business time is effectively faster for low-volume stocks and slower for high-volume stocks. When a transition order in a low-volume stock is being executed, there are therefore probably fewer other bets being executed at the same time; this makes the R^2 of the regression higher. Over the same period of calendar time, more bets are being executed for the high-volume stocks, making the R^2 lower than for low-volume stocks. The more patient business-time pace of execution for high-volume stocks may explain why the dummy variable estimates are noisier for high-volume stocks than for low-volume stocks. This may also explain why the execution costs of high-volume stocks appear to be slightly less expensive than low-volume stocks.

Transaction Cost Estimates in Non-Linear Regression. Next, we test the hypothesis $\alpha = -1/3$ in transaction-cost regression equation (29) while simultaneously calibrating a specific functional form for the cost function $f^*(.)/L^*$. We assume that this function has a particular parametric functional form equal to the sum of a constant bid-ask spread term and a market impact term which is a power of $\phi \cdot I$, similar to equation (27). For this particular specification, the non-linear regression (29) can be written as

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \kappa_0^* \cdot \left[\frac{W_i}{W^*}\right]^{\alpha_1} + \mathbb{I}_{BS,i} \cdot \kappa_I^* \cdot \left[\frac{W_i}{W^*}\right]^{\alpha_2} \cdot \left[\frac{\phi I_i}{0.01}\right]^{z} + \tilde{\epsilon}_i, \tag{31}$$

where $\phi I_i/0.01 = [W_i/W^*]^{2/3} \cdot X_i/[0.01 V_i]$. The explanatory variables are scaled so that, for the benchmark stock, execution of one percent of daily volume has price-impact cost of κ_I^* and fixed bid-ask spread of κ_0^*, both measured as a fraction of the value traded. Equation (31) nests for empirical testing both the linear model of equation (17) ($z = 1$) and the square-root model of equation (18) ($z = 1/2$).

First, we report estimates of the six parameters (β_{mkt}, z, α_1, κ_0^*, α_2, and κ_I^*) in equation (31) using non-linear regression. Second, we calibrate the three-parameter linear impact model of equation (17) with parameters (β_{mkt}, κ_0^*, κ_I^*) by imposing the additional invariance restrictions $\alpha_1 = \alpha_2 = -1/3$ and the linear cost restriction $z = 1$. Third, we also calibrate the three-parameter square-root model of equation (18) with parameters (β_{mkt}, κ_0^*, κ_I^*) by imposing the alternate restriction $z = 1/2$. Finally, we examine a twelve-parameter generalization of equation (31) which replaces powers α_1 and α_2 of trading activity W_i with powers of volatility σ_i, price P_i, volume V_i, and monthly turnover ν_i. Although statistical tests reject invariance, the results indicate that the predictions of invariance are economically significant, with the square-root version of invariance explaining transaction costs better than the linear version.

The parameter estimates for the six parameters β_{mkt}, κ_0^*, z, α_1, κ_I^*, α_2 in the non-linear regression (31) are reported in table 4.

For the coefficient β_{mkt}, which multiplies the market return $R_{mkt,i}$, the estimate is $\hat{\beta}_{mkt} = 0.65$ with standard error 0.013. The fact that $\hat{\beta}_{mkt} < 1$ suggests that many transition orders are executed early on the first day.[11]

The point estimate of the estimated bid-ask spread exponent is $\hat{\alpha}_1 = -0.49$, with standard error 0.050, three standard errors lower than the predicted value $\alpha_1 = -1/3$. In comparison with invariance, this result implies higher spread costs for less actively traded stocks and lower spread costs for more actively traded stocks. Note, however, that the factor involving α_1 is multiplied by $2 \cdot \kappa_0^*$, and κ_0^* is only of marginal statistical significance since it differs from zero by about two standard errors; excluding the bid-ask spread component of prices reduces the R^2 from 0.1010 to 0.1006 (not reported in table). This result may have something to do with the minimum tick size of one cent being a binding constraint for some stocks.

The point estimate for α_2 is $\hat{\alpha}_2 = -0.32$ with standard error 0.015. Since invariance implies $\alpha_2 = -1/3$, this result strongly supports invariance. When the four parameters are estimated separately for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, the estimated coefficients for α_2 are -0.40, -0.33, -0.41, and -0.29, respectively.

The estimate for the market impact curvature parameter z is $\hat{z} = 0.57$ with standard error 0.039. This suggests that a square-root specification ($z = 1/2$) may describe observed transaction costs better than a linear specification ($z = 1$). Note that invariance does not place any restrictions on the parameter z itself.

The point estimate of the market impact coefficient κ_I^* is $\hat{\kappa}_I^* = 10.69 \cdot 10^{-4}$ with standard error $1.376 \cdot 10^{-4}$. The estimates of κ_I^* are higher for buy orders than for sell orders ($12.08 \cdot 10^{-4}$ versus $9.56 \cdot 10^{-4}$ for NYSE; $12.33 \cdot 10^{-4}$ versus $9.34 \cdot 10^{-4}$ for

[11] The fact that $\hat{\beta}_{mkt} = 0.65$ is close to 2/3 is a coincidence; it is not implied by invariance.

NASDAQ).

Note that the point estimate of the market impact coefficient κ_I^* is different from zero with much greater statistical significance than the bid-ask spread coefficient κ_0^*; this is consistent with the interpretation that market impact costs are of considerably greater economic importance than bid-ask spread costs.

Results for the five-parameter linear specification, regression equation (31) with the parameter restriction $z = 1$ (linear impact), are reported in table A.3 in the Appendix. The estimate of the bid-ask spread cost κ_0^* is $6.28 \cdot 10^{-4}$ with standard error $0.890 \cdot 10^{-4}$, and the estimate of the exponent α_1 is $\hat{\alpha}_1 = -0.39$ with standard error 0.020. The estimate of the market impact cost κ_I^* is $2.73 \cdot 10^{-4}$ with standard error 0.252. The estimate of the exponent α_2 is $\hat{\alpha}_2 = -0.31$ with standard error of 0.028; thus, under the restriction of linear price impact ($z = 1$), the additional restriction imposed by invariance ($\alpha_2 = 1/3$) is not statistically rejected.

Model Calibration. Next, we calibrate transaction-cost models under the assumption of invariance and the assumption of either linear or square-root specification for the cost function.

Table 5 presents estimates for the three parameters β_{mkt}, κ_0^*, and κ_I^* in equation (31), imposing the invariance restrictions $\alpha_1 = \alpha_2 = -1/3$ and also imposing either a linear transaction-cost model $z = 1$ or a square-root model $z = 1/2$.

In the linear specification with $z = 1$, the point estimate for market impact cost $\hat{\kappa}_I^*$ is equal to $2.50 \cdot 10^{-4}$, and the point estimate for bid-ask spread cost $\hat{\kappa}_0^*$ is equal $8.21 \cdot 10^{-4}$.

In the square-root specification with $z = 1/2$, the point estimate for market impact cost $\hat{\kappa}_I^*$ is equal to $12.08 \cdot 10^{-4}$, and the point estimate for half bid-ask spread $\hat{\kappa}_0^*$ is equal to $2.08 \cdot 10^{-4}$.

For the benchmark stock, these estimates imply that the total cost of a hypothetical trade of one percent of daily volume incurs a cost of about 10.71 basis points in the linear model and 14.16 basis points in the square-root model.

The benchmark stock would belong to volume group 7, and the corresponding average quoted spread in table 1 for that group is 12.04 basis points. The implied spread estimate of about 16.42 basis points for the linear model is close to the quoted spread; the implied spread estimate of 4.16 basis points for the square-root model may be biased downwards due to collinearity between the constant term and the square-root term in the regression in the region close to zero.

Economic Interpretation. We examine the economic significance of our results by comparing the R^2 of different specifications for transaction-cost models.

The R^2 is equal to 0.0847 in the transaction-cost regressions with market return only (not reported). This implies that a substantial part of realized transaction costs is explained by overall market dynamics. The transaction-cost models improve the R^2s by only one or two percent.

A comparison of the R^2s in table 4 and table 5 provides strong support for the invariance hypothesis. When the coefficient on W_i/W^* is fixed at the invariance-implied value of $-1/3$ and only two transaction-cost parameters κ_I^* and κ_0^* are estimated (table 5), the R^2 is 0.0991 for a linear specification and 0.1007 for a square-root specification. The square-root specification performs better than the linear specification. Compared with the square-root specification, adding the three additional parameters α_1, α_2 and z modestly increases the R^2 from 0.1007 to 0.1010 (table 5). The modest increase strongly supports the economic importance of invariance.

We also consider a more general specification with eleven estimated coefficients. The exponents on the three components of trading activity W_i (volatility σ_i, price P_i, volume V_i) as well as the exponent on the monthly turnover ν_i are allowed to vary freely. The estimated regression equation is

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} \mathbb{I}_{BS,i} \cdot \kappa_0^* \cdot \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot \frac{\sigma_i^{\beta_1} \cdot P_i^{\beta_2} \cdot V_i^{\beta_3} \cdot \nu_i^{\beta_4}}{(0.02)(40)(10^6)(1/12)} + \\ + \mathbb{I}_{BS,i} \cdot \kappa_I^* \cdot \left[\frac{\phi I_i}{0.01}\right]^z \cdot \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot \frac{\sigma_i^{\beta_5} \cdot P_i^{\beta_6} \cdot V_i^{\beta_7} \cdot \nu_i^{\beta_8}}{(0.02)(40)(10^6)(1/12)} + \tilde{\epsilon}_i. \tag{32}$$

Because the exponents on the W-terms are set to be $-1/3$, the invariance hypotheses predict $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$.

Table 5 shows that despite increasing the number of estimated parameters from four to eleven, the R^2 in the aggregate regression increases from 0.1010 to only 0.1016. The estimates of β_1, β_2, β_3, β_4, β_5, β_6, β_7, and β_8 are shown in table A.4 in the Appendix. The estimates of β_1, β_2, β_3, β_4 are often statistically significant, but these explanatory variables are multiplied by statistically insignificant coefficient κ_0^*. Almost all estimates of β_5, β_6, β_7, and β_8 are statistically insignificant, both for the pooled sample as well as the four sub-samples.

In all three specifications, separate regressions for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells suggest that price-impact costs are higher for buy orders than for sell orders. This is consistent with the hypothesis, discussed Obizhaeva (2009), that the market believes that buy orders—in particular, buy orders in portfolio transitions—contain more information than sell orders.

OLS Estimates for Quoted Spread. Finally, we present results of statistical tests based on the data on quoted bid-ask spread for portfolio transition orders.

Since invariance implies that bid-ask spread costs are proportional to $\bar{\sigma}_i \cdot \bar{W}_i^{-1/3}$, intuition suggests that quoted spreads may also have this invariant property. As a supplement to our empirical results on transaction costs, we test this prediction using data on quoted spreads, supplied in the portfolio transition data as pre-trade information for each transition order.

Let s_i denote the dollar quoted spread for order i. Using equations (1) and (2), we can write equation (19) as the log-linear OLS regression

$$\ln\left[\frac{s_i}{P_i \cdot \sigma_i}\right] = \ln \bar{s} + \alpha_3 \cdot \ln\left[\frac{W_i}{W^*}\right] + \tilde{\epsilon}_i, \tag{33}$$

where invariance implies $\alpha_3 = -1/3$. The constant term $\ln[\bar{s}] := \ln[s^*/(40 \cdot 0.02)] + 2/3 \cdot \ln[\psi/\psi^*] - 1/3 \cdot \ln[\zeta^*/\zeta]$ quantifies the dollar spread s^* for the benchmark stock as a fraction of dollar volatility $P^* \cdot \sigma^*$, under the identifying assumptions $\zeta = \zeta^*$ and $\psi = \psi^*$.

Table 6 presents the regression results. The point estimate $\hat{\alpha}_3 = -0.35$ has standard error 0.003. For sub-samples of NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, the estimates are -0.31, -0.32, -0.40, and -0.39, respectively. Although the hypothesis $\alpha_3 = -1/3$ is usually rejected statistically, the estimates are economically close to the value of $-1/3$ predicted by invariance. The point estimate of $\ln[\bar{s}]$ is equal to -3.07, implying a quoted spread of $\exp(-3.07) \cdot 0.02 \approx 9 \cdot 10^{-4}$ for the benchmark stock. This number is similar to the median spread of 8.12 basis points for volume group 7 in table 1.

It can be shown that an implicit spread proportional to $\bar{\sigma}_i \cdot \bar{W}_i^{-1/3}$, as implied by invariance, provides a better proxy for the actually incurred spread costs than the quoted spread itself. When regression equation (31) is estimated with linear impact $z = 1$, using only the 436,649 observations for which quoted bid-ask spread data is supplied, we find the R^2 is equal to 0.0992. Now replace the invariance-implied spread cost proportional to $\bar{\sigma}_i \cdot \bar{W}_i^{-1/3}$ with the quoted half spread $1/2 \cdot s_i/P_i$ in equation (17). The estimated equation is

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \beta_S \cdot \frac{1}{2} \cdot \frac{s_i}{P_i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \kappa_I \cdot \left[\frac{\phi I_i}{0.01}\right] \cdot \left[\frac{W_i}{W^*}\right]^{\alpha_2} + \tilde{\epsilon}_i. \tag{34}$$

We find that the R^2 drops from 0.0992 to 0.0976 (table A.5 in the Appendix). The point estimate of the coefficient on the quoted half-spread coefficient is $\hat{\beta}_S = 0.71$. The estimates are equal to 0.61, 0.74, 0.61, and 0.75, when estimated for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells, respectively.

One interpretation of the estimate of 0.71 is that transition managers incur as a transaction cost only 71% of the quoted half-spread. The values are consistent with the intuition in Goettler, Parlour and Rajan (2005) that endogenously optimizing traders capture a fraction of the bid-ask spread by mixing between market orders and limit orders. Another interpretation is that noise in the quoted spread biases the coefficient towards zero and reduces the explanatory power of the regression.

Bouchaud, Farmer and Lillo (2009) and Dufour and Engle (2000) report that the quoted bid-ask spread is proportional to the standard deviation of percentage returns between trades; this result is implied by microstructure invariance under the assumption that the rate at which trades occur is proportional to the rate at which bets arrive. Stoll (1978*a*) proposes a theory that the percentage dealer bid-ask spread in NASDAQ stocks is proportional to variables including the dealer holding period and the returns variance of the stock; this captures the spirit of invariance if the dealer holding period is proportional to the rate at which bets arrive. Stoll (1978*b*) also tests this theory using data on dealer spreads for NASDAQ stocks, and his estimates are consistent with our findings as well.[12]

Discussion. Figure 5 plots the estimated coefficients for the 100 dummy variables, along with their 95% confidence intervals, estimated from the dummy variable regression equation (30) by pooling the data across all 10 volume groups. The linear and square-root cost functions with parameters calibrated in table 5 are superimposed. The linear specification is $2.50 \cdot 10^{-4} \cdot \phi I/0.01 + 8.21 \cdot 10^{-4}$ (solid black line), and the square-root specification is $12.07 \cdot 10^{-4} \cdot \sqrt{\phi I/0.01} + 2.08 \cdot 10^{-4}$ (solid grey line). Both specifications result in estimates economically close to each other.

Consistent with the higher reported R^2 for the square-root model than the linear model in table 5, the square-root specification fits the data slightly better than the linear specification, particularly for large orders in the order size bins from 90th to 99th percentiles. Consistent with our results, most studies find that total price impact is best described by a concave function.[13] For example, Almgren et al. (2005) obtain an estimate $\hat{z} = 0.60$ for their sample of almost 30,000 U.S. stock orders executed by Citigroup between 2001 and 2003; this is comparable to our estimate of $\hat{z} = 0.56$ when the constraint $\alpha_1 = \alpha_2 = -1/3$ is imposed in regression equation (31). To differentiate temporary impact from permanent impact of earlier executed trades, Almgren et al. (2005) assume a particular execution algorithm with a constant rate of trading. We do not quantify these cost components separately but rather focus on total costs.

Intuition might suggest that for gigantic orders, the square-root model would predict dramatically lower transaction costs than the linear model, making it easy to distinguish the predictions of one model from the other. As the superimposed estimated linear and square-root cost functions for the ten plots in figure 4 make clear, both specifications estimate similar transaction costs for the bin representing the largest 1% or orders (because the graphs of the linear and square-root functions in figure 5 cross near the bin representing the largest 1% of orders). Furthermore, the transaction-cost dummy variable for the largest 1% of orders fits both the linear and

[12] Stoll (1978*b*) reports an R^2 of approximately 0.82 in an OLS regression of percentage bid-ask spread on the logs of various variables including dollar volume, stock price, returns variance, turnover, and number of dealers. Using the standard deviations and correlation matrix for the variables (p. 1165), it can be shown that imposing coefficients of $-1/3$ on dollar volume and $+1/3$ on returns variance (to mimic the definition of $1/L_i$), while imposing coefficients of zero on all other explanatory variables except a constant term, results in an R^2 equal to 0.66. This result is similar to our results in table 6.

[13] Since we plot the log of order size on the horizontal axis but do not take the log of the transaction cost on the vertical axis (to make standard errors have similar magnitudes for different observations), both the linear model and the concave square-root model show up as exponential functions; the graph of the linear model is more convex than the graph of the square-root model.

square-root models well. For the largest 1% of orders in the highest-volume group in figure 4, the estimated dummy variable fits the higher cost estimates of the linear model better than the square-root model.

We have developed notation which allows the possibility that the parameters ζ_{jt}, ψ_{jt}, and δ_{jt} vary across stocks j and time t. Although our parameter estimates are economically close to the ones predicted by invariance, it is possible that ζ_{jt}, ψ_{jt}, and δ_{jt} do vary across j and t enough to statistically reject the invariance hypotheses in some cases. The notation we have developed can be used to investigate this possibility in future research.

6 Implications

The invariance relationships (7), (8), and (16) are like a structural model which describes the implications of market microstructure for bet sizes and transaction costs. The model is fully specified by constants describing the moments of $\tilde{I}$ and the shape of the un-modeled function $C_B(.)$, which determines the constant $\bar{C}_B$. These constants can be inferred from the estimates in section 4 and section 5, but their economic interpretation depends on assumptions about the volume multiplier ζ, the volatility multiplier ψ, and the deflator δ.

Our empirical tests provide not only evidence in favor of the invariance hypotheses but also inputs for calibration. Our empirical results can be summarized as follows. The distribution of portfolio transition orders $|\tilde{X}|$—expressed as a fraction of volume—is approximately a log-normal. It is therefore fully described by two parameters, the log-mean for the benchmark stock estimated to be -5.71 and the log-variance estimated to be 2.53 (table 3). The following formula shows how these estimates can be extrapolated to stocks with other levels of trading activity $W_{jt} = \sigma_{jt} \cdot P_{jt} \cdot V_{jt}$ and volume V_{jt}:

$$\ln\left[\frac{|\tilde{X}_{jt}|}{V_{jt}}\right] \approx -5.71 - \frac{2}{3} \cdot \ln\left[\frac{W_{jt}}{(0.02)(40)(10^6)}\right] + \sqrt{2.53} \cdot \tilde{Z}, \qquad \tilde{Z} \sim N(0,1) \tag{35}$$

$$\gamma_{jt} = 85 \cdot \left[\frac{W_{jt}}{(0.02)(40)(10^6)}\right]^{2/3}. \tag{36}$$

The last equation for the number of bets γ_{jt} follows directly from equation (35) under the assumption that the volume multiplier $\zeta = 2$ and the portfolio transition size multiplier $\delta = 1$. These two equations fully describe the order-flow process.

Our empirical results also suggest that transaction-cost functions can be described by either a linear model or a square-root model. Since both models also have a constant bid-ask spread term, each model is described by two parameters. For an order of 1% of average daily volume in the benchmark stock, the estimates imply market impact costs of $\kappa_I = 2.50 \cdot 10^{-4}$ and spread costs of $\kappa_0 = 8.21 \cdot 10^{-4}$ for the linear model as well as market impact costs of $\kappa_I = 12.08 \cdot 10^{-4}$ and spread costs of

$\kappa_0 = 2.08 \cdot 10^{-4}$ for the square-root model (table 5). The following formulas show how these estimates can be extrapolated to execution costs of an order of X shares for stocks with other levels of trading activity W_{jt}, volume V_{jt}, and volatility σ_{jt}:

$$C_{jt}(X) = \frac{\sigma_{jt}}{0.02}\Big(\frac{8.21}{10^4} \cdot \Big[\frac{W_{jt}}{(0.02)(40)(10^6)}\Big]^{-1/3} + \frac{2.50}{10^4} \cdot \Big[\frac{W_{jt}}{(0.02)(40)(10^6)}\Big]^{1/3} \frac{X}{(0.01)V_{jt}}\Big). \tag{37}$$

$$C_{jt}(X) = \frac{\sigma_{jt}}{0.02}\Big(\frac{2.08}{10^4} \cdot \Big[\frac{W_{jt}}{(0.02)(40)(10^6)}\Big]^{-1/3} + \frac{12.08}{10^4} \cdot \Big[\frac{X}{(0.01)V_{jt}}\Big]^{1/2}\Big). \tag{38}$$

These two equations fully describe the transaction-cost models.

To summarize, formulas (35), (36), (37), and (38) provide a simple way to calculate the number of bets, different percentiles of bet sizes, and transaction costs. The only stock-specific inputs required are expected dollar volume and volatility.[14]

Our results can also be used to calibrate the distribution of $\tilde{I}$ and the invariant cost function $C_B(\cdot)$, which further implies specific quantitative relationships concerning various market microstructure variables such as the number of bets per day, bet sizes, bid-ask spread, and market impact as functions of easily observable trading activity and its components. These implications ultimately depend on values assigned to the volatility multiplier ψ, the volume multiplier ζ, and the deflator δ.

In a more complicated exercise left for future research, this handful of parameters would make it possible to triangulate the value of parameters measuring the fraction of trading volume due to long-term investors rather than intermediaries $1/\zeta$, the fraction of returns volatility generated by bets ψ, and the ratio of the size of bets to the size of portfolio transition orders δ. In the future, careful thinking about calibration of the invariants and estimation of multipliers will be necessary to sharpen predictions based on invariance hypotheses.

7 Conclusion

We have shown that the predictions based on market microstructure invariance are economically consistent with estimates from portfolio transitions data for U.S. equities. We conjecture that predictions based on invariance may hold in other data as well, such as quotes and trades in the TAQ dataset, institutional holdings recorded in 13-F filings, institutional trades reported in the Ancerno dataset, and other datasets.

[14]If one attempts to apply these formulas to assets or time periods where the values for the volume multiplier ζ and the volatility multiplier ψ may be different from the ones relevant for our sample of portfolio transitions, then a simple adjustment to the formulas must be implemented. First, one needs to deflate trading volume and trading volatility in those formulas by appropriate multipliers in order to write those formulas in terms of bet volume and bet volatility for observations in portfolio transitions data. Second, one needs to plug into the modified formulas bet volume and bet volatility appropriate for the market for which calculations are made.

For example, we conjecture that data on news articles can help to show that information flows take place in the same business time as trading.

We conjecture that predictions of market microstructure invariance may generalize to other markets such as bond markets, currency markets, and futures markets, as well as to other countries. Whether market microstructure invariance applies to other markets poses an interesting set of issues for future research.

We do not expect invariance to hold perfectly across different markets and different times periods. Differences in trading institutions across markets might make the volume multiplier and the volatility multiplier vary across markets. We expect transaction costs, particularly bid-ask spread costs (but perhaps not market impact costs), to be influenced by numerous institutional features, such as government regulation (e.g., short sale restrictions, customer order handling rules), transaction taxes, competitiveness of market making institutions, efficiency of trading platforms, market fragmentation, technological change, and tick size. For example, if minimum tick size rules affect bid-ask spread costs, we believe that market microstructure invariance can be used as a benchmark against which the effect of tick size on bid-ask spread costs can be evaluated.

To conclude, market microstructure invariance implies simple scaling laws which lead to sharp statistical hypotheses about bet size and transaction costs. Its implications explain an economically significant portion of the variation in portfolio transition order size and transaction costs when the scaling laws are imposed on the data. The scaling laws enable us to derive simple operational formulas describing order-size distributions and transaction costs; thus, they provide simple benchmarks from which past research can be evaluated and open up new lines of research in market microstructure.

References

Aitchison, J., and J. A. C. Brown. 1957. *The Lognormal Distribution with Special Reference to Its Uses in Economics.* Cambridge University Press.

Allais, Maurice. 1956. "Explication del Cycles Economiques Par Un Modele Non Linéaire a Régulation Retardée." *Metroeconomica*, 8(1): 4–83.

Allais, Maurice. 1966. "A Restatement of the Quantity Theory of Money." *American Economic Review*, 56(5): 1123–1157.

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li. 2005. "Direct Estimation of Equity Market Impact." *Risk*, 18(7): 58–62.

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5: 31–56.

Ané, Thierry, and Helyette Geman. 2000. "Order Flow, Transaction Clock, and Normality of Asset Returns." *Journal of Finance*, 55(5): 2259–2284.

Back, Kerry, and Shmuel Baruch. 2004. "Information in Securities Markets: Kyle meets Glosten and Milgrom." *Econometrica*, 72: 433–465.

Bouchaud, Jean-Philippe, Doyne J. Farmer, and Fabrizio Lillo. 2009. "How Markets Slowly Digest Changes in Supply and Demand." In *Handbook of Financial Markets: Dynamics and Evolution*. 47–160. North Holland.

Breen, William, Laurie Simon Hodrick, and Robert A. Korajczyk. 2002. "Predicting Equity Liquidity." *Management Science*, 48: 470–83.

Brennan, Michael, and Avanidhar Subrahmanyam. 1998. "The Determinants of Average Trade Size." *Journal of Business*, 71(1): 1–25.

Campbell, John Y., and Albert S. Kyle. 1993. "Smart Money, Noise Trading, and Stock Price Behavior." *Review of Economic Studies*, 60: 1–34.

Clark, Peter. 1973. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica*, 41: 135–155.

Derman, Emanuel. 2002. "The Perception of Time, Risk and Return during Periods of Speculation." *Quantitative Finance*, 2: 282–296.

Dufour, Alfonso, and Robert F. Engle. 2000. "Time and the Price Impact of a Trade." *Journal of Finance*, 55(6): 2467–2598.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and Eugene H. Stanley. 2006. "Institutional Investors and Stock Market Volatility." *Quarterly Journal of Economics*, 121: 461–504.

Glosten, Lawrence R., and Paul R. Milgrom. 1985. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics*, 14: 71–100.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan. 2005. "Equilibrium in a Dynamic Limit Order Market." *Journal of Finance*, 60(5): 2149–2192.

Gopikrishnan, Parameswaran, Martin Meyer, Luis A. Amaral, and Eugene H. Stanley. 1998. "Inverse Cubic Law for the Distribution of Stock Price Variations." *European Physical Journal*, 3: 139–140.

Grinold, Richard C., and Ronald N. Kahn. 1995. "Active Portfolio Management." In *Active Portfolio Management*. Probus Publishing.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*, 70(3): 393–408.

Grossman, Sanford J., and Merton H. Miller. 1988. "Liquidity and Market Structure." *Journal of Finance*, 43(3): 617–633.

Hasbrouck, Joel. 1999. "Trading Fast and Slow: Security Market Events in Real Time." NYU Working Paper No. FIN-99-012, SSRN-id1296401, paper available at `http://archive.nyu.edu/bitstream/2451/26851/2/wpa99012.pdf`.

Hasbrouck, Joel. 2007. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading.* Oxford University Press, USA.

Holden, Craig W, Stacey E. Jacobsen, and Avanidhar Subrahmanyam. 2015. "The Empirical Analysis of Liquidity." *Foundations and Trends in Finance*, forthcoming.

Kolmogorov, Andrei N. 1941*a*. "The Local Structure of Turbulence in Incompressible Viscous Fluids at Very Large Reynolds Numbers." *Doklady Akademii Nauk SSSR*, 30: 301–305. Reprinted in *Procedings of the Royal Society* London A 434, 9–13, 1991.

Kolmogorov, Andrei N. 1941*b*. "On The Log-Normal Distribution of Particle Sizes during Break-up Process." *Doklady Akademii Nauk SSSR*, 31(2): 99–101.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.

Kyle, Albert S., and Anna A. Obizhaeva. 2016*a*. "Large Bets and Stock Market Crashes." Working Paper, available at `http://dx.doi.org/10.2139/ssrn.2023776`.

Kyle, Albert S., and Anna A. Obizhaeva. 2016*b*. "Market Microstructure Invariance: A Dynamic Equilibrium Model." Working Paper, University of Maryland, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2749531`.

Lee, Charles, and Mark Ready. 1991. "Inferring Trade Direction from Intraday Data." *Journal of Finance*, 46: 733–746.

Loeb, Thomas. 1983. "Trading Cost: The Critical Link between Investment Information and Results." *Financial Analysts Journal*, 39: 39–44.

Mandelbrot, Benoît, and Howard M. Taylor. 1967. "On the Distribution of Stock Price Differences." *Operations Research*, 15(6): 1057–1062.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review*, 48: 261–297.

Obizhaeva, Anna. 2009. "Portfolio Transitions and Price Dynamics." Working Paper, New Economic School, available at
`http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2591294`.

Obizhaeva, Anna. 2012. "Liquidity Estimates and Selection Bias." Working Paper, New Economic School, available at
`http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1178722`.

Perold, André. 1988. "The Implementation Shortfall: Paper vs. Reality." *Journal of Portfolio Management*, 14(3): 4–9.

Plerou, Vasiliki, Parameswaran Gopikrishnan, Luis A. Amaral, Xavier Gabaix, and Eugene H. Stanley. 2000. "Economic Fluctuations and Anomalous Diffusion." *Physical Review E*, 62: 3023–3026.

Schwarzkopf, Yonathan, and Doyne J. Farmer. 2010. "Empirical Study of the Tails of Mutual Fund Size." *Physical Review E*, 81: 066113.

Stambough, Robert F., and Lubos Pastor. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111: 642–685.

Stoll, Hans R. 1978*a*. "The Supply of Dealer Services in Securities Markets." *Journal of Finance*, 33(4): 1133–1151.

Stoll, Hans R. 1978*b*. "The Pricing of Security Dealer Services: An Empirical Study of NASDAQ Stocks." *Journal of Finance*, 33(4): 1153–1172.

Torre, Nicolo G. 1997. *BARRA Market Impact Model Handbook.* BARRA Inc., Berkeley.

Treynor, Jack. 1995. "The Only Game in Town." *Financial Analysts Journal*, 51(1): 81–83 (reprinted from the Financial Analysts Journal 22, 1971, 12–14, 22).

Wang, Jiang. 1993. "A Model of Intertemporal Asset Prices Under Asymmetric Information." *Review of Economic Studies*, 60(2): 249–282.

Table 1: Descriptive Statistics.

Panel A: Properties of Stocks.

	All	1	2	3	4	5	6	7	8	9	10
Med($V \cdot P$) $\times 10^6$	18.72	1.13	5.10	9.92	15.93	23.87	31.41	42.12	60.25	101.60	212.85
Med(σ) $\times 10^2$	1.93	2.16	2.04	1.94	1.98	1.90	1.86	1.80	1.78	1.77	1.76
Med($Sprd$) $\times 10^4$	12.04	40.96	18.72	13.70	12.02	10.32	9.42	8.12	7.21	5.92	4.83
Med($Turn$) $\times 10^2$	12.72	6.58	12.36	14.80	15.95	16.80	16.36	15.37	14.41	14.92	11.12

Panel B: Properties of Portfolio Transitions Orders.

	All	1	2	3	4	5	6	7	8	9	10
Avg(X/V) $\times 10^2$	4.20	16.23	4.54	2.62	1.83	1.37	1.18	1.08	0.88	0.69	0.49
Med(X/V) $\times 10^2$	0.57	3.33	1.36	0.79	0.53	0.40	0.34	0.30	0.25	0.20	0.14
Avg(X/Cap) $\times 10^4$	1.72	3.55	2.68	2.04	1.59	1.26	1.06	0.91	0.72	0.56	0.37
Med(X/Cap) $\times 10^4$	0.35	0.98	0.80	0.58	0.42	0.32	0.27	0.23	0.19	0.15	0.09
Avg $C(X)$ $\times 10^4$	16.79	44.95	21.46	14.53	12.62	11.70	5.58	9.27	3.99	7.37	6.16
Avg Comm $\times 10^4$	7.43	14.90	9.30	7.86	7.00	6.15	5.49	4.93	4.34	3.62	2.68
Avg SEC fee $\times 10^5$	2.90	3.26	3.02	3.00	2.85	2.84	2.76	2.76	2.73	2.68	2.56
# Obs	439,765	71,000	68,689	41,238	49,000	28,073	29,330	29,778	34,409	40,640	47,608
# Stks	2,583	1,108	486	224	182	106	126	90	102	81	78

Table reports the characteristics of stocks and transition orders. Panel A shows the median average daily dollar volume (in \$ million), the median daily volatility (percent), the median percentage spread (in basis points), the median monthly turnover rate (in percent). Panel B shows the average and median order size (in percent of daily volume and in basis points of market capitalization) as well as average implementation shortfall (in basis points), the average commission (in basis points), and the average SEC fee for sell orders (in percent per 10 basis points). The thresholds of ten volume groups correspond to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar volume for common stocks listed on the NYSE. Group 1 (Group 10) contains orders in stocks with lowest (highest) dollar trading volume. The sample ranges from January 2001 to December 2005.

Table 2: OLS Estimates of Order Size.

		NYSE		NASDAQ	
	All	Buy	Sell	Buy	Sell
$\ln\left[\bar{q}\right]$	-5.67	-5.68	-5.63	-5.75	-5.65
	(0.017)	(0.023)	(0.018)	(0.035)	(0.032)
α_0	-0.62	-0.63	-0.59	-0.71	-0.59
	(0.009)	(0.011)	(0.008)	(0.019)	(0.015)
R^2	0.3167	0.2587	0.2646	0.4298	0.3542
$Q^*/V^* \cdot \delta^{-1} \times 10^{-4}$	34.62	34.14	35.98	31.80	34.78
#Obs	439,765	131,530	150,377	69,871	87,987

Table presents the estimates $\ln[\bar{q}]$ and α_0 for the regression:

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

Each observation corresponds to transition order i with order size X_i, pre-transition price P_i, expected daily volume V_i, expected daily volatility σ_i, trading activity W_i. The parameter $\bar{q}$ is the measure of order size such that for $\delta = 1$, $Q^*/V^* \cdot \delta^{-1} \times 10^{-4}$ measures the median bet size for the benchmark stock, in basis points of average daily volume. The benchmark stock has daily volatility of 2%, share price of \$40, and daily volume of one million shares. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The sample ranges from January 2001 to December 2005.

Table 3: OLS Estimates for Order Size: Model Calibration.

	All	NYSE Buy	NYSE Sell	NASDAQ Buy	NASDAQ Sell
	Restricted Specification: $\alpha_0 = -2/3,\ b_1 = b_2 = b_3 = b_4 = 0$				
$\ln[\bar{q}]$	-5.71	-5.70	-5.68	-5.70	-5.77
	(0.019)	(0.023)	(0.019)	(0.042)	(0.039)
$Q^*/V^* \cdot \delta^{-1} \times 10^{-4}$	33.13	33.46	34.14	33.46	31.20
MSE	2.53	2.61	2.54	2.32	2.56
R^2	0.3149	0.2578	0.2599	0.4278	0.3479
	Unrestricted Specification With 5 Degrees of Freedom: $\alpha_0 = -2/3$.				
R^2	0.3229	0.2668	0.2739	0.4318	0.3616
#Obs	439,765	131,530	150,377	69,871	87,987

Table presents the estimates $\ln[\bar{q}]$ and the mean squared error (MSE) for the regression:

$$\ln\left[\frac{X_i}{V_i}\right] = \ln[\bar{q}] + \alpha_0 \cdot \ln\left[\frac{W_i}{W^*}\right] + b_1 \cdot \ln\left[\frac{\sigma_i}{0.02}\right] + b_2 \cdot \ln\left[\frac{P_i}{40}\right] + b_3 \cdot \ln\left[\frac{V_i}{10^6}\right] + b_4 \cdot \ln\left[\frac{\nu_i}{1/12}\right] + \tilde{\epsilon}_i.$$

with α_0 restricted to be $-2/3$ as predicted by invariance and $b_1 = b_2 = b_3 = b_4 = 0$. Each observation corresponds to transition order i with order size X_i, pre-transition price P_i, expected daily volume V_i, expected daily volatility σ_i, trading activity W_i, and monthly turnover rate ν_i. The parameter $\bar{q}$ is the measure of order size such that for $\delta = 1$, $Q^*/V^* \cdot \delta^{-1} \times 10^{-4}$ measures the median bet size for the benchmark stock, in basis points of average daily volume. The benchmark stock has daily volatility of 2%, share price of $40, and daily volume of one million shares. The R^2s are reported for restricted specification with $\alpha_0 = -2/3, b_1 = b_2 = b_3 = b_4 = 0$ as well as for unrestricted specification with coefficients $\ln[\bar{q}]$ and b_1, b_2, b_3, b_4 allowed to vary freely. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The sample ranges from January 2001 to December 2005.

Table 4: Transaction Cost Estimates in Non-Linear Regression.

		NYSE		NASDAQ	
	All	Buy	Sell	Buy	Sell
β_{mkt}	0.66	0.63	0.62	0.76	0.78
	(0.013)	(0.016)	(0.016)	(0.037)	(0.036)
$\kappa_0^* \times 10^4$	1.77	-0.27	1.14	0.77	3.55
	(0.837)	(2.422)	(1.245)	(4.442)	(1.415)
α_1	-0.49	-0.37	-0.50	0.53	-0.44
	(0.050)	(1.471)	(0.114)	(1.926)	(0.045)
$\kappa_I^* \times 10^4$	10.69	12.08	9.56	12.33	9.34
	(1.376)	(2.693)	(2.254)	(2.356)	(2.686)
z	0.57	0.54	0.56	0.44	0.63
	(0.039)	(0.056)	(0.062)	(0.051)	(0.086)
α_2	-0.32	-0.40	-0.33	-0.41	-0.29
	(0.015)	(0.037)	(0.029)	(0.035)	(0.037)
R^2	0.1010	0.1118	0.1029	0.0945	0.0919
#Obs	439,765	131,530	150,377	69,871	87,987

Table presents the estimates for $\beta_{mkt}, z, \alpha_1, \kappa_0^*, \alpha_2$, and κ_I^* in the regression:

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \kappa_0^* \cdot \left[\frac{W_i}{W^*}\right]^{\alpha_1} + \mathbb{I}_{BS,i} \cdot \kappa_I^* \cdot \left[\frac{W_i}{W^*}\right]^{\alpha_2} \cdot \left[\frac{\phi I_i}{0.01}\right]^{z} + \tilde{\epsilon}_i. \tag{39}$$

where $\phi I_i/0.01 = X_i/(0.01 V_i) \cdot (W_i/W^*)^{2/3}$. S_i is implementation shortfall. $R_{mkt,i}$ is the value-weight market return for the first day of transition. The trading activity W_i is the product of expected volatility σ_i, pre-transition price P_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 i. The parameter $\kappa_I^* \times 10^4$ is the market impact cost of executing a trade of one percent of daily volume in the benchmark stock; and $\kappa_0^* \times 10^4$ is the effective spread cost; both are measured in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The sample ranges from January 2001 to December 2005.

Table 5: Transaction Costs: Model Calibration.

	All	NYSE Buy	NYSE Sell	NASDAQ Buy	NASDAQ Sell
		Linear Model: $z = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$.			
$\kappa_0^* \times 10^4$	8.21	7.19	6.77	9.18	9.27
	(0.578)	(1.122)	(0.794)	(1.563)	(0.781)
$\kappa_I^* \times 10^4$	2.50	3.37	1.92	3.46	2.46
	(0.190)	(0.370)	(0.265)	(0.395)	(0.327)
R^2	0.0991	0.1102	0.1012	0.0926	0.0897
		Square-Root Model: $z = 1/2$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$.			
$\kappa_0^* \times 10^4$	2.08	-1.31	0.92	2.28	4.65
	(0.704)	(1.278)	(0.926)	(2.055)	(0.824)
$\kappa_I^* \times 10^4$	12.08	15.65	11.10	13.50	10.41
	(0.742)	(1.218)	(1.298)	(1.456)	(1.207)
R^2	0.1007	0.1116	0.1027	0.0941	0.0911
		Unrestricted Specification With 12 Degrees of Freedom.			
R^2	0.1016	0.1121	0.1032	0.0957	0.0944
#Obs	439,765	131,530	150,377	69,871	87,987

Table presents the estimates κ_0^* and κ_I^* for the regression:

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} \mathbb{I}_{BS,i} \cdot \kappa_0^* \cdot \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot \frac{\sigma_i^{\beta_1} \cdot P_i^{\beta_2} \cdot V_i^{\beta_3} \cdot \nu_i^{\beta_4}}{(0.02)(40)(10^6)(1/12)} + \\ + \mathbb{I}_{BS,i} \cdot \kappa_I^* \cdot \left[\frac{\phi I_i}{0.01}\right]^z \cdot \left[\frac{W_i}{W^*}\right]^{-1/3} \cdot \frac{\sigma_i^{\beta_5} \cdot P_i^{\beta_6} \cdot V_i^{\beta_7} \cdot \nu_i^{\beta_8}}{(0.02)(40)(10^6)(1/12)} + \tilde{\epsilon}_i.$$

where invariant $\phi I_i/0.01 = X_i/(0.01V_i) \cdot (W_i/W^*)^{2/3}$. S_i is implementation shortfall. $R_{mkt,i}$ is the value-weight market return for the first day of transition. The trading activity W_i is the product of expected volatility σ_i, pre-transition price P_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 i. The parameter $\kappa_I^* \times 10^4$ is the market impact cost of executing a trade of one percent of daily volume in the benchmark stock; and $\kappa_0^* \times 10^4$ is the effective spread cost; both are measured in basis points. The R^2s are reported for restricted specification as well as for unrestricted specification with twelve coefficients β_{mkt}, z, $\kappa_I^*, \kappa_0^*, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ allowed to vary freely. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The sample ranges from January 2001 to December 2005.

Table 6: OLS Estimates of Log of Quoted Spread.

		NYSE		NASDAQ	
	All	Buy	Sell	Buy	Sell
$\ln[\bar{s}]$	-3.07	-3.09	-3.08	-3.04	-3.04
	(0.008)	(0.008)	(0.008)	(0.013)	(0.012)
α_3	-0.35	-0.31	-0.32	-0.40	-0.39
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
R^2	0.4744	0.3545	0.3964	0.5516	0.5721
$e^{\ln[\bar{s}]} \cdot 0.02 \times 10^4$	9.28	9.10	9.19	9.57	9.57
#Obs	434,920	130,700	149,197	68,833	86,190

Table presents the estimates $\ln[\bar{s}]$ and α_3 for the regression:

$$\ln\left[\frac{s_i}{P_i \cdot \sigma_i}\right] = \ln[\bar{s}] + \alpha_3 \cdot \ln\left[\frac{W_i}{W^*}\right] + \tilde{\epsilon}_i,$$

Each observation corresponds to order i. The left side variable is the logarithm of the quoted bid-ask spread s_i/P_i as a fraction of expected returns volatility σ_i. The trading activity W_i is the product of expected daily volatility σ_i, pre-transition price P_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 quantity $\exp(\ln[\bar{s}]) \cdot 0.02 \times 10^4$ measures the median percentage spread for the benchmark stock in basis points. The standard errors are clustered at weekly levels for 17 industries and shown in parentheses. The sample ranges from January 2001 to December 2005.

Figure 1: Invariant Order Size Distribution.

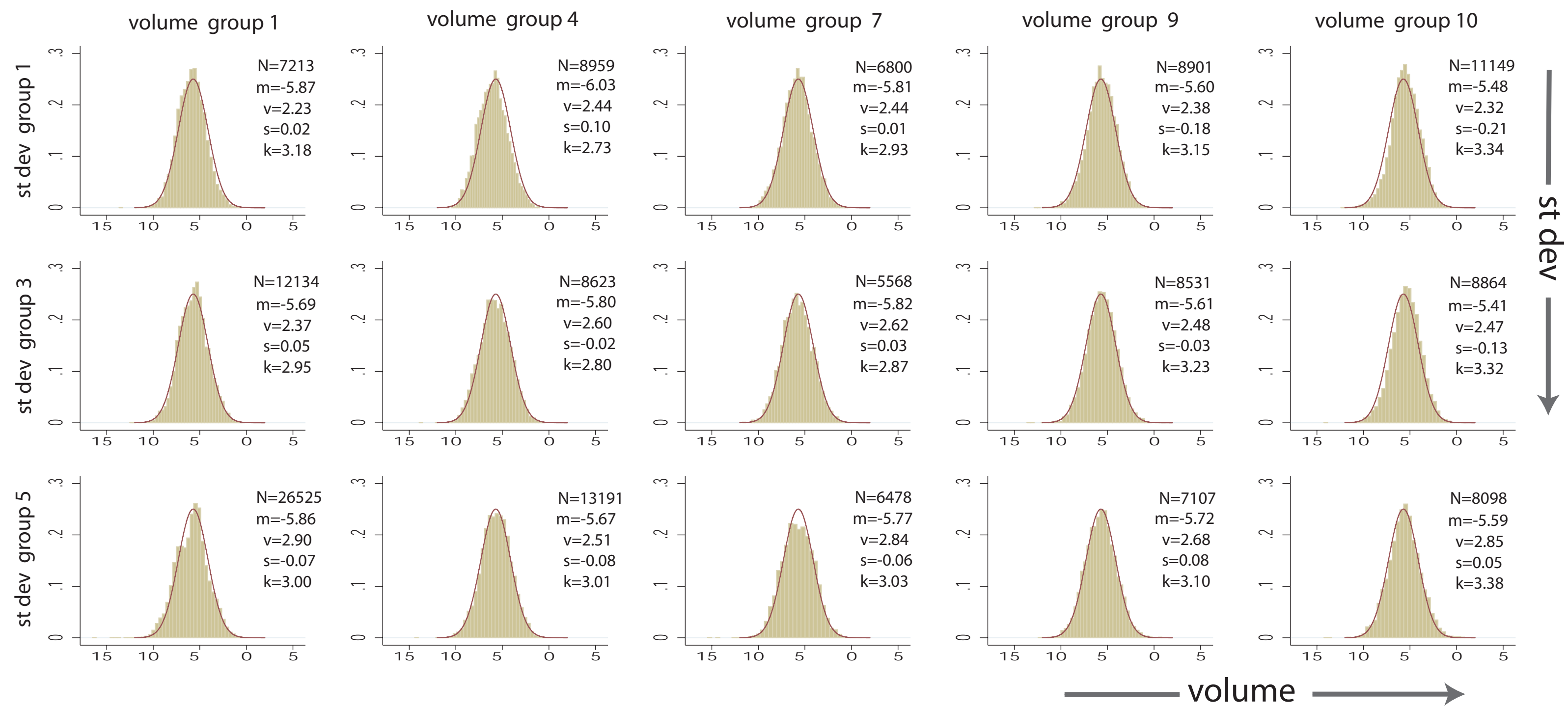
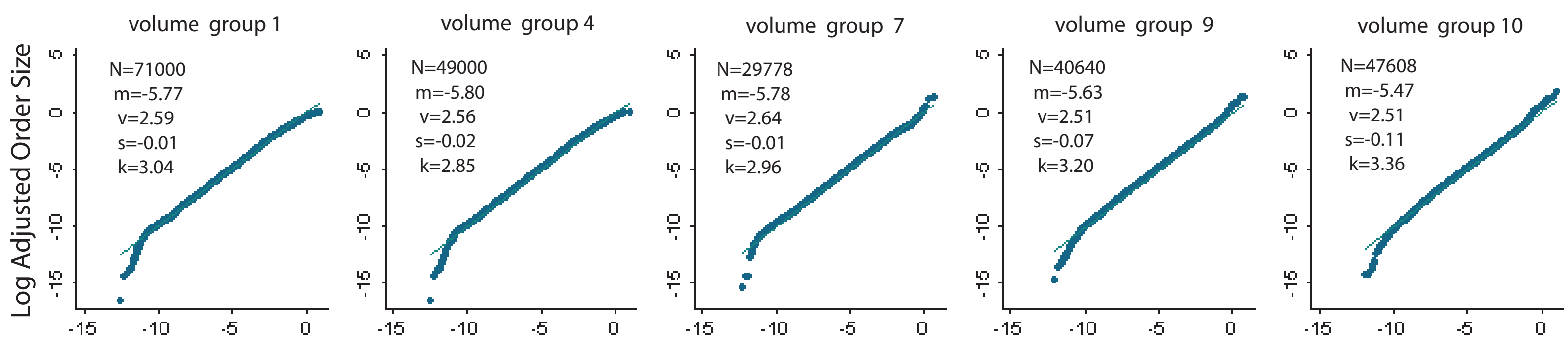


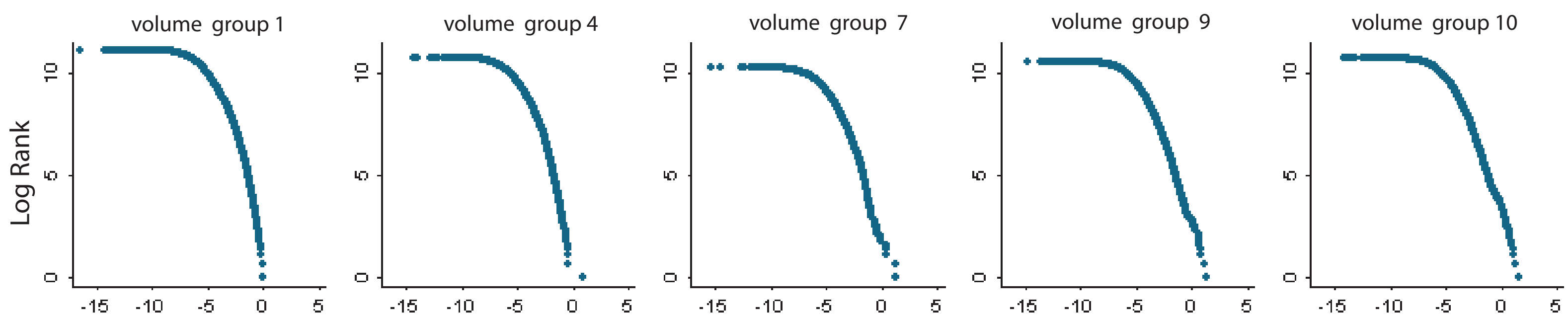
Figure shows distributions of $\ln[\tilde{X}/V]+2/3\cdot\ln[W_i/W^*]$ for stocks sorted into 10 volume groups and 5 volatility groups (only volume groups 1, 4, 7, 9, 10 and volatility groups 1, 3, 5 are reported). X_i is an order size in shares, V_i is the average daily volume in shares, and W_i is the trading activity. The thresholds of ten volume groups correspond to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar volume for common stocks listed on the NYSE. Volume group 1 (group 10) contains orders in stocks with lowest (highest) dollar volume. The thresholds of five volatility groups correspond 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 -5.71 and variance of 2.54 is imposed on each subplot. The common mean and variance are calculated as the mean and variance of distribution over the entire sample. The sample ranges from January 2001 to December 2005.

Figure 2: Invariant Order Size Distribution.

Panel A: Quantile-to-Quantile Plot for Empirical and Lognormal Distribution.

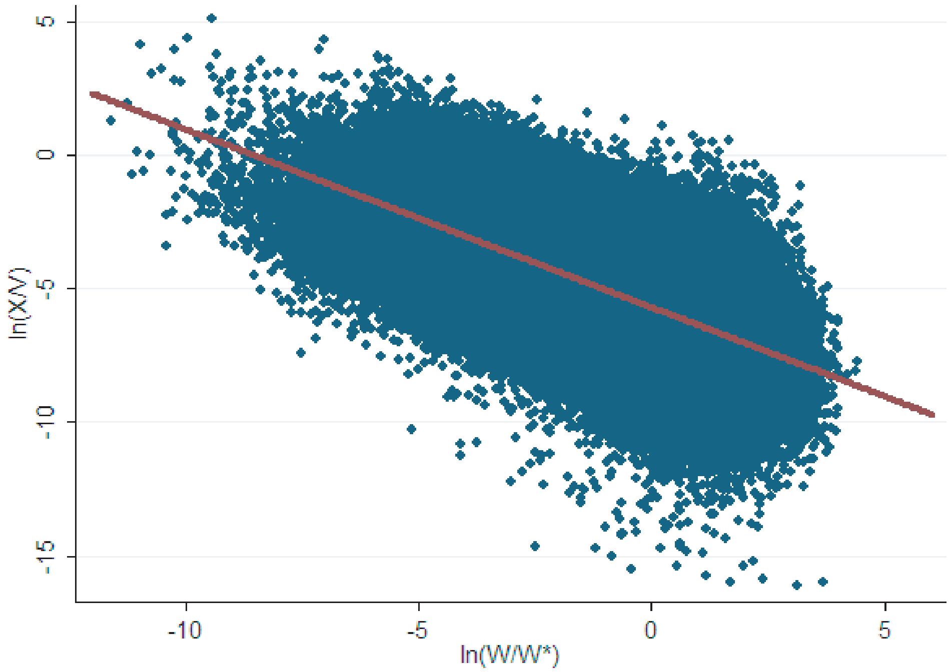


Panel B: Logarithm of Ranks against Quantiles of Empirical Distribution.



Panel A shows quantile-quantile plots of empirical distributions of $\ln[X_i/V_i] + 2/3 \cdot \ln[W_i/W^*]$ 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 that distribution. 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 3: Order Size and Trading Activity.



The figure plots $\ln[X_i/V_i]$ on the vertical axis against $\ln[W_i/W^*]$ on the horizontal axis, where X_i is portfolio transition order size in shares, V_i is average daily volume in shares, and $W_i = P_i \cdot V_i \cdot \sigma_i$ is trading activity. The fitted line is $\ln[X_i/V_i] = -5.705 - 2/3 \cdot \ln[W_i/W^*]$, where the intercept is estimated from an OLS regression with the slope fixed at $-2/3$. There are 400,000+ data points from January 2001 to December 2005.

Figure 4: Invariant Transaction-Cost Functions.

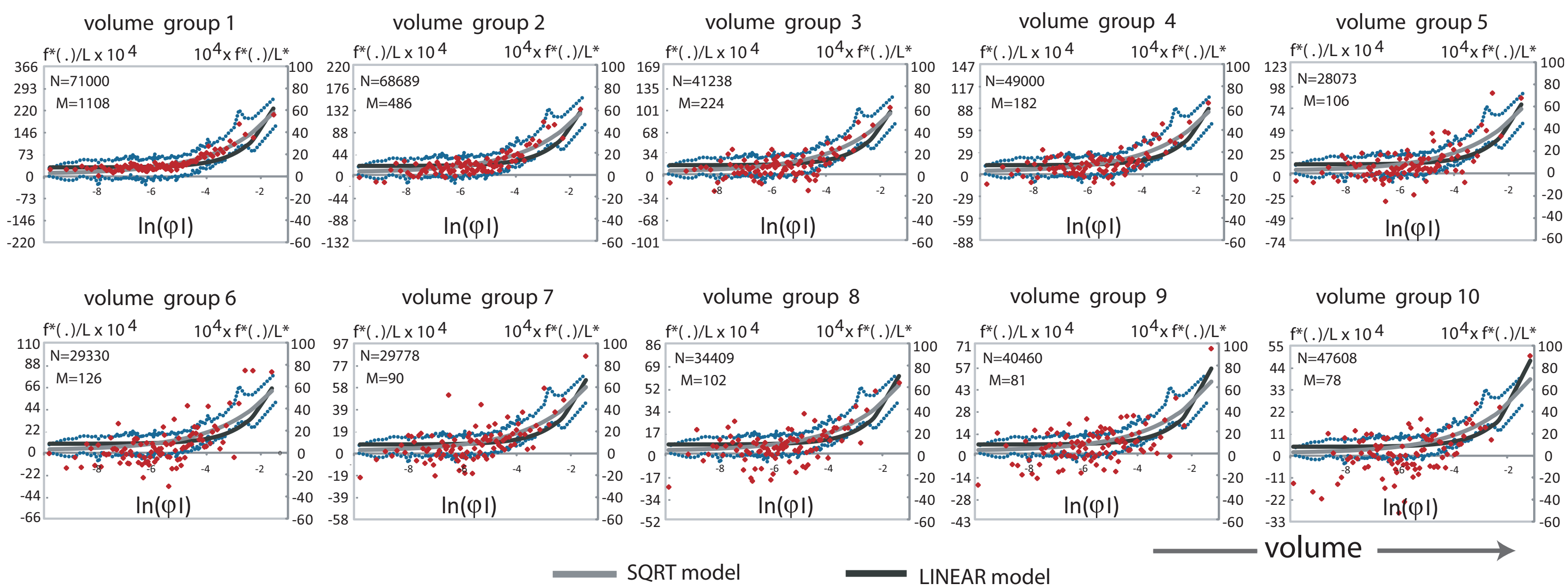


Figure shows estimates of transaction-cost functions for stock sorted into 10 volume groups. On the horizonal axis, there are 100 equally spaced bins based on re-scaled order sizes, $\phi I = \tilde{X}/V \cdot (W_i/W^*)^{2/3}$. For each volume group $k = 1, \ldots, 10$, the subplot contains 100 estimates of dummy variables $f^*(k,h)/L^*, h = 1, \ldots, 100$ from regression (30). On the right side vertical axis, there are units of scaled transaction cost $f^*(.)/L^*$ for a benchmark stock. On the left side vertical axis, there are units of actual transaction cost $f^*(.)/L^k$ for a benchmark stock, where $1/L^k$ is the illiquidity measure for orders in volume group k. The 95th percent confidence interval estimates based on the entire sample are imposed on each subplot (blue dotted lines). The common linear and square-root functions are imposed on each subplot with the parameter estimated on the entire sample. A linear function is $2.50 \cdot 10^{-4} \cdot \phi I/0.01 + 8.21 \cdot 10^{-4}$ (black solid line). A square-root function is $12.07 \cdot 10^{-4} \cdot \sqrt{\phi I/0.01} + 2.08 \cdot 10^{-4}$ (grey solid line). The thresholds of ten volume groups correspond to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar volume for common stocks listed on the NYSE. Group 1 (group 10) contains orders in stocks with lowest (highest) dollar volume. Each subplot also shows the number of observations N and the number of stocks M (for the last month). The sample ranges from January 2001 to December 2005.

Figure 5: Transaction-Cost Functions.

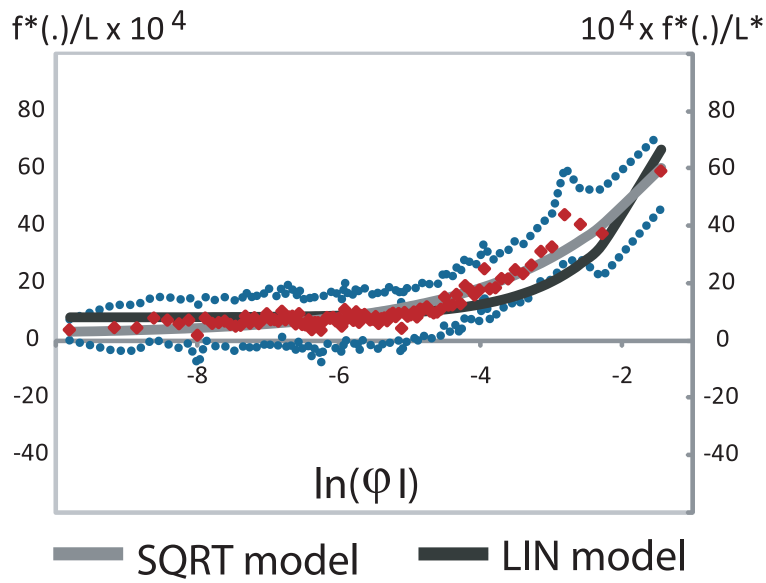


Figure shows estimates of transaction-cost functions based on entire sample. On the horizonal axis, there are 100 equally spaced bins based on re-scaled order sizes, $\phi I = \tilde{X}/V \cdot \Big(W_i/W^*\Big)^{2/3}$. The plot contains 100 estimates $f^*(k,h)/L^*, h = 1, \ldots, 100$ from the regression

$$\mathbb{I}_{BS,i} \cdot S_i \cdot \frac{(0.02)}{\sigma_i} = \beta_{mkt} \cdot R_{mkt,i} \cdot \frac{(0.02)}{\sigma_i} + \mathbb{I}_{BS,i} \cdot \Big[\frac{W_i}{W^*}\Big]^{-1/3} \cdot \sum_{h=1}^{100} D_i^*(k,h) \cdot f^*(k,h)/L^* + \tilde{\epsilon}_i.$$

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 vertical axis presents estimated transaction-cost invariant $f^*(.)/L_i$ in basis points. The 95th percent confidence interval are superimposed (dotted lines). A linear function is $2.50 \cdot 10^{-4} \cdot \phi I/0.01 + 8.21 \cdot 10^{-4}$ (black solid line). A square-root function is $12.07 \cdot 10^{-4} \cdot \sqrt{\phi I/0.01} + 2.08 \cdot 10^{-4}$ (grey solid line). The sample ranges from January 2001 to December 2005.