Selection Bias in Liquidity Estimates

Anna Obizhaeva*

March 22, 2011

Abstract

This paper studies the trading costs estimated from the data on portfolio transitions. Our estimates avoid a selection bias problem, which is endemic to most estimates based on price changes and quantities traded. Since traders often employ price dependent strategies and cancel expensive orders, the conventional estimates tend to overestimate available liquidity. We find that the liquidity is lower than it is usually believed, especially in high volume markets, as illustrated by the Flash Crash in May 2010. High trading costs have implications for the assessment of the viability of trading strategies, the performance of money managers, and the actual limits to arbitrage in financial markets. Our bias-free estimates also allow us to calibrate the existing measures of liquidity and to examine the degree of non-linearity in market impact functions.

Keywords: market impact, spread, trading costs, selection bias.

JEL classification: G14

*Robert H. Smith School of Business, University of Maryland, e-mail:obizhaeva@rhsmith.umd.edu. I am thankful to Charles Jones and Laurence Lescourret who discussed this paper at NBER 2008 and EFA 2008. I thank Yakov Amihud, James Berkovec, Tarun Chordia, Marc Lipson, Mark Loewenstein, Nagpurnanand Prabhala and especially Pete Kyle. I am also grateful to Ross McLellan and Simon Myrgren for their help as well as to Mark Kritzman for his kind support.

1 Introduction

Financial markets are less liquid than many traders think. On May 6, 2010, for example, the execution of a single sell order in the E-mini market has led to a five-percent collapse of market prices. Since the order size was only 3% of the daily trading volume and only 9% of the contemporaneously executed volume, many observers believed that the order was too small to cause the Flash Crash and the E-mini market had plenty of liquidity to absorb it.

This misjudgement prevails because, as emphasized by Hasbrouck (2007), it is difficult to obtain reliable estimates of the actual trading costs. The fundamental problem is that traders usually employ price-dependent strategies and often choose not to execute their orders entirely. Consequently, the estimates of trading costs based on the data of price changes and quantities traded are subject to a selection bias arising from the endogeneity of quantities traded, making markets seem to be more liquid than they actually are. In this paper, we examine the estimates of trading costs based on the sample of portfolio transitions, the institutional specifics of which ensure that our data do not suffer from a selection bias problem. We find that the liquidity is lower than it is usually believed, especially in high volume markets. We discuss the important implications of our findings for the assessment of the viability of trading strategies, the performance of money managers, and the limits to arbitrage in financial markets.

The following example illustrates that the estimates obtained based on the sample of price changes and quantities traded can be significantly distorted by a selection bias. Consider a trader who intends to buy 100,000 shares of stock. At the time the order is placed, the price (benchmark) is $40 per share. The trader purchases 80,000 shares at an average price of $40.20. Suppose that the price then moves adversely to $45 per share, at which point the trader cancels the remaining 20,000 shares of the order. In typical situations, a database of trades may contain the 80,000 shares executed at an average price of $40.20 but not contain any indication that 20,000 shares were not executed at a price which would have been about $45 per share. In this situation, the trading costs of executed trades are calculated as 50 basis points for 80,000 shares. A 50 basis point number is a biased estimate of actual trading costs, because it fails to take account of the cost of at least 1250 basis points that would have been incurred on the 20,000 share portion of the order that was canceled and thus not observed in the data.

Empirically, large fractions of contemplated orders are canceled. Hasbrouck and Saar (2005), for example, report that for non-marketable limit orders, which account for 90% of all incoming orders for Nasdaq-listed stocks traded in Island, only 13% are filled. This suggests that the endogeneity problem can be substantial. It is difficult, however, to assess the magnitude of the selection bias: Unexecuted trades are usually not observed, trading intentions are impossible to reconstruct from sequences of executed transactions, and the exact dependencies of quantities traded on price dynamics are unknown.

Prior literature adopts a number of valuable strategies to handle the selection bias problem. One line of research relies on multivariate linear models for joint dynamics of trades and prices. The examples include Hasbrouck (1988, 1991a, 1991b), Dufour and Engle (2000), and Madhavan, Richardson and Roomans (1997). The resolution of contemporaneous effects in these papers is achieved by imposing a restriction that innovations in trades can affect innovations in price-changes but not vice versa. This restriction, however, leaves out economically plausible scenarios for which quantities executed depend on contemporaneous price dynamics (e.g., limit orders). Alternative approach is to model specific order submission strategies (e.g., Lesmond, Ogden, and Trczinka (1999)) or to predict quantities traded. Cheng and Madhavan (1997) and Conrad, Johnson and Wahal (2003), for example, use endogenous switching regressions to adjust for a selection bias. These procedures involve predicting the choice of trading mechanism and quantities traded in a first-stage regression, and then using the fitted values in a second-stage regression. Yet, as the authors point out, a diverse price–path dependency of trading strategies confounds the estimates of trading costs, making it difficult to assign a structural interpretation to these estimates as ex ante trading costs faced by traders. Other papers suggest examining proprietary transaction data to better identify intended orders (e.g., Almgren et al. (2005), Chan and Lakonishok (1993, 1995), Keim and Madhavan (1997)). A different approach is to altogether disregard the selection bias problem. The representative examples include Breen, Glosten and Harris (1988), Glosten and Harris (1988), Holthausen, Leftwich and Mayers (1987, 1990), and Lillo, Farmer and Mantegna (2003). These methods help to obtain more precise estimates of trading costs, but most likely, they offer only loose bounds on the true trading costs. The reason is that non-zero costs of foregone trades cannot be accurately measured without detailed understanding of the individual trader's trading strategies and, in particular, how these strategies depend on prices.

The prominent study of Amihud (2002) advocates another convenient way to obtain liquidity estimates, referred to as illiquidity ratios, from prices and trading volume in the CRSP data set. Under the assumption that markets are sufficiently resilient so that any random price changes dissipate by the end of the trading day, these estimates might avoid the selection bias problem. Unfortunately, these liquidity estimates are not calibrated, i.e., they implicitly assume that only one order is executed per day and its price impact can be inferred from the ratio of price change to daily volume. Of course, even though this assumption may be valid for small stocks, it is not valid for large stocks, for which numerous trades are executed over a course of a day.

We address the selection bias problem in a different way. We exploit a proprietary database of portfolio transitions. Portfolio transitions are economically significant transactions initiated by institutional sponsors wishing to transfer funds from one portfolio to another. These transfers occur when sponsors replace their fund managers, rebalance their asset classes, or accommodate large cash inflows and outflows. Institutional sponsors usually delegate portfolio transitions to a transition manager. The transition manager replaces the incumbent's legacy portfolio with a new portfolio by selling a portfolio held by the incumbent manager and buying a portfolio chosen by the new manager. The portfolio transition database includes 2,680 portfolio transitions that have been carried out by a leading vendor of portfolio transition services. It involves more than 400,000 individual orders executed over the period from 2001 to 2005.

The advantage of portfolio transitions data is that its sample does not suffer from the selection bias problem because of several unique institutional properties. The transition manager gets the lists of orders to be sold and to be bought the night before the transition begins. He executes entire transition orders and cannot cancel requested transactions. This implies that there are no unexecuted trades. The feedback effect between quantities traded and price changes is thus broken because the former do not depend on the price dynamics during execution. Also, the timing of transitions themselves is determined by a schedule of investment committee meetings of institutional sponsors, who make decisions to undertake transitions; portfolio transition follow shortly after the decision is made. This implies that portfolio transition trades are unlikely to be correlated with short-term price dynamics of individual securities during the period of the transition.

We obtain the estimates of the price impact and the effective spread. On average, the half price impact and the half effective spread are estimated to be 0.30 and 19.20 basis points, respectively. The total bias-free trading costs are thus equal to 19.50 basis points. These estimates are expressed as trading costs for trading one percent of average daily volume, not one dollar or one share, for a benchmark stock with the volatility of two percents. The estimates vary significantly across stocks with different levels of trading activity. The half price impact increases ten-folds from 0.25 basis points for stocks with low trading volume to 2.37 basis points for stocks with high trading volume. The half effective spread drops from 39.29 basis points for stocks with low trading volume to 7.85 basis points for stocks with high trading volume. This implies a particular composition of total trading costs across stocks. For inactively traded stocks, the total trading costs are almost invariant with respect to trade sizes, with spread-related payments being their largest fraction. For actively traded stocks, the total trading costs are, in contrast, highly sensitive to trade sizes, with spread-related payment being less significant.

The comparison of our estimates with those in previous studies is problematic due to a diversity in estimation procedures, definitions, and time periods. Naively comparing our estimates with those in the previous literature, we see that their magnitude is similar. Breen, Glosten and Harris (2002), Hasbrouck (1991 a,b), Glosten and Harris (1988) find, for instance, that the average price impact of a 1000-share trade ranges from 18 to 30 basis points. The average price change after a comparable portfolio transition trade is 13 basis points for actively traded stocks and 40 basis points for inactively traded stocks (1000 shares roughly corresponds to two percents of average daily volume). However, most estimates in previous work are estimated for a pre-decimalization period, while our estimates correspond to a post-decimalization period with significantly lower trading costs. The similarity thus indicates that previous estimates overestimate liquidity in a pre-decimalization period.

The E-mini futures contracts are traded intensively. Our estimates suggest that for actively traded securities, contrary to a conventional intuition, the price impact is high and price changes are very sensitive to sizes of executed orders. The reason is that high trading volume does not indicates deep markets. Rather, high trading volume consists of numerous small trades which tend to cancel each other. The execution of sizable orders therefore may have big price impact. The infamous order executed during the Flash Crash, was large enough to lead to significant price changes.

Note that the illiquidity ratios would significantly overestimate liquidity available in high volume markets, as these ratios do not properly account for multiple independent orders executed during the day. The comparison of our bias-free estimates with the illiquidity ratios reveals how many independent orders, on average, are executed in various markets. We find one order is executed per day in low volume markets, conditioning on non-zero daily volume, and about 27 orders are executed per day in high volume markets. These implied numbers allow us to calibrate the illiquidity ratios. For high volume stocks, for example, the illiquidity ratios overestimate available liquidity by about $\sqrt{27}$ times.

Our estimates provide new insights on trading costs incurred by institutional managers, which they do not report in any publicly available documents. Based on the actual holdings and turnover of the U.S. mutual funds, we conclude that the implicit trading costs of implementing their strategies are usually comparable to management fees they collect, being several times larger for high turnover funds. Our analysis also indicates that market participants should feel being significantly market constraint: Seemingly profitable strategies can absorb only a limited amount of capital before becoming economically infeasible. For example, we estimate that the value-weighted 5/1/1 momentum strategy becomes unprofitable already after \$3 billion investment. Note that the break-even point will be several times higher, if we use the uncalibrated illiquidity ratios as measures of trading costs.

We also examine the non-linearity of price impact functions. In most theoretical studies, price impact functions are linear in equilibrium, whereas most empirical studies document concave price impact functions. The concavity is often attributed to a selection bias, when inexpensive orders are executed as a whole and expensive orders are broken into a sequence of smaller trades. Examining our bias-free price impact functions, we find that they are almost linear for most securities, except for stocks with high trading volume. The non-linearity "discount" for small orders is comparable to the magnitude of bid-ask spreads. This makes us conjecture that ex ante price impact functions are linear, but realized price impact functions are concave, because traders often postpone execution trying to fill their orders on the opposite side of the book and earn the spread providing liquidity. This strategy is especially reasonable during executions of small orders in high volume markets.

The remainder of this paper is structured as follows. Section 2 explains the selection bias problem. It also suggests a novel method for estimation of trading costs based on the total implementation shortfall that accounts for the opportunity costs of foregone trades, related to both bid-ask spread and their potential price impact. Section 3 describes the data. Section 4 reports the estimates. Section 5 discusses the implications. Section 6 concludes.

2 Trading Costs Estimation

We first describe a general methodology of how liquidity parameters - the price impact and the effective spread - can be estimated using data on quantities traded and associated trading costs. We show why this procedure delivers biased estimates of trading costs when applied to samples with endogenous quantities traded and non-random cancelations, features shared by most available samples. We then suggest how to improve this methodology by incorporating the opportunity costs of foregone trades and considering the total implementation shortfall of intended orders.

2.1 How to Estimate Parameters of Trading Costs?

The parameters of trading costs are often estimates from the quantities traded and the realized trading costs associated with these transactions. Let $\bar{Q}_i$ denote the unsigned number of shares in the *ith* order intended to be executed over a fixed time period $[0, T_i]$ and $\mathbb{I}_{BS,i}$ denote the trading direction for order i, which is equal to 1 for buy orders and -1 for sell orders. Let $Q_{t,i}$ denote the number of shares traded during time interval $[0,\, t)$ with $Q_{0,i} = 0$.

If the ith order is executed by time T_i, then $Q_{T_i,i} = \bar{Q}_i$. Often, orders are only partially executed and $Q_{T_i,i} < \bar{Q}_i$.

The execution of the ith order incurs trading costs, quantified by the price impact and the bid-ask spread. The price impact parameter λ reflects how strongly midquote prices respond to trades of different sizes. It describes the variable costs. The dynamics of a midquote $P_{t,i}$ depends on the quantities traded, $Q_{t,i}$, and the news about fundamentals,$d\widetilde{Z}_t$,

$$P_{t,i} - P_{0,i} = \lambda \cdot (\mathbb{I}_{BS,i} \cdot Q_{t,i}) + \sigma_{P,i} \cdot \widetilde{Z}_{t,i}, \tag{1}$$

where $\sigma_{P,i}$ is the volatility of daily price changes and $\widetilde{Z}_{t,i}$ is the cumulative news arriving during $[0,\ t)$. We fix a pre-trade price at a level of $P_{0,i}$ and assume that $\tilde{Z}_{0,i}$ is equal to zero.

The actual transaction price $\hat{P}_{t,i}$ may deviate from the midquote price $P_{t,i}$ by a fixed component, a half bid-ask spread $k/2$,

$$\hat{P}_{t,i} = P_{t,i} + \mathbb{I}_{BS,i} \cdot 1/2k. \tag{2}$$

This fixed component exists if market makers, being not perfectly competitive, charge a fee for providing liquidity. The bid-ask spread k may be related to price impact λ: The spread may be proportional to both price impact λ and a typical trade size. To allow for a more general specification, however, we consider separate parameters for the price impact λ and the spread k.

Our specification of transaction prices effectively includes the permanent and temporary price effects. The permanent price changes are quantified by the price impact λ and assumed to be linear in quantities traded. We examine the linearity assumption later in detail. The transitory price changes are quantified by the spread k. We assume that financial markets are sufficiently resilient, and the spread-related price changes dissipate prior to subsequent transactions. Both assumptions are broadly consistent with the empirical evidence on the form of price impact functions. Indeed, permanent price impact functions are often found to be linear, as discussed, for example, in Almgren et al. (2005), and temporary price deviations are often found to be short-lived, as reported in Coppejans, Domowitz, Madhavan (2004).

Let $\Pi_{t,i}$ denote the realized trading costs in dollars during the execution of the ith order relative to a pre-trade benchmark price $P_{0,i}$,

$$\Pi_{t,i} = \mathbb{I}_{BS,i} \cdot (\hat{P}_{t,i} - P_{0,i}) \cdot Q_{t,i}. \tag{3}$$

Plugging (1) and (2) into (3), we obtain the following relation between trading costs Π_t and the parameters of price impact λ and bid-ask spread k,

$$\frac{\Pi_{t,i}}{P_{0,i}Q_{t,i}} = 1/2\tilde{\lambda} \cdot Q_{t,i} + 1/2k + \sigma_r \cdot \tilde{\epsilon}_{t,i}, \tag{4}$$

where daily returns volatility $\sigma_{r,i} = \sigma_{P,i}/P_{0,i}$ and $\tilde{\epsilon}_{t,i} = \mathbb{I}_{BS,i} \cdot (\widetilde{Z}_{t,i} - \widetilde{Z}_{0,i})$. Equation (4) shows that the realized trading costs $\Pi_{t,i}$ per \$1 traded by time t, or the price "slippage", has several terms. The first term reflects the price impact of shares $Q_{t,i}$ traded; it is linear in the number of executed shares, and its magnitude is determined by the price impact λ. The second term reflects the fixed payment of the half bid-ask spread k. The last term $\sigma_r \cdot \tilde{\epsilon}_{t,i}$

captures unexpected price changes, potentially due to trades of other market participants or arrival of news during $[0, t)$ with a typical magnitude determined by volatility $\sigma_{r,i}$.

In equation (4), the parameters of price impact λ and bid-ask spread k are divided by 2. Costs due to price impact are divided by two because a trader is assumed to walk up or down the demand curve, generating an average cost equal to a half of a marginal cost. Costs due to bid-ask spread are divided by 2 because the bid-ask spread represents a cost for a round-trip trade, while the order $Q_{t,i}$ is either a buy order or a sell order, but not both.

The regression equation (4) and its variations have been used in the prior literature to estimate price impact λ and spread k. The examples include Breen, Hodrick, and Korajczyk (2002) and Glosten and Harris (1988), among many others.

2.2 The Selection Bias Problem

In equations (4), the necessary condition for the estimates to be unbiased is the orthogonality condition, $E(\tilde{\epsilon}_{t,i}|Q_{t,i}) = 0$. This condition does not hold in most available samples because of a natural endogeneity of the regressors $Q_{t,i}$. Indeed, the quantities traded are choice variables, often unknown before the trading begins and potentially influenced by price changes $\tilde{\epsilon}_{t,i}$.

Liquidity estimates are biased, if traders implement price-dependent investment strategies and their trading decisions $\bar{Q}_i$ are related in a systematic way to short-term price dynamics. For traders who trade on a short-lived private information, for example, trading costs will seem to be too high, as the information is being revealed during their trading as well.

The liquidity estimates are also biased when traders adapt their trading strategies and modify quantities traded in response to unfolding market conditions. For instance, traders often cancel their orders if stock prices move against them or trade more aggressively if prices move in their favor. Thus, they can easily manipulate the magnitude of realized trading costs simply by adjusting the quantities traded. In this case, the difference between the average execution price $\hat{P}_{t,i}$ and the pre-trade price $P_{0,i}$ will differ from its unconditional level, $1/2\lambda \cdot Q_{t,i} + 1/2k$.

The sign of the selection bias of the price impact λ depends on the correlation between price changes and quantities traded, $\tilde{\epsilon}_{t,i}$ and $Q_{t,i}$. If traders speed up their trading during unfavorable market conditions, then a positive correlation between these variables makes us overestimate the price impact. In contrast, if traders slow down their trading or cancel orders when the market runs away from them, then the most expensive orders are not executed and therefore not observed in the sample of actual trades. The negative correlation makes us underestimate the price impact. We expect that the later scenario is likely in most samples of actually realized transactions, such as the TAQ and Plexus data sets. In these sample, the price impact λ will appear to be too low comparing to its actual levels. (In Plexus data initial orders are reconstructed based on the quantities traded over a 30-day window).

Given the price impact, the estimates of the bid-ask spread k are such that the average of the right-hand side in equations (4) matches the average of the left-hand side. The estimates of the spread can be therefore biased as well, their bias will depend on the bias of the price impact as well as on a part of $E(\tilde{\epsilon}_{t,i}|Q_{t,i})$ unrelated to quantities $Q_{t,i}$ traded. Note that if $E(\tilde{\epsilon}_{t,i}|Q_{t,i})$ does not depend on $Q_{t,i}$ at all, then only the estimates of the bid-ask spread are biased, whereas the estimates of the price impact do not suffer from selection bias.

In this paper, we exploit the data on portfolio transitions whose unique properties ensure that the orthogonality condition in regression (4) holds. This allows us to obtain and examine the bias-free estimates of liquidity.

2.3 How to Improve Estimation Procedure?

Sometimes, not only executed quantities $Q_{i,t}$, $[0,\ t)$, but also the intended quantities $\bar{Q}_i$ are observed. The knowledge of trading intentions is not an unrealistic assumption. Most traders typically know their own trading intentions and can potentially use this information to get more precise estimates of trading costs. This data may be also available to researchers, if they examine samples with limit orders or they are interested in the trading costs for certain parts of trading packages. In these situations, the estimates can be improved, if the implementation shortfall is used.

The implementation shortfall is a traditional metric of execution quality introduced by Perold (1988). The implementation shortfall $IS_{t,i}$ of ith order at time t is equal to the difference between two returns: the dollar return of an ideal "paper" portfolio where an entire order $\bar{Q}_i$ is transacted instanteneously at pre-trade benchmark prices $P_{0,i}$ and the dollar return of an actual portfolio where $Q_{t,i}$ shares are transacted at actual average execution prices $\hat{P}_{t,i}$ during $[0, t)$,

$$IS_{t,i} = \mathbb{I}_{BS,i} \cdot \bar{Q}_i \cdot (P_{t,i} - P_{0,i}) - \mathbb{I}_{BS,i} \cdot Q_{t,i} \cdot (P_{t,i} - \hat{P}_{t,i}). \tag{5}$$

The implementation shortfall can be also written as the sum of two components,

$$IS_{t,i} = \Pi_{t,i} + OC_{t,i}. \tag{6}$$

The first component $\Pi_{t,i}$ is the trading costs of executed trades $Q_{t,i}$, defined in (3). The second component is the opportunity cost for unexecuted trades $\bar{Q}_i - Q_{t,i}$,

$$OC_{t,i} = \mathbb{I}_{BS,i} \cdot (P_{t,i} - P_{0,i}) \cdot \left(\bar{Q}_i - Q_{t,i}\right). \tag{7}$$

It captures losses of potential gains for the remaining $\bar{Q}_i - Q_{t,i}$ shares. Although these shares have not been traded yet, a trader may have already lost on their execution, because security prices have already deviated from pre-trade benchmark prices $P_{0,i}$ in response to previous transactions.

Plugging (1) and (2) into (5), we obtain the following relation between the implementation shortfall $IS_{t,i}$ marked-to market at time t and the parameters of trading costs,

$$\frac{IS_{t,i}}{P_{0,i}\bar{Q}_i} = 1/2\lambda \cdot \frac{\bar{Q}_i^2 - (\bar{Q}_i - Q_{t,i})^2}{\bar{Q}_i} + 1/2k \cdot \frac{Q_{t,i}}{\bar{Q}_i} + \sigma_r \cdot \tilde{\epsilon}_{t,i}, \tag{8}$$

where $\sigma_{r,i} = \sigma_{P,i}/P_{0,i}$ and $\tilde{\epsilon}_{t,i} = \mathbb{I}_{BS,i} \cdot (\widetilde{Z}_{t,i} - \widetilde{Z}_{0,i})$. Equation (8) shows what determines the implementation shortfall $IS_{t,i}$ per \$1 of the initial order $\bar{Q}_i$ marked-to-market at time t, taking into account losses on unexecuted shares.

This equation has a simple interpretation. Conditioning on $Q_{t,i}$ shares executed by time t, the expected realized implementation shortfall is the difference between the expected costs of executing the total intended order $\bar{Q}_i$ and the expected costs of finishing off its unexecuted part $\bar{Q}_i - Q_{t,i}$, namely, the difference between $1/2\lambda\bar{Q}_i^2 + 1/2k\bar{Q}_i$ and $1/2\lambda(\bar{Q}_i - Q_{t,i})^2 + 1/2k(\bar{Q}_i - Q_{t,i})$. Note that this formula accounts for both price impact and bid-ask spread associated with finishing off remaining shares rather than assuming that they will be executed at a post-trade midquote price or the opposite side of quote (e.g. the ask price for buy order).

Our method improves the approach used in several papers, including Handa and Schwartz (1995) as well as Harris and Hasbrouck (1996), that analyze the trading costs for samples of limit orders. Since many of orders are only partially executed, the authors have suggested to impute to the canceled portion of an order a fill at the opposite-side quote prevailing at the time of cancelation. This method certainly improves the accuracy of estimates comparing to those based solely on executed trades, but it still underestimates the trading costs because unexecuted transactions would have been executed at much less favorable prices than the post-trade ones due to their price impact. Our method accounts for both the spread and the price impact of unexecuted trades.

The regression equation (8) can be used to estimate the price impact λ and the spread k. As before, the estimates will be often biased because the orthogonality condition usually does not hold. The additional information on initial trading intentions $\bar{Q}_i$ exploited in (8), however, helps to obtain more accurate estimates, comparing to those from (4). This method is especially valuable in situations where the number of shares traded does not coincide with intended orders. Note that when these quantities coincide, both methods are equivalent. The intuition is that, even if by conditioning their trading on prices, traders may defer trading costs in time through price dependent trading strategies, these costs will be inevitably realized when intended orders are completed.

3 Data

3.1 Portfolio Transition Data

We exploit a proprietary database of portfolio transitions from a leading vendor of portfolio transition services. During the evaluation period, this portfolio transition vendor supervised more than 30 percent of outsourced U.S. portfolio transitions. The sample includes about 2,680 portfolio transitions executed over the period from 2001 to 2005. This database is derived from the post-transition reports prepared by transition managers for their U.S. clients. These reports contain detailed information on execution of portfolio transitions, thoroughly verified and discussed by both transition managers and institutional sponsors after their implementation.

The portfolio transitions database contains the data on individual transactions. Each observation has the following fields: a trade date, an identifier of a portfolio transition, its starting and ending dates, the name of the stock traded, the number of shares traded, buy or sell indicator, the average execution price, the pre-transition benchmark price, commissions, and fees. The data is given on separate lines for three trading venues: internal crossing networks, external crossing networks, and open market transactions. It is also given separately for each trading day in a trading package. Old and new portfolios usually overlap. For example, both portfolios may have positions in some large and therefore widely held securities. Instead of first selling overlapping holdings from legacy portfolios and then acquiring them into target portfolios, these positions are transferred from one account to another one as "in-kind" transactions, which do not incur any trading costs. For example, if old portfolio had 10,000 shares of IBM and new portfolio had 4,000 shares of IBM in a portfolio transition, then 4,000 shares are transferred in-kind and recorded as in-kind transactions. The rest 6,000 shares will be sold. If transition manager sells these shares in two days with open market trades on the first day (1,000 shares) and both external crosses and open market trades on the second day (2,000 and 3,000 shares), then there will be 4 lines in the database corresponding to IBM stock in a given portfolio transition.

The original data is further grouped so that daily observations are treated separately and one observation corresponds to the triple of a transition, a stock, and a trading day. In-kind transactions are omitted. In the aforementioned example, the execution original data is combined into two lines. The first line corresponds to the first trading day for the IBM order in a portfolio transition and has data on 6,000 shares of IBM to be executed from the first day to the last day of transition. The second line corresponds to the second trading day for the IBM order in a portfolio transition and has data on 5,000 shares of IBM to be executed during the last day of transition. Each line contains the name of the stock, the pre-transition benchmark price, buy or sell indicator, the number of shares executed over different trading venues, the average execution price for each of them. Effectively, we assume that a transition manager faces a new transition task each day. We have chosen to consider daily observations separately because it enables us to analyze potential biases in liquidity estimates, when only parts rather than total trading packages are considered.

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

After exclusions, there are 702,406 daily observations (334,035 buy orders and 368,371 sell orders).

Portfolio transition data has several important properties which make it particularly advantageous for estimating trading costs.

For each stock in a portfolio transition, the quantities to be traded are known precisely at a specific time before the trades are actually executed. The composition of legacy and target portfolios is fixed in the mandates that transition managers receive the night before portfolio transitions begin. Transition managers then execute entire orders regardless of the price dynamics. There are no order cancelations or modifications. This institutional specifics ensures that traded quantities $Q_{i,T}$ coincide with intended quantities $\bar{Q}_i$. The traded quantities are thus not influenced by price changes between the time orders are placed and the time they are executed. Consequently, there is no endogenous relation between quantities traded and realized price changes during portfolio transitions.

The intended quantities $\bar{Q}_i$ are not affected by short-term price dynamics either. The timing of portfolio transitions is likely determined by a schedule of investment committee meetings of institutional sponsors, who make decisions to undertake transitions. The investment committee meets regularly on schedules set well in advance of the meetings. Among the issues boards discuss are the replacement of fund managers and the changes of asset mix. If a decision is made to replace a portfolio manager, then a portfolio transition is arranged shortly after the meeting. These decisions are unlikely to be correlated with short-term price dynamics of individual securities during the period of the transition. These important properties allows us to obtain liquidity estimates that are not distorted by the endogeneity of quantities traded.

These properties of portfolio transitions are not often shared by other data. Consider a database built up from trades by a mutual fund, a hedge fund, or a proprietary trading desk at an investment bank. In such samples, the trading intentions of traders may not be recorded in the database. Furthermore, trading intentions before traders begin trading may not correspond 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 by canceling parts of their orders, thus hard-wiring into their strategies a selection bias problem for using such data to estimate trading costs. The trading intentions themselves can be significantly affected by overall price dynamics, for instance, traders may be following trends or playing contrarian strategies. This dependence of traded quantities on prices, consequently, makes it unfeasible to use implementation shortfall to obtain unbiased estimates of market depth and bid-ask spreads.

The data make it possible to deal with another major problem associated with using implementation shortfall to estimate trading costs, the problem of low statistical power. Suppose, for example, that a trade of one percent of average daily volume has a transactions cost of 20 basis points, but the stock has a price volatility of 200 basis points per day. If we think of the 20 basis points as a random variable which could be positive or negative depending on whether the underlying transition order is a buy or a sell, then the transition order adds about 1% to the variance of the stock's return. This implies that a properly specified regression to estimate trading costs using implementation shortfall is going to have an R^2 of about 0.01, and statistical power is going to be low. Clearly, larger trades with higher trading costs reduce this problem and make the transactions cost easier to estimate.

The portfolio transition database, examined in this paper, addresses the problem of low statistical power in two ways. First, the data involves more than 400,000 individual orders executed by a leading vendor of portfolio transition services over the period 2001-2005. The large number of degrees of freedom increases the statistical power of the estimates. Second, some of the orders are large enough to induce relatively significant trading costs; this increases statistical power as well. As a result, our estimation procedure is powerful enough to deliver precise estimates of price impact and spread.

3.2 CRSP Data: Prices, Volume, and Volatility

Our estimation procedures will require variables such as the expected daily volume, the volatility of daily returns, and pre-trade share price. We calculate these variables are calculated from the Center for Research in Security Prices (CRSP) data.

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

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

As a measure of expected volatility, it would be reasonable to use the implied volatility for our estimation, but this measure is not available for most securities in our sample. We therefore estimate the expected volatility of daily returns, denoted σ_i for ith order, using past daily CRSP returns for the stock involved in the ith trade. We use two different estimates of volatility, a simple estimate equal to average daily volatility from the past month and a more complicated estimate from an ARIMA model.

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

$$\sigma^m_{i,t} = \Big[\sum_{k=1}^{N_{i,t}} r^2_{i,t,k}\Big]^{1/2} \tag{9}$$

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

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

$$\sigma^h_{i,t} = \frac{1}{N^{1/2}_{i,t}} \sigma^m_{i,t}. \tag{10}$$

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

$$(1 - L) \ln \sigma^m_{i,t} = \Theta_0 + (1 - \Theta_1 L - \Theta_2 L^2 - \Theta_3 L^3) u_t. \tag{11}$$

The conditional forecast for the volatility of daily returns is

$$\sigma^e_{i,t} = \frac{1}{N^{1/2}_{i,t}} \exp\left[\ln \sigma^m_{i,t} + \frac{1}{2}\hat{V}(u)\right] \tag{12}$$

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

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

The volatility estimates themselves are probably affected by both spread and price impact that we estimate later. However, since we use the volatility only as a scaling variable in our regressions, these effects should not affect the implied price impact functions under general assumptions such as, for example, that spread and price impact are proportional to fundamental volatility and proportionality constants are not correlated with other variables in the regressions.

3.3 Mutual Funds Data

We also use the CRSP data on the returns and holdings of U.S. mutual funds from 2001 to 2005 to assess the magnitude of actual implicit trading costs in the mutual funds industry. These data allows us to observe monthly data on net returns as well as annual data on the equity holdings of each fund, the total net assets under management, the self-declared investment objective, the management fee, the expense ratios, and the turnover reported annually. We include only mutual funds with no missing data and the investment objective that would require that the main part of their holdings are the U.S. equities, for which we will be able to get the estimates of the trading costs. Our sample include 34,487 monthly observations for 7,168 mutual funds.

3.4 Descriptive Statistics

Table 1 reports statistical characteristics of both individual transition orders and securities traded during portfolio transitions. Statistics are calculated for all securities in aggregate as well as separately for ten groups sorted by average daily dollar volume. Instead of dividing the securities into ten deciles with the same number of securities, volume break points are set at the 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of trading volume for the universe of stocks listed on the NYSE with CRSP share codes of 10 and 11. Group 1 contains stocks in the bottom 30th percentile by dollar trading volume. Group 10 contains stocks in the top 5th percentile. Finer percentiles for the more active stocks make it possible to focus on the stock which are most important economically. For each month, the thresholds are recalculated and the stocks are reshuffled across bins.

Panel A of Table 1 reports statistical properties of securities in the sample. There is a column for each of the ten groupings as well as a column which reports aggregate statistics. For the entire sample of stock, the median trading volume is $19.99 million per day, ranging from $1.22 million for the lowest volume decile to $212.55 million for the highest volume decile. Since the average dollar volume ranges over more than two orders of magnitude, this variation in the data should create statistical power helpful in determining how trading costs and biases of their estimates vary with dollar volume. The median volatility for all stocks is a standard deviation in returns of 1.85 percent per day. Volatility tends to be slightly higher in the lower volume groups than the higher ones. The volatility for the lowest volume group is 2.04 percent, and it is 1.76 percent for the highest volume group.

Panel A of Table 1 reports that the mean bid-ask spread, a quoted spread obtained from the transition database, is 23.67 basis points. From lowest volume group to highest volume group, the bid-ask spread declines monotonically across groups from 64.05 basis points to 7.46 basis points. We will later see that the quoted spreads are by and large similar to statistical estimates of spreads based on our analysis of the execution data.

Panel B of Table 1 reports properties of orders in the portfolio transition data. The mean portfolio order is 3.90 percent of average daily volume of the stock traded. The means decline monotonically across the ten volume groups from 15.64 percent in the smallest group to 0.49 percent in the largest. The median portfolio transitions order is 0.56 percent of average daily volume. The median also declines monotonically across the ten volume groups, from 3.48 percent in the smallest to 0.14 percent in the largest. The fact that the medians are much smaller than the means indicates that the order size is skewed to the right. This is to be expected, since the order size is a non-negative number, and there may be some very small orders from highly diversified portfolios involving smaller transitions as well as very large trades from less diversified portfolios involving larger transitions. The larger order size in the lower deciles generates more statistical power for using implementation shortfall to estimate market impact. Thus, our estimates for the entire sample will be more representative of liquidity for securities in the lower volume groups than in the higher ones.

3.5 Empirical Implementation

To implement the estimation procedures, several adjustments to equation (4) are made, two based on economics and one based on statistics. First, in order to make estimated parameters have intuitive interpretation, we define an arbitrary "benchmark stock" as a stock with volatility of 2% per day. We also re-scale price impact and the bid-ask spread so that they are expressed as trading costs in basis points for trading one percent of average daily volume for the benchmark stock. We chose to report our results for trades as a percentage of volume rather than for a dollar traded, because this is consistent with how most traders think about trading costs; for example, they often specify their strategies by restricting their trading to be equal to a given percentage of trading volume. We denote these constants, $\bar{\lambda}$ and $\bar{k}$, respectively.

Second, transition managers use different trading venues. The bid-ask spread on these venues may be different. In fact, in an internal cross, one of the transition manager's customers buys from the other at some price. It is possible that both the buyer and the seller represent different portfolio transitions, but internal crosses with other types of customers also occur. Since the buyer and the seller pay the same price, it seems reasonable to expect that there is no effective spread incurred for internal crosses. Concerning external crosses and open market transactions, is is assumed that the transition manager optimally chooses the percentages of the orders not crossed internally to execute via open market transactions and external crosses. To the extent that external crosses are cheaper than open market transactions, this is expected to show up as a larger percentage of the orders being executed with external crosses than open market transactions, not as lower market impact and spread costs on external crosses. The fact that both external crosses and open market transactions are used in a significant proportion of orders suggests that there are significant pools of liquidity in both crossing networks and open markets, i.e., neither dominates the other. It is thus assumed that there is price impact $\bar{\lambda}$ associated with internal crosses, of the same magnitude as for external crosses and open market trades and that there is effective spread $\bar{k}$ associated with external crosses and open market trades but not with internal crosses.

Third, since the errors in the regression based on (4) are likely to be proportional in size to the price volatility of the stock, both the right-hand-side and left-hand-side variables are divided by price volatility. This has the effect of making a correction for a heteroscedasticity problem which would otherwise occur.

We implement the estimation in the following way. Observation i corresponds to a triple of a transition, a stock, and a trading day. Let $\bar{X}_i$ denote the unsigned number of shares executed during the period from a given trading day to the last trading day for a given transition order in the ith observation. For multi-day trading sequences, we assume that each day corresponds to a new observation because the transition manager effectively faces a new order to implement with the size equal to the number of shares remained unexecuted at the beginning of that day. For example, if an order was executed over two days, then we have two observations. For $i = 1$, the order $\bar{X}_1$ corresponds to the total number of shares executed in day one and in day two. For $i = 2$, the order $\bar{X}_2$ corresponds to the number of shares left to be executed in day two. Our estimation procedure is then based on the sample of realized trading costs for intended orders, $\bar{X}_1$ and $\bar{X}_2$. Furthermore, let $\bar{X}_{omt,i}$, $\bar{X}_{ec,i}$ and $\bar{X}_{ic,i}$ denote the number of these shares executed in open market transactions, external crosses and internal crosses, respectively, during the period from a given trading day to the end of trading sequence in ith observation. The price impact $\bar{\lambda}^{bf}$ and spread $\bar{k}^{bf}$ can be estimated from the following regression,

$$\frac{\bar{\Pi}_i}{P_{0,i}\bar{X}_i} \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}_i^e} = \frac{1}{2}\bar{\lambda}^{bf} \cdot \frac{\bar{X}_i}{(0.01)\hat{V}_i^e} + \frac{1}{2}\bar{k}^{bf} \cdot \frac{\bar{X}_{omt,i} + \bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i. \tag{13}$$

This regression is based on equation (4) applied to total transition order $\bar{X}_i$ as a proxy for $\bar{Q}_i$ with previously mentioned adjustments. In this regression equation, the left-hand side is the realized trading costs $\bar{\Pi}_i = \mathbb{I}_{BS,i} \cdot (\hat{P}_{ex,i} - P_{0,i}) \cdot \bar{X}_i$ of implementing order $\bar{X}_i$. $P_{0,i}$ is the benchmark price established the night before the trading period, and $\hat{P}_{ex,i}$ is the average execution price for shares traded during the period from a given trading day to the last trading day of transition order. In this regression, the observed data items have subscript i: $\bar{\Pi}_i$, $P_{0,i}$, $\bar{X}_i$, $\bar{X}_{omt,i}$, $\bar{X}_{ec,i}$, $\hat{V}_i^e$, σ_i^e. The term $(0.02)/\hat{\sigma}_i^e$ with $\hat{\sigma}_i^e$ being a proxy for daily volatility adjusts for heteroscedasticity. The term $\bar{X}_i/(0.01)\hat{V}_i^e$ is the size of the trade relative to expected daily volume $\hat{V}_i^e$, scaled so that the size is one for a trade of one percent of expected daily volume.

The variables are scaled so that $\bar{\lambda}^{bf}/2$ estimates in basis points the market impact costs of a trade of one percent of average daily volume, and $\bar{k}^{bf}/2$ estimates in basis points the effective spread cost for open markets and external crossing networks. To be precise, if a trade, representing one percent of average daily volume in the benchmark stock with two-percent daily volatility, incurs 8 basis points of expected costs due to price impact and 3 basis points of expected costs due to spread, then $\bar{\lambda}^{bf}/2 = 8$ and $\bar{k}^{bf}/2 = 3$. The total transactions cost adds up to 11 basis points. Midpoint changes by 16 basis points.

As we have argued above, institutional specifics of portfolio transition orders imply that the orthogonality condition holds and the sample of total transition orders does suffer from the selection bias problem. The superscript "bf" of the estimates $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$ thus indicates that these estimates are bias-free. When calculating other estimates, we will be using similar adjustments.

4 Empirical Results

4.1 The Bias-Free Estimates.

Relying on the unique properties of the portfolio transitions, we obtain the bias-free estimates of the price impact and the spread and then discuss their properties.

Table 6 reports the bias-free estimates of market impact and spread based on the regression (13). The first column of the table reports the results of a regression (13) pooling all the data. The four other columns in the table report results for four separate regressions in which the two liquidity parameters are estimated separately for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells. To adjust standard errors for positive contemporaneous correlation in returns, the 702,446 observations are pooled by week over the period from 2001 to 2005 into 4,389 cluster across 17 industry categories using the pooling option on Stata.

The estimate for half market impact is $\bar{\lambda}^{bf}/2$=0.30 with standard errors of 0.05 and the estimate for half spread is $\bar{k}^{bf}/2$= 19.20 with standard errors of 1.42. These estimates imply that a hypothetical trade in the benchmark stock equal to one percent of daily volume incurs a market impact cost of 0.30 basis points and a spread cost of 19.20 basis points. The total cost is 19.50 basis points.

The estimate for the effective spread $\bar{k}^{bf}$ is double the point estimate for the half-spread $\bar{k}^{bf}/2$, i.e. 38.40 basis points. This estimate is slightly higher than the mean spread of 23.67 basis points reported in Table 1. Similarly, the estimate for $\bar{\lambda}^{bf}$ is double the estimate of 0.30 for $\bar{\lambda}^{bf}/2$, i.e., it is 0.60. This means that a trade of one percent of the average daily volume in the benchmark stock induces the price change of 0.60 basis points.

The reported estimates of trading costs are more representative of those for the transactions in the lower volume groups, because the lower deciles generate more statistical power than the higher deciles. Indeed, a market impact costs are small relative to a spread cost for the lower groups. We will see later that the composition of trading costs differs significantly across the volume groups.

The estimates of trading costs seem to be broadly consistent with the estimates from other studies. For instance, Jones and Lipsons (2001) report in their Table 3 that the proportional effective spread is equal 51 basis points for transactions with less than 1000 shares traded (less than 0.35% of average daily volume) and 70 basis points for transactions with more than 100,000 shares traded (more than 35% of average daily volume). This implies that the half market impact of a trade equivalent to one percent of average daily volume is about $1/2(70-51)/35 = 0.30$ basis points. A fixed part, i.e. half effective spread, is equal to 51 basis points. Both numbers are similar to our estimates of 0.30 and 38.40 basis points, respectively. Note that the estimates of Jones and Lipsons (2001) are based on the pre-decimalization sample in which trading costs are known to be higher than in the post-decimalization sample. The similarity of our estimates may therefore imply that Jones and Lipsons (2001) overestimates the level of available liquidity in their pre-decimalization sample.

The disaggregated results for NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells suggest that buying is more expensive than selling. For NYSE and NASDAQ, both estimated market impact costs and estimated spread costs are larger for buy orders than for sell orders by margins that are economically meaningful if not statistically significant. For example, the market impact for NYSE buys is estimated to be more than twice as large as the market impact for NYSE sells. This is consistent with the idea that the market believes that buy orders contain more information than sell orders. Buy orders in transitions are initiated based on requests to acquire securities chosen by fund managers, who are in turn selected by institutional sponsors as skillful candidates. Buy orders therefore may forecast future price changes and introduce a selection bias into the sample. This double-selection mechanism can lead to the higher estimates for the sample of buy orders. In contrast, sell orders in transitions are liquidation of positions of fund managers, who are terminated by institutional sponsors as unskilled managers. The estimates for the sell orders therefore may be lower. See Obizhaeva (2009) for further discussion of this idea.

One may argue that portfolio transitions might be not representative of institutional trades, because some of them are likely to be uninformed. Even if our estimates give a low bound on the estimate of illiquidity, it does not change our main message that less liquidity is available in financial markets than many people think.

Also, consistent with the previous literature, the effective spread tends to be larger for the Nasdaq-listed securities than for the NYSE/Amex-listed ones (e.g. Christie and Schultz (1994), Huang and Stoll (1996), Keim and Madhavan (1997), and NYSE Market Quality report (2006)) for more recent data.

The Bias-Free Estimates for The Ten Volume Groups. Table 2 presents the bias-free estimates of transaction estimates for ten volume groups. Volume groups are based on the pre-transition trading volume with thresholds 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. We obtain these estimates from regression equation (13) modified in a following way. We replace $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$ with the dummy variables for each of the ten volume groups associated with a half market impact parameter and a half spread parameter for each volume group j, $\bar{\lambda}_j^{bf}$ and $\bar{k}_j^{bf}$. The result is a regression with twenty coefficients, two coefficients for each volume bin, with one coefficient for half market impact and one coefficient for half spread. The regression equation can be written as,

$$\frac{\Pi_i}{P_{0,i}\bar{X}_i}10^4\frac{(0.02)}{\hat{\sigma}_i^e} = \left(\sum_{j=1}^{10}\mathbb{I}_{j,i}\cdot {}^1\!/\!_2\bar{\lambda}_j^{bf}\right)\cdot\frac{\bar{X}_i}{(0.01)\hat{V}_i^e} + \left(\sum_{j=1}^{10}\mathbb{I}_{j,i}\cdot {}^1\!/\!_2\bar{k}_j^{bf}\right)\cdot\frac{\bar{X}_{omt,i}+\bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i \quad (14)$$

where $\mathbb{I}_{j,i}$ is an indicator equal to one if ith order is executed in a stock from jth volume group, $j = 1, ..10$. To adjust standard errors for positive contemporaneous correlation in returns, the 702,446 observations are pooled by weeks over the period from 2001 to 2005 into 4,389 clusters across 17 industry categories using the pooling option on Stata. We chose to consider one regression rather than a set of separate regressions for each volume group, since this enables us to take into account potential correlation of residuals across volume groups.

Table 2 shows that market impact $\bar{\lambda}^{bf}$ increases with trading volume. Its estimates increase from $\bar{\lambda}_1^{bf} = 0.25$ for inactively traded stocks (Group 1) to $\bar{\lambda}_{10}^{bf} = 2.37$ for actively traded stocks (Group 10). The estimate for $\bar{\lambda}_j^{bf}$ is double the estimate for $\bar{\lambda}_j^{bf}/2$. This implies that, on average, trading one percent of daily volume triggers the change of 0.50 basis points in midquotes of securities with low trading volume and the change of 4.74 basis points for securities with high trading volume, if daily volatility is equal to 2%. The market impact is thus ten times higher for actively traded stocks than for inactively traded stocks. This positive association between the market impact and trading volume is observed not only for the entire sample but also by and large for disaggregated subsamples of NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells.

Table 2 also shows that spread $\bar{k}^{bf}$ decreases with trading volume. Its estimates decreases from $\bar{k}_1^{bf} = 39.29$ basis points for inactively traded stocks (Group 1) to $\bar{k}_{10}^{bf} = 7.85$ basis points for actively traded stocks (Group 10). The spread therefore drops from the especially high level for the lowest volume group to the level five times lower for the highest volume group.

To summarize, we find that transactions in stocks with *higher* volume tend to be associated with *lower* spread but with *larger* market impact. The first finding is consistent with a conventional intuition. The second one seems to contradict it, since it is commonly believed that market impact decreases with trading volume. Indeed, market makers require smaller compensation when trading larger stocks, as they face smaller risk of information asymmetry and can easier either find a hedge for these securities or reallocate them to other market participants.

Although the positive association between the market impact and overall trading volume appears to be counterintuitive, similar patterns have been mentioned in the previous literature. For instance, Breen, Hodrick, Korajyzyk (2002) estimate the price impact using transactions from the TAQ data set. They found that the market impact increases with the market capitalization, if market impact is defined as the price change after a trade equivalent to one percent of turnover. To avoid these "unreasonable" results, the authors chose to focus on the market impact defined as the price change after a $1-trade. The later estimates "reasonably" decreases with the market capitalization (see also Hasbrouck (1991b), Chen, Stanzl and Watanabe (2005)). This suggests that the specification of market impact is not inconsequential. Moreover, for a more meaningful specification of the market impact per percent of daily order flow, the estimates tend to increase rather than decrease with trading volume. We will discuss possible explanations for these patterns in Section 5.

4.2 The Selection Bias and Improved Estimates

We also report two other sets of liquidity estimates based on the realized trading costs and the implementation shortfall for daily portfolio transition trades rather than for total orders. We expect these estimates to be different from the bias-free estimates, because transition managers most likely implement price-dependent trading strategies introducing a selection bias into the sample. The sign of the bias should be related to the types of strategies that transition managers employ. This exercise also allows to examine the improvement in the estimates when they are based on the actual implementation shortfall rather than the realized trading costs.

The Estimates Based on the Realized Trading Costs. The first set of estimates, $\bar{\lambda}^{rtc}$ and $\bar{k}^{rtc}$, is based on the realized trading costs of daily transition trades. Each observation i corresponds to a triple of a transition, a stock, and a trading day. The variable X_i denote the number of shares executed during a given trading day for a given transition order in ith observation. For example, if an order was executed during two days, then we have two observations. For $i = 1$, the order X_1 corresponds to the number of shares executed in day one. For $i = 2$, the order X_2 corresponds to the number of shares executed in day two. The estimation procedure is based on the realized trading costs of daily trades, X_1 and X_2. The variable $X_{omt,i}$, $X_{ec,i}$ and $X_{ic,i}$ denote the number of these shares executed in open market transactions, external crosses and internal crosses, respectively, during a given trading day in ith observation.

$$\frac{\Pi_i}{P_{0,i}X_i} \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}_i^e} = \frac{1}{2}\bar{\lambda}^{rtc} \cdot \frac{X_i}{(0.01)\hat{V}_i^e} + \frac{1}{2}\bar{k}^{rtc} \cdot \frac{X_{omt,i} + X_{ec,i}}{X_i} + \tilde{\epsilon}_i \qquad (15)$$

This equation is similar to equation (13). The observed data items have subscript i: Π_i, $P_{0,i}$, X_i, $X_{omt,i}$, $X_{ec,i}$, $\hat{V}_i^e$, σ_i^e. The variables are scaled so that $\bar{\lambda}^{rtc}/2$ estimates in basis points the market impact costs of a trade of one percent of average daily volume, and $\bar{k}^{rtc}/2$ estimates in basis points the effective spread cost for open markets and external crossing networks. The superscript "rtc" shows that these estimates of the price impact and bid-ask spread are based on the realized trading costs.

Since we are more interested in examining the bias of these estimates relative to the bias-free estimates $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$, we implement the following procedure. We stacked regressions (13) and (15) together and impose the two constraints on coefficients: $\bar{\lambda}^{rtc} = \bar{\lambda}^{bf} + \Delta_{\lambda}^{rtc}$ and $\bar{k}^{rtc} = \bar{k}^{bf} + \Delta_{k}^{rtc}$. The estimates of Δ_{λ}^{rtc} and Δ_{k}^{rtc} therefore will quantify the bias of $\bar{\lambda}^{rtc}$ and $\bar{k}^{rtc}$ relative to the bias-free estimates $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$. We also estimates these differences for each of the ten volume group as well, similarly to equation (14).

Table 4 reports the result. The estimate of $\bar{\lambda}^{rtc}$ are biased upwards by 0.09. The estimate of $\bar{k}^{rtc}$ are biased downwards by -5.96. The biases $\Delta_{\lambda,j}^{rtc}$ are estimated to be positive for each volume group. Their magnitude ranges from 16% to 54%. For some volume groups, these biases are statistically significant, even though the selection bias problem is expected to be of a limited importance in the sample of portfolio transitions: most transactions do not take more than one day, making both procedures (13) and (15) largely overlap. The bias of spread $\Delta_{k,j}^{rtc}$, in contrast, is negative for each volume group, partially undoing the upward bias of market impact.

Our results reveal that the sign of the market impact bias is different for buy and sell orders (not reported). For buy orders, Δ_{λ}^{rtc} is either negative or insignificant. For sell orders, Δ_{λ}^{rtc} is either positive or insignificant. The direction of bias is related to the correlation between trading strategies and the price dynamics. It tells us how transition managers adjust their trading strategies in response to changes in stock prices. The positive bias for sell orders implies that transition managers tend to speed up selling securities from old portfolios during adverse price changes. The negative bias for buy orders implies that transition managers prefer to slow down buying securities into new portfolios in the same situation. These trading patterns are consistent with the agency-based approach of managing portfolio transitions, when managers do not employ their own capital but have to use proceeds of their sales to fund their purchases.

The Estimates Based on the Implementation Shortfalls. The second set of estimates, $\bar{\lambda}^{is}$ and $\bar{k}^{is}$, is based on the implementation shortfall of daily portfolio transition trades rather than their realized trading costs. It properly accounts for the opportunity costs of trades unexecuted by a given trading day. As before, observation i corresponds to a triple of a transition, a stock, and a trading day. Let $\bar{X}_i$ denote the total number of shares executed during the period from a given trading day to the last trading day for a given transition order in ith observation. Let X_i denote the number of shares executed during a given trading day for a given transition order in ith observation. For example, if an order was executed during two days, then we have two observations. For $i = 1$, $\bar{X}_1$ corresponds to the total number of shares executed in day one and two and X_1 corresponds to the total number of shares executed in day one. For $i = 2$, $\bar{X}_2 = X_2$ corresponds to the number of shares executed in day two. The estimation is based on the marked-to-market implementation shortfall of daily trades. Let $X_{omt,i}$, $X_{ec,i}$ and $X_{ic,i}$ denote the number of these shares executed in open market transactions, external crosses and internal crosses, respectively, during a given trading day in ith observation. The price impact $\bar{\lambda}^{is}$ and spread $\bar{k}^{is}$ are estimated from the following regression:

$$\frac{IS_i}{P_{0,i}X_i} \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}_i^e} = \frac{1}{2}\bar{\lambda}^{is} \cdot \frac{\bar{X}_i^2 - (\bar{X}_i - X_i)^2}{(0.01)\hat{V}_i^e} + \frac{1}{2}\bar{k}^{is} \cdot \frac{X_{omt,i} + X_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i \qquad (16)$$

This regression is based on equation (8) applied to daily transition trades X_i as a proxy for Q_i and $\bar{X}_i$ as a proxy for $\bar{Q}_i$ with previously mentioned adjustments. The left-hand side is the implementation shortfall $IS_i = \mathbb{I}_{BS,i} \cdot (\hat{P}_{ex,i} - P_{0,i})X_i + \mathbb{I}_{BS,i} \cdot (\hat{P}_{end,i} - P_{0,i})(\bar{X}_i - X_i)$ with $\bar{X}_i$ shares intended to be traded and X_i shares actually traded. $P_{0,i}$ is the benchmark price established the night before the trading period, $\hat{P}_{ex,i}$ denotes the average execution price for shares traded during a given trading day, and $P_{end,i}$ is the price at the end of a given trading day. Other adjustments are similar to those done for equation (13). In this regression, the observed data items have subscript i: IS_i, $P_{0,i}$, $P_{end,i}$, $\bar{X}_i$, X_i, $X_{omt,i}$, $X_{ec,i}$, V_i^e, σ_i^e. The variables are scaled so that $\bar{\lambda}^{is}/2$ estimates in basis points the market impact costs of a trade of one percent of average daily volume, and $\bar{k}^{is}/2$ estimates in basis points the effective spread cost for open markets and external crossing networks. The superscript "is" shows that these estimates of the price impact and bid-ask spread are based on the implementation shortfall.

Again, we stacked regressions (13) and (16) together and impose the two constraints on coefficients: $\bar{\lambda}^{is} = \bar{\lambda}^{bf} + \Delta_{\lambda}^{is}$ and $\bar{k}^{is} = \bar{k}^{bf} + \Delta_{k}^{is}$. The estimates of Δ_{λ}^{is} and Δ_{k}^{is} therefore will quantify the bias of $\bar{\lambda}^{is}$ and $\bar{k}^{is}$ relative to the bias-free estimates $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$. We also estimates these differences for each of the ten volume group as well, similarly to equation (14).

Table 4 reports the result. The estimate of $\bar{\lambda}^{is}$ are biased upwards by 0.03. The estimate of $\bar{k}^{is}$ is biased downwards by -5.63. Although these biases are statistically significant, their magnitude is, however, smaller comparing to those of the estimates $\bar{\lambda}^{rtc}$ and $\bar{k}^{rtc}$. The reduction in the bias is observed across all disaggregated subsamples of NYSE Buys, NYSE Sells, NASDAQ Buys, and NASDAQ Sells (not reported). These results illustrate that, when the data on intended orders are available, the estimation method based on the total implementation shortfall helps to improve to liquidity estimates based on the realized trading costs only, with no adjustment for the opportunity costs.

5 Implications

5.1 Expected Execution Costs

It is important to know the magnitude of trading costs. One may be interested in studying whether seemingly profitable trading strategies remain profitable after adjusting their returns for trading costs. For example, Chordia et al. (2009) examine the profitability of the long-short strategies based on the earning surprises. Cooper, Gutierrez, and Marcum (2005) analyze the profitability of strategies based on book-to-market ratios, firm sizes, and lagged returns. One may be also interested in studying whether abnormal performance of some market participants can survive incurred trading costs. The example of this strand of literature is the work on mutual fund performance by Wermers (2000). All these papers adjust calculated gross returns for potentially incurred trading costs relying on the estimates from the studies of earlier pre-decimalization samples such as the work of Keim and Madhavan (1997).

This paper suggests a simple formula for the expected trading costs with separate estimates for both their fixed and variable parts. Our estimates are obtained from data on the post-decimalization period from 2001 to 2005, when trading costs are known to decline substantially. Because of institutional specifics of portfolio transitions, these estimates also do not suffer from a selection bias problem. In general, our estimates of trading costs are usually bigger than the conventional estimates, which are either influenced by the selection bias or uncalibrated. Later, when we analyze aforementioned questions using the bias-free estimates, we conclude that trading costs have more significant effect in financial markets than previously thought.

Let X denote the number of shares in the order. Let $C(X)$ denote the expected cost of trading X shares of some stock, measured in basis points. Using reported bias-free estimates $\bar{\lambda}^{bf}/2 = 0.30$ and $\bar{k}^{bf}/2 = 19.20$, we can write $C(X)$ as follows:

$$C(X) = 0.30 \cdot \frac{\sigma_r^e}{0.02} \cdot \frac{X}{(0.01)V^e} + 19.20 \cdot \frac{\sigma_r^e}{0.02}. \tag{17}$$

where σ_r^e and V^e are expected volatility of daily returns and expected daily volume in shares. In this equation, the first term of trading costs on the right-hand-side quantifies the size of the variable costs due to market impact, and the second term quantifies the size of fixed costs due to bid-ask spread. The ratio of X to $(0.01)V^e$ is one when the trade size is one percent of expected daily volume. For example, if a trade X represents one percent of expected daily volume in the benchmark stock with volatility 2%, then its expected execution cost adds up to 19.50 basis points. The term $\frac{\sigma_r^e}{0.02}$ adjusts the estimates if volatility of a given stock is different from volatility of a benchmark stock, for which bias-free estimates are reported.

Expected Costs for Different Volume Groups. Our findings show that the trading costs differ substantially across stocks. The more precise estimates for expected trading costs can be obtained if formulas are presented for individual volume groups. The expected trading costs $C_j(X)$ in basis points as functions of a trade size, $\frac{X}{(0.01)V^e}$, in percents of expected daily volume for stocks in the ten volume groups, $j = 1, ..10$,

$$C_j(X) = \frac{1}{2}\bar{\lambda}_j^{bf} \cdot \frac{\sigma_r^e}{0.02} \cdot \frac{X}{(0.01)V^e} + \frac{1}{2}\bar{k}_j^{bf} \cdot \frac{\sigma_r^e}{0.02}. \tag{18}$$

For stocks with low trading volume (Group 1), one has to plug into (18) the estimates $\bar{\lambda}_1^{bf}/2 = 0.25$ and $\bar{k}_1^{bf}/2 = 39.29$. For stocks with high trading volume (Group 10), one has to plug into (18) the estimates $\bar{\lambda}_{10}^{bf}/2 = 2.37$ and $\bar{k}_{10}^{bf}/2 = 7.85$. The expected trading costs $C(X)$ for other volume groups can be obtained using the corresponding estimates from Table 2.

Figure 2 plots the price impact functions for the volume groups 1, 5, and 10. These functions are similar to equations (18) but there is no 1/2 multiplier in front of the price impact. The spread is reflected in their intercept of the price impact functions. The market impact is shown by their slopes. We assume that the daily volatility σ_r^e is equal to the average volatility for a given group from Table 1.

The figure reveals that the composition of expected trading costs differ significantly across stocks with low and high trading volume. For Group 1, the market impact is a relatively small fraction of the total costs $C_1(X)$, making them almost invariant with respect to a trade size, whereas the spread-related payments account for the large fraction of $C_1(X)$. For Group 10, in contrast, the market impact is large relative to the spread-related payments, making the total costs $C_{10}(X)$ very sensitive to a trade size. The price impact of orders increases significantly, as the bigger fraction of daily volume is being executed. Note that, regardless of larger market impact, trading in stocks with high volume will be still less expensive than trading in stocks with low volume for a wide range of trade sizes. Furthermore, a 1\$-transaction in active stocks is cheaper than a 1%-transaction in inactive stocks, since the former corresponds to a much smaller percent of daily volume than the latter.

5.2 Market Impact in High Volume Markets

We find that market impact is increasing with the overall trading volume: In particular, the execution of 1% of average daily volume for large stocks is ten times more expensive than the execution of the same percentage for small stocks, holding spread-related payments constant.

We deliberately state our results in this form, as we believe that most traders think about execution costs in terms of costs for executing a given percentage of trading volume; this approach underlies, for example, the popular VWAP trading of executing a given fraction of daily volume. Moreover, most traders have a wrong intuition about the cost of trading in liquid markets, having a tendency to underestimate transaction costs in these markets.

The recent flash crash perfectly illustrates this point. On May 6, a large trader, identified by journalists as Waddell & Reed Financial, has sold 75,000 S&P 500 E-mini contracts on the Chicago Mercantile Exchange. This order corresponded to only 3.75% of a daily trading volume or about 9% of volume traded contemporaneously. According to the conventional intuition, this order was too small to cause the five-percent drop in the market price. The CFTC and the SEC have issues a joint report stating, however, that the flash crash was triggered by a single large sale, even though it may seem to be a counterintuitive explanation.

In this paper, we would like to emphasize similar intuition. More intensive competition in larger markets manifests itself in more expensive trading. Large stocks are traded intensively, the competition among market participants is significant, and the trading volume is high, but this trading volume is usually two-sided, buys alternates with sells. For these stocks, a one-sided trade equivalent to one percent of the average daily volume is a large trade representing a distinctive event. The market impact of this trade is likely to be substantial, just as it was during the flash crash. On the other hand, small stocks are not traded a lot, and the trading volume is small, but this trading volume constitutes mostly one-sided trading. For these stocks, the information content of a trade equivalent to one percent of the average trading volume is less significant. The market impact of this trade is most likely to be small.

A simple example can illustrate our intuition about a relationship between the market impact and the degree of competition. Let us assume that there are γ traders operating in the market. These traders have different views about the true value of security. The differences in their views may be a created by the differences in their private information or in their interpretations of public news. Based on their views, traders place buy or sell orders. For simplicity, we assume that $\tilde{\epsilon}_i$ is 1 if trader i decides to buy a security, and -1 if he decides to sell it. Let us assume that traders submit orders of equal sizes. Then, the size of their orders is $Q = ADV/\gamma$, where ADV is the average number of shares traded at that day. The stock price change in response to these orders. The change of price ΔP over a given day is determined as,

$$\Delta P = \sum_{i=1}^{\gamma} \tilde{\epsilon}_i \cdot Q \cdot \lambda,$$

where λ relates to the price elasticity. If trades $\tilde{\epsilon}_i$ are independent, then the daily volatility, σ_P, is equal to

$$\sigma_P = \sqrt{\gamma} \cdot Q \cdot \lambda = \sqrt{\gamma} \cdot \frac{ADV}{\gamma} \cdot \lambda$$

Consequently, the price impact of a trade equivalent to one percent of daily volume expressed as a percent of daily volatility is equal to $\sqrt{\gamma}$. This example illustrates that market impact increases with the number of traders γ. In other words, more intensive competition between market participants, existing in financial market for large stocks, leads to higher market impact.

Several other factors may explain our counterintuitive finding about market impact and trading volume. An alternative explanation could be a possible concavity of the price impact functions. Indeed, Table 1 shows that the average order size (in percents of average trading volume) is lower for stocks with high trading volume. If market impact is concave making larger orders "cheaper" than smaller ones, then trading in inactively traded stocks may appear to be less expensive that trading in actively traded stocks. We examine the non-linearity of price impact functions below.

The resilience could also explain smaller market impact for transactions in the lower-volume groups. In the portfolio transition data, the order execution for securities with low trading volume is slower than for securities with high trading volume. The former may benefit from price reversals and appear to be cheaper. However, trading horizons are not significantly different across securities. Moreover, slightly shorter trading horizons for large stocks may be potentially compensated by stronger forces of their resilience, thus mitigating the impact of resilience on the total costs.

5.3 Calibration of Existing Liquidity Measures

For some measures of illiquidity, the selection bias may be of a less concern, but these measures are often not calibrated. For example, Amihud (2002) suggests to estimate the illiquidity as the average ratio of the absolute value of daily returns to the daily volume. Under the assumption that the short-term price changes dissipate prior to the end of the trading day, the selection bias does not distort these popular estimates. The illiquidity ratios, however, effectively assume that daily price changes are caused by trading volume. These measures would be reasonable if there were only one trade executed per day. Obviously, this is not a valid assumption for most stocks. For example, illiquidity will be significantly underestimated for large stocks with numerous orders executed per day, sinc these order cancel each others price impacts. These measures are thus not calibrated.

In particular, suppose that γ independent orders arrive per day to the marketplace. It is easy to see that the calibrated illiquidity ratios can be calculated as,

$$ILLIQ_i = \overline{\frac{|R_i|}{V_i} \cdot \sqrt{\gamma_i}}, \tag{19}$$

where Amihud (2002) assumes that $\gamma = 1$. Unfortunately, γ is unknown. Moreover, its estimation is challenging, because trading decisions of market participants can be correlated and each trading decision usually generates multiple trades. We can find out the multiplier γ, however, if we compare the conventional illiquidity ratios with the bias-free estimates of the price impact from the the portfolio transition sample.

Calibration of Illiquidity Ratios. We slightly modify the conventional illiquidity ratios so that they are measured in the same units as our bias-free estimates. In particular, the CRSP-based estimates of price impact $\bar{\lambda}^{crsp}$ are estimated from the following regression:

$$|R_i| \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}_i^e} = \bar{\lambda}^{crsp} \cdot \frac{Vol_i}{(0.01)\hat{V}_i^e} + + \tilde{\epsilon}_i. \tag{20}$$

In this regression, the observed data items have subscript i: Vol_i, R_i, V_i^e, $\hat{\sigma}_i^e$. The expression $R_i \cdot 10^4$ is a stock return in basis points. The variable Vol_i is the number of shares traded for a given stock during a trading day for the *ith* observation. The term $(0.02)/\hat{\sigma}_i^e$ with $\hat{\sigma}_i^e$ being a proxy for daily volatility adjusts for heteroscedasticity. Equation (20) is similar to equation (13) with quantities X_i traded being replaced by overall trading volume Vol_i and realized trading costs per dollar traded being replaced by daily returns. Note that we do not divide the price impact by 2 as in (13), because returns on the left-hand side reflect the total price effect of trades rather than average cost of their execution. We do not estimate the spread; the estimate $\bar{\lambda}^{crsp}$ is obtained under the assumption that the price recorded at the close is, on average, a midquote price. The estimates $\bar{\lambda}^{crsp}$ and $\bar{\lambda}^{bf}$ differ from each other only if γ differs from one,

$$\bar{\lambda}^{bf} = \sqrt{\gamma} \cdot \bar{\lambda}^{crsp}.$$

Table 3 reports the CRSP-based estimates $\bar{\lambda}^{crsp}$ of market impact based on the regression (20) as well as these estimates for the ten volume groups. It also reports the implied level of γ. For lowest deciles, the estimates $\bar{\lambda}^{crsp}$ and $\bar{\lambda}^{bf}$ are very similar, implying that only one order is executed per day in markets of the lower-volume stocks and their γ is equal to one. For higher deciles, $\bar{\lambda}^{crsp}$ are much smaller than $\bar{\lambda}^{bf}$. Their difference implies that the number of independent orders executed per day increases substantially. For the highest decile, for example, the implied value of γ is equal to 27.

The usage of the uncalibrated illiquidity ratios as proxies for illiquidity can be misleading. Figure 2 shows that the price impact functions based on the uncalibrated estimates have much lower slope than those based on the bias-free estimates. The magnitude of a price drop during the flash crash in May 2010 would seem to be surprisingly large, because the uncalibrated proxies make us significantly overestimate the liquidity available in the markets with high trading volume.

5.4 The Magnitude of Trading Costs for Mutual Funds

Many mutual fund investors can easily find information on how much funds charge as management fees and other miscellaneous fees by simply looking at the reported expense ratio at the end of the year. There are, however, other costs that are not reported in any documents. These are costs paid by mutual fund managers whenever they actually implement their strategies and trade securities. The magnitude of these implicit costs and its relation to delivered returns as well as to the observed expense ratios are the important questions.

We use our bias-free estimates to assess the magnitude of trading costs for the sample of the U.S. mutual funds from 2001 to 2005. For each mutual fund and for each month from June of year t to June of year $t+1$, we infer the expected trading costs based on its holdings and actual turnover ratio reported at the end of year t. In particular, at the end of each year, we examine what percentage of securities belongs to each of the ten volume groups and use equations (18), adjusted for the turnover, to calculate the implied trading costs. Also, following Daniel et al. (1997) and Wermers (2004), we assign the benchmark returns for each mutual fund and for each month from June through May, based on its holdings reported for that year.[1]

[1]The DGTW benchmarks are available via www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm

Table 7 reports the mean and median values of mutual funds returns, returns on their benchmarks, management fees, and implied expected costs per year. The expected costs are further decomposed into the price impact related payments and spread-related payments. The results are reported for mutual funds with different levels of turnover rates. Our general conclusion is that the magnitude of the trading costs is comparable to a magnitude of management fees. For mutual funds with high turnover ratio, the former can be even bigger than the latter.

We can also compare two models for the trading costs, one based on the bias-free estimate and another based on the uncalibrated CRSP-based estimates. While the spread-related payments are similar across both models, the price impact related payments are very different, implying more than three times larger payments if we use the bias-free estimates comparing to the uncalibrated CRSP-based ones. For example, for mutual funds with the turnover greater than 200%, the average annual returns are 11.60%. The average management fee is 1.86%. The spread related payments are similar across both models and equal to about 1%. The impact related payments are 4.14% per year if the bias-free estimates are used and only 1.29% if the uncalibrated CRSP-based estimates are employed. Of course, if the actually expenses are larger than we usually think, the implied alpha has to be larger as well in order to cover these expenses. For instance, given the benchmarks returns of 15.05% for high turnover funds, the implied alphas had to be 3.89% and 0.68% for both models, respectively.

The reported implicit trading costs seem to be higher comparing to the results in Wermers (2000), even though the latter studies these issues for more "expensive" pre-decimalization period. For the top decile of mutual funds ranked based on their turnover, for example, we estimate the average implicit costs as 5.49%, whereas Wermers (2000) reports the estimates of only 2.65% (see page 1688) based on Keim and Madhavan (1997). Thus, our analysis shows that the implicit costs, not included in the expense ratios, can be substantial in the mutual funds industry.

5.5 The Profitability of Trading Strategies and Fund Sizes

The knowledge of the trading costs is important in asset pricing for evaluation of the profitability of trading strategies and studying market efficiency. As our paper shows, the trading costs are larger than other studies have reported, mostly due to the selection bias. Our analysis shows that market participants should feel being especially market constraint, i.e. the seemingly profitable trading strategies are difficult to exploit because of significant trading costs.

We illustrate our claim using the example of the momentum-based strategies for the period from 2001 to 2005. We construct the monthly returns for the 5/1/1 momentum strategy. We filter out the securities with missing data and prices below $5. Each month, we assign the signal of one to top 10% and the signal of minus one to bottom 10% of stocks ranked according to their return over the last five month, skipping one month. The equally-weighted and value-weighted signals are then aggregated over the last six months to determine the weights of securities.

We calculate the expected trading costs using the bias-free estimates and the uncalibrated illiquidity ratios. Each month, we examine how much trading is required at the beginning of the month to turn the prior portfolio into a given portfolio as well as at the end of the month to turn it into the subsequent portfolio. Assuming a given amount of dollars invested into the long and short positions, we calculate the expected trading costs: Each securities is assigned to one of the ten volume groups and relevant trading costs are calculated from equation (18). We repeat these calculations for the uncalibrated CRSP-based estimates as well.

Figure 1 shows the average monthly return, both unadjusted and adjusted for trading costs, for different initial sizes of a hypothetical fund. For the equally-weighted strategies, the raw returns are 12.51%. After adjustment for the trading costs, the break-even points are as low as $300 million for the bias-free estimates and $800 million for the uncalibrated CRSP-based estimates. For the value-weighted strategies, the raw returns are 14.03%. The break-even points are about $3 billion and $13 billion, respectively.

As we see, only a limited amount of capital can be invested into the 5/1/1 momentum strategies, when trading costs are taken into account. Market participants are thus severely market constraint. Good investment ideas are often hard to exploit. This is true especially if we look at the bias-free estimates of trading costs. These concerns can be even more pressing, since financial institutions often magnify the size of available capital through the leverage.

5.6 The Non-Linearity of Price Impact

In most theoretical models, the linear price impact functions often represent equilibrium relation between trades and prices. For example, a linear price impact is consistent with a stylized model of Kyle (1985), in which an informed trader trades on his information with a risk-neutral market maker in the presence of a noise trader. Many other authors showed that a linear rule is a solution in their models with other structures of information asymmetry as well (e.g., Admatti and Pfleiderer (1988), Back (1992), Foster and Vishwanathan (1996)). The assumption of a linear price impact makes the theoretical analysis tractable and elegant. Huberman and Stanzl (2000) showed that only the linear price impact functions rule out a certain form of arbitrage.

Whether the price impact is linear or not, of course, remains ultimately an empirical question. Despite theoretical appeal, the financial data strongly often suggests that the price impact functions are concave. The numerous examples include Almgren et al. (2005), Coppejans, Domowitz and Madhavan (2001), Chen, Stanzl and Watanabe (2001), Hasbrouck (1991), Hausman et al. (1992), Keim an Madhavan (1996), Kempf and Korn (1999), and Lilo et al. (2003).

The interpretation of these findings is somewhat problematic, because most researchers document the non-linearity of price impact functions based on the ex post data rather than the ex ante one. The non-linearity can then be simply a result of a selection bias, if large inexpensive orders are executed at once and large expensive orders are shredded into smaller trades executed over time. We next examine the non-linearity of the price impact functions of the portfolio transition orders, which as we have discussed do not have a selection bias. We consider two specifications for the non-linear price impact functions: the power specification and the inverse quadratic specification. Both include the linear price impact as a special case.

The Power Law Specification A conventional way to model non-linear price impact functions is to use the power law functions. The magnitude of an exponent parameter indicates the convexity if this parameter is greater than 1, concavity if it is less than 1, or linearity if it is equal to one. The examples of this approach include Lillo et al. (2003), Almgren et al. (2003), among others. We consider the following specification:

$$\Delta P_{t,i} = \lambda \cdot \mathbb{I}_{BS,i} \cdot (\Delta Q_{t,i})^{z+1} + \sigma^P_{t,i} \cdot \widetilde{Z}_t. \tag{21}$$

Plugging (21) into (3) and re-scaling parameters, we derive the regression equation for the power law specification,

$$\frac{\Pi_i}{P_{0,i}\bar{X}_i} \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}^e_i} = \lambda^p \cdot \frac{1}{z^p+2} \cdot \Big(\frac{\bar{X}_i}{(0.01)\hat{V}^e_i}\Big)^{z^p+1} + \frac{1}{2}\bar{k}^p \cdot \frac{\bar{X}_{omt,i} + \bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i. \tag{22}$$

This equation is similar to equation (13). The two parameters characterize the price impact functions: the magnitude λ^p and the curvature z^p. The superscript "p" shows that these estimates are for the power specification of price impact functions. Naturally, the test for non-linearity of price impact functions is equivalent to comparing z^p to zero. The price impact is convex if $z^p > 0$, concave if $z^p < 0$ and linear if $z^p = 0$.

The Inverse Quadratic Specification. The power law specification restricts the marginal cost of infinitesimal trades to be infinite. To avoid this restrictive assumption, we also consider an inverse quadratic model with price impact specified at two levels of order sizes. We derive its functional form from the equation $x = g(y) = ay^2 + by + c$, where x stands for the order size and y stand for the price changes. For convenience, it is expressed in terms of parameters λ_0 and λ_1 using the following restrictions:

$$\begin{cases} g^{-1}(0) = 0, \\ \frac{\partial}{\partial x} g^{-1}(x)\big|_{x=x_0} = \lambda_0, \\ \frac{\partial}{\partial x} g^{-1}(x)\big|_{x=x_1} = \lambda_1. \end{cases}$$

The first equation ensures that the price impact of zero-size trade is equal to zero. The last two equations are definitions of λ_0 and λ_1. They quantify the slope of the price impact functions at x_0 and x_1, respectively. If $\lambda_0 = \lambda_1$, then this specification is identical to the linear one. The inverse quadratic price impact function can be represented as

$$\Delta P_{t,i} = \mathbb{I}_{BS,i} \cdot \lambda_0 \cdot \alpha \cdot Q_{t,i} \left(\frac{1}{2}\sqrt{1 + \beta \cdot Q_{t,i}} + \frac{1}{2} \right)^{-1} + \sigma^P_{t,i} \cdot \widetilde{Z}_t, \tag{23}$$

with parameters

$$\beta = \frac{\lambda_0^2 - \lambda_1^2}{x_1\lambda_1^2 - x_0\lambda_0^2} \qquad \text{and} \qquad \alpha = \sqrt{\frac{\lambda_1^2(x_1 - x_0)}{x_1\lambda_1^2 - x_0\lambda_0^2}}.$$

The constants x_0 and x_1 as well as the variables λ_0 and λ_1 characterize the magnitude and curvature of the price impact functions.

Plugging (23) into (3) and re-scaling parameters, we derive the regression equation for an inverse quadratic specification:

$$\frac{\Pi_i}{P_{0,i}\bar{X}_i} \cdot 10^4 \cdot \frac{(0.02)}{\hat{\sigma}_i^e} = \lambda_0^q \cdot \frac{X_{t,i}}{\hat{V}_i^e} \frac{\alpha\left(1 + 4/3\beta\frac{X_{t,i}}{\hat{V}_i^e}\right)}{\left(1+\beta\frac{X_{t,i}}{\hat{V}_i^e}\right)^{3/2} + 3/2\beta\frac{X_{t,i}}{ADV} + 1} + \frac{1}{2}\bar{k}^q \cdot \frac{\bar{X}_{omt,i} + \bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i. \quad (24)$$

where α and β are defined in (23). We arbitrarily pick x_0 and x_1 at 10 bps and 1% of the average daily volume; these assumptions are not crucial. The variables λ_0^q and λ_1^q stands for the price impact at that points. The superscript "q" shows that these estimates are for the inverse quadratic specification of the price impact functions.

Results for Non-Linear Estimations. We estimate regression equation (22) and (24) similarly to how we have estimated the linear specification (13). Our goal is to test whether $z^p = 0$ in equation (22) and whether $\lambda_0^q = \lambda_1^q$ in equation (24). These are tests of linearity.

Table 5 reports the results. For the power law specification, the curvature coefficient z^p is is statistically significant and equal to -0.46 for the total sample. Its values range from -0.46 to -0.21 across different volume groups. Thus, the price impact functions are concave and well approximated by a square root function which is often used by practitioners (Barra model). For the inverse quadratic specification, the marginal price impact λ_0 at x_0 is larger than the marginal price impact λ_1 at x_1, thus implying the statistically significant concavity of the price impact as well.

Figure 3 depicts the implied non-linear price impact functions for various volume groups. We plot the difference between the execution price and the pre-trade price as a function of trade size in percents of the average daily volume. We assume that the daily volatility σ_r^e is equal to the average volatility for a given group from Table 1. The solid line tracks the inverse quadratic specification. The dashed line tracks the power specification.

The figure shows that the power specification is indeed too restrictive: The marginal price impact of infinitesimal trades is usually not infinite, thus justifying our use of the inverse quadratic functions. Also, for most volume groups, the concavity of price impact functions is not economically significant. For the volume groups 1 to 7, for example, the price impact signature are almost linear. Only small orders in large stocks are executed at discounts relative to a linear model.

How can we interpret these results? We suggest that these patterns may be consistent with the world in which the "true" price impact functions are linear but traders have the ability to earn the spread when trading small orders. When traders have to execute small orders (as a fraction of daily volume) over a fixed period, they may postpone their trading hoping to get a more favorable execution by providing liquidity to other market participants. For instance, these traders may place limit orders on the opposite side of the book rather than market orders. When their orders are picked off, not only do they execute original orders but also gain the spread by providing liquidity to others. Certainly, these strategies are more meaningful for more actively traded securities, for which buy orders frequently alternate sell orders. In contrast, when traders have to execute large orders, they ought to trade more aggressively and their options to postpone trading are limited. This explanation is consistent with the results of Griffiths et al. (2000) who report that the implementation shortfall of small limit orders is negative for securities with high market capitalization and, consequently, traders who want to minimize overall costs might initially want to enter a buy limit order at the bid or a sell limit order at the ask.

To support our hypothesis, we superimpose the average spread from Table 1 on each subplot of Figure 3. For all volume groups except for the stocks with the largest volume, the deviations from the linear specification are certainly less then the average quoted spread. For low volume stocks, the quoted spread is much more substantial than the discount for small orders.

To summarize, most price impact functions are close to being linear. The ex post concavity of price impact functions for large stocks is consistent with the ability of market participants to earn spread when executing small orders in frequently traded securities. Our findings thus provide the empirical validation for theoretical models that often assume the linear price impact rules.

6 Conclusion

This paper examines the estimates of the trading costs, the price impact and the spread, obtained from the data on portfolio transitions. These estimates do not suffer from a selection bias problem, because the quantities traded during portfolio transitions are by and large exogenous - initial orders are not based on the short-term information and they are executed without any cancelations or modifications. This is in a sharp contrast with most other data samples, in which the quantities traded are often endogenous, since traders use price-dependent investment and trading strategies, typically canceling orders if prices run away from them. This natural endogeneity usually makes one overestimate available liquidity if he analyzes the sample of actual price changes and quantities traded.

We find that our bias-free estimates tend to be bigger than other conventional estimates, which are either distorted by the selection bias or not properly calibrated, such as illiquidity ratios. The half market impact is equal to 0.30 basis points for a trade equivalent to one percent of daily volume for a stock with a two-percent daily volatility. The half-spread is equal to 19.20 basis points. Furthermore, trading costs vary significantly across stocks. The market impact increases tenfold and spread decreases fivefold across stocks sorted by trading volume in an ascending order.

Our results have important implications. They suggest, for example, that the liquidity available in high volume markets is not as abundant as many traders think. Indeed, the flash crash in May 2010 has demonstrated that one order can lead to significant price changes, even when its size is not particularly large relative to the overall trading volume. The intuition is that even though the competition among market participants can substantially increase trading volume, it does not increase the market depth at the same rate.

Examining the actual turnover of the U.S. mutual funds, we find that the trading costs incurred by institutional fund managers are comparable to the management fees they charge and the differences between delivered returns and their benchmarks. High trading costs also imply that fund managers are substantially constraint by the market; good investment ideas can absorb only a limited amount of capital. Finally, we show how to calibrate the popular proxies for illiquidity such as the illuquidity ratios by adjusting them for the number of independent bets arriving to the market place per day. Due to numerous orders executed in the markets, these conventional measures significantly overestimate available liquidity, especially in the high volume markets.

References

Admati, Anat, and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies*, 1, 3–40.

Almgren, Robert, Chee Thum, Emmanuel Hauptmann, and Hong Li, 2005, Equity market impact, *Risk*, 57–62.

Almgren, Robert, Execution costs, 2008, *Encyclopedia of Quantitative Finance.*

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Anderson, Theodore, Kaitai Fang, and Ingram Olkin, 1994, Multivariate analysis and its applications.

Back, Kerry, 1992, Insider trading in continuous time, *Review of Financial Studies* 5(3), 387-409.

Barra, 1997, Market impact model handbook.

Breen, William, Laurie Hodrick, and Robert Korajczyk, 2002, Predicting Equity Liquidity, *Management Science*, 48, 470-483.

Chan, Louis, and Josef Lakonishok, 1993, Institutional trades and intra-day stock price behavior, *Journal of Financial Economics* 33, 173–199.

Chan, Louis, and Josef Lakonishok, 1995, The behavior of stock prices around institutional trades, *Journal of Finance* 50(4), 1147–1174.

Chen, Zhiwu, Werner Stanzl, and Masahiro Watanabe, 2001, Price impact costs and the limit of arbitrage, *ICF Working Paper* No. 00–66.

Cheng, Minder, and Ananth Madhavan, 1997, In search of liquidity: Block trades in the upstairs and downstairs markets, *Review of Financial Studies* 10, 175–203.

Chiyachantana, Chiraphol, Pankaj Jain, Christine Jiang, and Robert Wood, 2006, Price history and asymmetry in the impact of institutional trades, *Working paper.*

Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007, The cross-section of expected trading activity, *Review Financial Studies.*

Chordia, Tarun, Amit Goyal, Ronnie Sadka, Gil Sadka, and Lakshmanan Shivakumar, 2009, Liquidity and the post-earnings-announcement drift, Financial Analysts Journal 65, 18–32.

Christie, William, and Paul Schultz, 1994, Why do NASDAQ market makers avoid odd-eighth quotes?, *Journal of Finance* 49(5), 1813–1840.

Conrad Jennifer, Kevin Johnson, and Sunil Wahal, 2003, Institutional trading and alternative trading systems, *Journal of Financial Economics*, 70, 99-134.

Cooper, Michael, Roberto Gutierrez, and William Marcum, 2005, On the predictability of stock returns in real time, Journal of Business, 78, 469–499.

Coppejans, Mark T., Ian H. Domowitz, and Anath Madhavan, 2004, Resiliency in an automated auction, *ITG Working Paper*.

Daniel Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring Mutual Fund Performance with Characteristic Based Benchmarks, *Journal of Finance*, 52(3), 1935–1058.

Dufour, Alfonso and Robert Engle, 2000, Time and the price impact of a trade, *Journal of Finance*, 55(6), 2467-2498.

Foster, Douglas, and Viswanathan, S, 1996, Strategic trading when agents forecast the forecasts of others, *Journal of Finance*, 51(4), 1437–78.

French, Kenneth, William Schwert, and Robert Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics*, 19(1), 3–29.

Glosten, Lawrence, and Lawrence Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics*, 21(1), 123–142.

Griffiths, Mark, Brian Smith, Alasdair Turnbull, 2000, The costs and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65–88.

Handa, Puneet, and Robert Schwartz, 1996, Limit order trading, *Journal of Finance*, 51(5), 1835–1861.

Harris, Lawrence, and Joel Hasbrouck, 1996, Market vs. limit orders: The SuperDOT evidence on order submission strategy, *Journal of Financial and Quantitative Analysis*, 31(2), 213–231.

Hasbrouck, Joel, 1988, Trades, quotes, inventories and information, *Journal of Financial Economics* 22, 229–252.

Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46(1), 179–207.

Hasbrouck, Joel, 1991b, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4(3), 571–595.

Hausman, Jerry, Andrew Lo, and Craig MacKinley, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics*, 31, 319–379.

Holthausen, Robert, Richard Leftwich, and David Mayers, 1987, The effect of large block transactions on security prices: A cross-sectional analysis, *Journal of Financial Economics* 19(2), 237–67.

Holthausen, Robert, Richard Leftwich, and David Mayers, 1990, Large-block transactions, the speed of response, and temporary and permanent stock-price effects, *Journal of Financial Economics* 26(1), 71–95.

Huang, Roger, and Hans Stoll, 1996, Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE, *Journal of Financial Economics*, 41(3), 313–357.

Huberman, Gur, and Werner Stanzl, 2004, Price manipulation and quasi-arbitrage, *Econometrica*, 72(4), 1247–1275.

Jones, Charles, and Marc Lipson, 2001, Sixteenths: direct evidence on institutional execution costs, *Journal of Financial Economics* 59(2), 253–278.

Keim, Donald, and Ananth Madhavan, 1995, Anatomy of the trading process empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37(3), 371–398.

Keim, Donald, and Ananth Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies*, 9(1), 1–36.

Keim, Donald, and Ananth Madhavan, 1997, Transactions costs and investment style: an inter-exchange analysis of institutional equity trades, *Journal of Financial Economics* 46(3), 265–292.

Kempf, Alexander, and Olaf Korn, 1999, Market depth and order size, *Journal of Financial Markets*, 2(1), 29–48.

Kyle, Albert, 1985, Continuous auctions and insider trading, *Econometrica*, 53(6), 1315–1335.

Lesmond, David, Joseph Ogden, and Charles Trzcinka, 1999, A new estimate of transaction costs, *Review of Financial Studies*, 12(5), 113–1141.

Lillo, Fabrizio, Doyne Farmer, and Rosario Mantegna, 2003, Master curve for price-impact function, *Nature*, 421, 129–130.

Madhavan, Ananth, Matthew Richardson, Mark Roomans, 1997, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies*, 10(4), 1035–1064.

NYSE Research, 2006, NYSE market quality: Evidence from the 67 NASDAQ transferring stocks during 2002-2005.

Obizhaeva, Anna, 2009, Portfolio transitions and price dynamics, *University of Maryland, Working paper.*

Perold, Andre, 1988, The implementation shortfall: Paper vs. reality, *Journal of Portfolio Management* 14(3), 4–9.

Saar, Gideon, 2001, Price impact asymmetry of block trades: An institutional trading explanation, *Review of Financial Studies* 14, 1153–1181.

U.S. Securities and Exchanges Commission, 2001, Report on the comparison of order executions across equity market structures.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, Journal of Finance 55(4), 1655–1695.

Wermers, Russ, 2004, Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence, *Working Paper.*

Table 1: The Descriptive Statistics.

Panel A: Properties of Securities

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

Panel B: Properties of Orders

	All	1	2	3	4	5	6	7	8	9	10
Avg(X/V) (%)	3.90	15.64	4.58	2.63	1.82	1.36	1.18	1.07	0.88	0.69	0.49
Med(X/V) (%)	0.56	3.48	1.39	0.80	0.54	0.40	0.35	0.30	0.25	0.20	0.14
Avg OMT Share	0.31	0.38	0.33	0.32	0.32	0.31	0.31	0.30	0.29	0.27	0.24
Avg EC Share	0.40	0.42	0.42	0.41	0.41	0.41	0.40	0.40	0.39	0.38	0.36
Avg IC Share	0.29	0.20	0.25	0.26	0.27	0.28	0.29	0.30	0.32	0.35	0.40
# Obs	441,865	65,081	68,545	41,559	49,532	28,621	30,087	30,710	35,733	42,331	49,666

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

Table 2: The Bias-Free Estimates Across Volume Groups.

	All	j: 1	2	3	4	5	6	7	8	9	10
$1/2\bar{\lambda}_j^{bf}$	0.30***	0.25***	1.24***	1.72***	1.80***	2.04***	2.79***	2.38***	2.54***	2.08***	2.37***
	(0.05)	(0.03)	(0.17)	(0.27)	(0.50)	(0.43)	(0.56)	(0.46)	(0.63)	(0.47)	(0.57)
$1/2\bar{k}_j^{bf}$	19.20***	39.29***	17.66***	10.93***	12.01***	7.79***	6.41**	8.26***	8.14***	9.31***	7.85***
	(1.42)	(2.89)	(2.34)	(2.32)	(2.20)	(2.15)	(2.11)	(2.08)	(2.20)	(1.79)	(1.84)

Table presents the bias-free estimates of $\bar{\lambda}_j^{bf}$ and $\bar{k}_j^{bf}$ for ten volume groups j for the regression (14),

$$\frac{P_{ex,i} - P_{0,i}}{P_{0,i}} 10^4 \frac{(0.02)}{\hat{\sigma}_i^e} = \left(\sum_{j=1}^{10} \mathbb{I}_{j,i} \cdot 1/2 \bar{\lambda}_j^{bf} \right) \cdot \frac{\mathbb{I}_{BS,i} \bar{X}_i}{(0.01) \hat{V}_i^e} + \left(\sum_{j=1}^{10} \mathbb{I}_{j,i} \cdot 1/2 \bar{k}_j^{bf} \right) \cdot \frac{\mathbb{I}_{BS,i} \bar{X}_{omt,i} + \mathbb{I}_{BS,i} \bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}_i.$$

Each observation corresponds to portfolio transition order i. The left-hand side variable is the implementation shortfall in basis points, where $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i^e$ adjusts for heteroscedasticity. X_i is the number of shares in the order i with $X_{omt,i}$ and $X_{ec,i}$ shares executed in open market and external crossing networks, respectively. $\mathbb{I}_{BS,i}$ is the buy/sell indicator. $\mathbb{I}_{j,i}$ is equal to one if order i is executed in a stock from group j and zero otherwise. $\bar{\lambda}_j^{bf}/2$ estimates in basis points the market impact costs of a trade of one percent of expected daily volume V_i^e in a benchmark stock with daily volatility 2% based on data for volume group j, $\bar{k}_j^{bf}/2$ estimates in basis points the effective spread cost for volume group j. Volume groups are based on the pre-transition dollar trading volume with thresholds 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume. The standard errors (in parentheses) are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively.

Table 3: Calibration of the Illiquidity Ratios for Ten Volume Groups.

Panel A: The Bias-Free Estimates, $\bar{\lambda}^{bf}$.

	All	j: 1	2	3	4	5	6	7	8	9	10
$1/2\bar{\lambda}_j^{bf}$	0.30***	0.25***	1.24***	1.72***	1.80***	2.04***	2.79***	2.38***	2.54***	2.08***	2.37***
	(0.05)	(0.03)	(0.17)	(0.27)	(0.50)	(0.43)	(0.56)	(0.46)	(0.63)	(0.47)	(0.57)

Panel B: The CRSP-Based Estimates, $\bar{\lambda}^{crsp}$.

	All	j: 1	2	3	4	5	6	7	8	9	10
$1/2\bar{\lambda}_j^{crsp}$	0.34***	0.23***	0.33***	0.40***	0.38***	0.38***	0.40***	0.42***	0.41***	0.45***	0.46***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
$\sqrt{\gamma}$	0.88	1.05	3.70	4.34	4.76	5.32	7.00	5.65	6.23	4.58	5.17
γ	1	1	14	19	23	28	49	32	39	21	27

Table presents the bias-free estimates of $\bar{\lambda}_j^{bf}$ from the the regression (14) as well as the biases CRSP-based estimates $\Delta_{\lambda,j}^{crsp}$ for ten volume groups j for the regression (20),

$$|R_i|10^4\frac{(0.02)}{\hat{\sigma}_i^e} = \left(\sum_{j=1}^{10}\mathbb{I}_{j,i}\cdot\bar{\lambda}_j^{crsp}\right)\cdot\frac{Vol_i}{(0.01)\hat{V}_i^e} + \tilde{\epsilon}_i,$$

Each observation corresponds to portfolio transition order i. The left-hand side variable is the daily returns R_i in basis points. The term $(0.02)/\sigma_i^e$ adjusts for heteroscedasticity. $\mathbb{I}_{j,i}$ is equal to one if order i is executed in a stock from group j and zero otherwise. $\bar{\lambda}_j^{crsp}/2$ estimates in basis points the market impact costs of a trade of one percent of expected daily volume V_i^e in a benchmark stock with daily volatility 2% based on data for volume group j. Parameter γ is the adjustment factor required to convert the CRSP-based illiquidity ratios into the bias-free estimates; it is related to the implied number of independent bets executed, on average, during the day. Volume groups are based on the pre-transition dollar trading volume with thresholds 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume. The standard errors (in parentheses) are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively.

Table 4: The Bias of the RTC-based and IS-Based Estimates and Volume Groups.

Panel A: The Bias of the RTC-Based Estimates, $\bar{\Delta}^{is}_{\lambda}$ *and* $\bar{\Delta}^{is}_{k}$

	All	j: 1	2	3	4	5	6	7	8	9	10
$1/2\bar{\Delta}^{rtc}_{\lambda}$	0.09***	0.06**	0.42***	0.04	0.50	0.45	0.44	0.58	0.65	1.05*	1.27*
	(0.03)	(0.02)	(0.12)	(0.23)	(0.33)	(0.38)	(0.50)	(0.55)	(0.56)	(0.52)	(0.61)
	30%	24%	34%	2%	28%	22%	16%	24%	26%	50%	54%
$1/2\bar{k}^{bf}$	-5.96***	-11.13***	-6.61***	-2.33	-3.72**	-1.75	-2.24*	-2.05	-3.25*	-2.80***	-2.60**
	(0.77)	(1.85)	(1.30)	(1.21)	(1.22)	(1.04)	(1.03)	(1.05)	(1.36)	(0.81)	(0.84)
	-31%	-28%	-37%	-21%	-31%	-22%	-35%	-25%	-40%	-30%	-33%

Panel B: The Bias of the IS-Based Estimates, $\bar{\Delta}^{is}_{\lambda}$ *and* $\bar{\Delta}^{is}_{k}$

	All	j: 1	2	3	4	5	6	7	8	9	10
$1/2\bar{\Delta}^{is}_{\lambda}$	0.03*	0.02*	0.16	-0.25	0.17	0.19	0.20	0.12	0.12	0.55	0.26
	(0.01)	(0.01)	(0.11)	(0.18)	(0.27)	(0.26)	(0.37)	(0.30)	(0.39)	(0.41)	(0.39)
	10%	8%	13%	-15%	9%	9%	7%	5%	5%	26%	11%
$1/2\bar{\Delta}^{is}_{k}$	-5.63*	-6.32**	-5.12***	-1.96	-4.47***	-3.08*	-3.26**	-3.57**	-4.20**	-2.86**	-1.99*
	(0.87)	(2.01)	(1.42)	(1.36)	(1.35)	(1.29)	(1.07)	(1.14)	(1.43)	(0.92)	(0.91)
	-29%	-16%	-29%	-18%	-37%	-40%	-51%	-43%	-52%	-31%	-25%

Table presents the biases $\Delta^{rtc}_{\lambda,j}$ and $\Delta^{is}_{\lambda,j}$ of RTC-based and IS-based estimates for ten volume groups j. The biases $\Delta^{rtc}_{\lambda,j}$ of the RTC-based estimates $\bar{\lambda}^{rtc}_j$ are estimated from the stacked regressions (14) and (15), with constraints $\bar{\lambda}^{rtc}_j = \bar{\lambda}^{bf}_j + \Delta^{rtc}_{\lambda,j}$ and $\bar{k}^{rtc}_j = \bar{k}^{bf}_j + \Delta^{rtc}_{k,j}$. The biases $\Delta^{is}_{\lambda,j}$ of the IS-based estimates $\bar{\lambda}^{is}_j$ are estimated from the stacked regressions (14) and (16), with constraints $\bar{\lambda}^{is}_j = \bar{\lambda}^{bf}_j + \Delta^{is}_{\lambda,j}$ and $\bar{k}^{is}_j = \bar{k}^{bf}_j + \Delta^{is}_{k,j}$. $\Delta^{rtc}_{\lambda,j}/2$ and $\Delta^{rtc}_{k,j}/2$ estimate in basis points the bias of RTC-based market impact and effective spread cost for volume group j. $\Delta^{is}_{\lambda,j}/2$ and $\Delta^{is}_{k,j}/2$ estimate in basis points the bias of IS-based market impact and effective spread cost for volume group j. Volume groups are based on the pre-transition dollar trading volume with thresholds 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume. The standard errors (in parentheses) are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively.

Table 5: The Estimates of Non-linear Price Impact Functions.

Panel A: The Power Specification of Price Impact Functions.

	All	1	2	3	4	5	6	7	8	9	10
λ^p	16.59***	15.59***	11.83***	8.82**	15.84***	13.28**	13.14**	15.09**	12.14**	17.00***	18.49***
	(1.88)	(2.51)	(2.61)	(2.80)	(4.64)	(4.74)	(4.37)	(4.62)	(4.15)	(4.01)	(3.80)
z^p	-0.46***	-0.46***	-0.32***	-0.21*	-0.39***	-0.32*	-0.22	-0.32**	-0.34	-0.36***	-0.40***
	(0.02)	(0.03)	(0.06)	(0.10)	(0.11)	(0.14)	(0.14)	(0.10)	(0.19)	(0.10)	(0.09)
$1/2k^p$	0.51	9.08*	0.65	1.15	-0.87	0.39	0.33	2.22	3.08	2.95	1.98
	(2.03)	(4.02)	(3.46)	(3.61)	(3.87)	(3.57)	(3.23)	(3.48)	(3.19)	(2.94)	(2.63)
Adj. R^2	0.011	0.030	0.014	0.005	0.005	0.003	0.004	0.004	0.002	0.003	0.002

Panel B: The Inverse Quadratic Specification of Price Impact Functions.

	All	1	2	3	4	5	6	7	8	9	10
λ_0^q	15.57***	11.45***	9.43***	7.61***	14.56*	10.33*	12.88*	14.92**	14.76*	22.12*	31.57***
	(3.85)	(3.34)	(2.07)	(1.79)	(6.22)	(4.14)	(5.12)	(5.76)	(7.32)	(9.09)	(9.51)
λ_1^q	9.49***	8.20***	8.22***	7.15***	10.38***	9.07***	11.36***	11.61***	9.81***	12.73***	12.04***
	(1.01)	(1.59)	(1.41)	(1.47)	(2.12)	(2.58)	(2.69)	(2.81)	(2.32)	(1.86)	(2.85)
$1/2k^q$	3.67	13.33***	2.49	1.85	1.70	2.47	0.99	3.64	3.76	4.06	2.49
	(1.97)	(3.94)	(3.12)	(3.34)	(3.09)	(2.88)	(2.84)	(3.01)	(2.74)	(2.63)	(2.58)
Adj. R^2	0.011	0.030	0.013	0.005	0.004	0.002	0.001	0.002	0.000	0.001	0.002
p-val ($\lambda_0^q = \lambda_1^q$)	0.036	0.07	0.08	0.17	0.33	0.44	0.56	0.29	0.40	0.25	0.06
F-stat ($\lambda_0^q = \lambda_1^q$)	4.40	3.37	3.12	1.87	0.96	0.59	0.33	1.12	0.72	1.34	3.66
# Obs	421505	62827	65537	39524	47001	27096	28642	29162	33991	40327	47398

Table presents the estimates of the power parameter z^p, the price impact λ^p, and the spread $1/2k^p$ in the power specification from the regression (22) and the estimates of the price impacts λ_0^q and λ_1^q as well as the spread $1/2k^q$ in the inverse quadratic specification from the regression (24). Results are presented for stocks with different dollar trading volume. Group 1 (Group 10) contains orders in stocks with lowest (highest) dollar trading volume. Each month, the observations are split into 10 bins according to stocks' dollar trading volume in pre-transition month. The thresholds correspond to 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of dollar trading volume for common stocks listed on the NYSE. The sample ranges from January 2001 to December 2005.***, **, *denotes significance at 1%, 5% and 10% levels, respectively.

Table 6: The Bias-Free Estimates of Trading Costs.

		NYSE		NASDAQ	
	All	Buy	Sell	Buy	Sell
$1/2\bar{\lambda}^{bf}$	0.30***	0.87***	0.32***	0.87***	0.24***
	(0.05)	(0.21)	(0.07)	(0.87)	(0.03)
$1/2\bar{k}^{bf}$	19.20***	21.80***	9.11***	22.30***	22.35***
	(1.42)	(2.20)	(2.41)	(3.38)	(3.68)
$R2$	0.010	0.015	0.004	0.019	0.017
# Obs	702,406	210,194	228,934	123,841	139,437

Table presents the bias-free estimates of $\bar{\lambda}^{bf}$ and $\bar{k}^{bf}$ for the regression (13):

$$\frac{P_{ex,i} - P_{0,i}}{P_{0,i}} 10^4 \frac{(0.02)}{\hat{\sigma}_i^e} = \frac{1}{2}\bar{\lambda}^{bf} \frac{\mathbb{I}_{BS,i}\bar{X}_i}{(0.01)\hat{V}_i^e} + \frac{1}{2}\bar{k}^{bf} \frac{\mathbb{I}_{BS,i}\bar{X}_{omt,i} + \mathbb{I}_{BS,i}\bar{X}_{ec,i}}{\bar{X}_i} + \tilde{\epsilon}.$$

Each observation corresponds to portfolio transition order i. The left-hand side variable is the implementation shortfall in basis points, where $P_{ex,i}$ and $P_{0,i}$ are the execution and benchmark prices. The term $(0.02)/\sigma_i^e$ adjusts for heteroscedasticity. X_i is the number of shares in the order i with $X_{omt,i}$ and $X_{ec,i}$ shares executed in open market and external crossing networks, respectively. $\mathbb{I}_{BS,i}$ is the buy/sell indicator. $1/2\bar{\lambda}^{bf}$ estimates in basis points the market impact costs of a trade of one percent of expected daily volume V_i^e in a benchmark stock with daily volatility 2%, and $1/2\bar{k}^{bf}$ estimates in basis points the effective spread cost. The results are presented for stocks listed at the NYSE/Amex and the Nasdaq as well as for buy and sell orders. The standard errors (in parentheses) are clustered at weekly levels for 17 industries. The sample ranges from January 2001 to December 2005. ***, **, *denotes significance at 1%, 5% and 10% levels, respectively.

Table 7: The Annual Returns and Costs of Mutual Funds.

	Turnover ≤ 100% Mean / Median	100% < Turnover ≤ 200% Mean / Median	Turnover > 200% Mean / Median
	Panel A: Returns on Mutual Funds, Their Benchmarks, and Fees (in %).		
R_{MF}	12.32/15.31	13.13/18.00	11.60/16.91
R_B	13.63/17.82	14.85/20.75	15.05/21.06
Fee	1.19/1.25	1.45/1.32	1.86/1.60
	Panel B.1: Transaction Costs (in %), Uncalibrated Estimates.		
Impact	0.07/0.01	0.38/0.07	1.29/0.31
Spread	0.08/0.05	0.28/0.20	1.00/0.65
Impl. Skills	0.03/ -1.21	0.39/ -1.16	0.68/-1.59
	Panel B.2: Transaction Costs (in %), Bias-Free Estimates.		
Impact	0.21/0.03	1.05/0.29	4.14/1.29
Spread	0.11/0.08	0.37/0.29	1.35/0.88
Impl. Skills	0.20/-1.16	1.15/-0.85	3.89/-0.39
# Obs	22,716	7,987	3,784
# Funds	4,842	1,577	749

Table presents the annual returns and estimated costs for the sample of mutual funds. Panel A shows the returns of the mutual funds, R_{MF}, the returns of the benchmark portfolios, R_B, and the fees, Fee. Panel B.1 shows the expected transaction costs based on the turnover ratio of each fund and the uncalibrated estimates of market impact and bid-ask spread as well as the implied level of skills. Panel B.2 shows the expected transaction costs based on the turnover ratio of each fund and the bias-free estimates of market impact and bid-ask spread as well as the implied level of skills, calculated as the sum of the fees, the expected transaction costs, and the difference between the mutual fund returns and the benchmark returns. The mean and median values of all variables are presented in percents per annum. Three subsamples are considered: funds with the annual turnover less than 100%, funds with the annual turnover greater than 100% but less than 200%, and funds with the annual turnover greater than 200%. The sample ranges from January 2001 to December 2005.

Figure 1: The Profitability of Momentum Strategies and Fund Size.

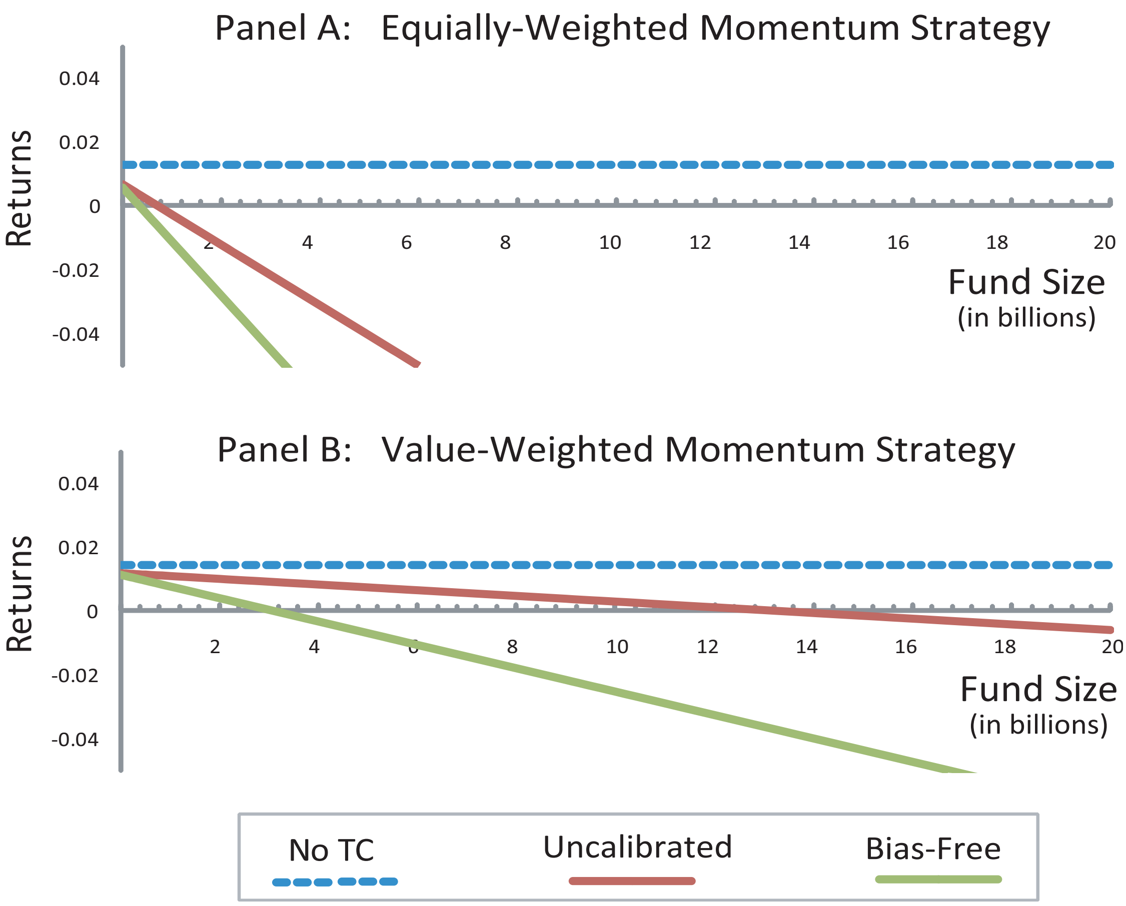

Figure shows the monthly returns on the 5/1/1 momentum strategy as a function of the fund size. For the momentum strategy, the signal of one is assigned to top 10% and the signal of minus one is assigned to bottom 10% of stocks ranked according to their return over the last five month, skipping one month. Each month, the equally-weighted and value-weighted signals aggregated over the last six months determine the weights of securities in the strategies. The dashed line corresponds to the gross monthly returns. The two solid lines corresponds to the net monthly returns, with transaction costs calculated in two ways: (1) using the bias-free estimates of market impact and spread, and (2) using their uncalibrated estimates from the CRSP dataset. The sample ranges from January 2001 to December 2005.

Figure 2: The Linear Price Impact Functions.

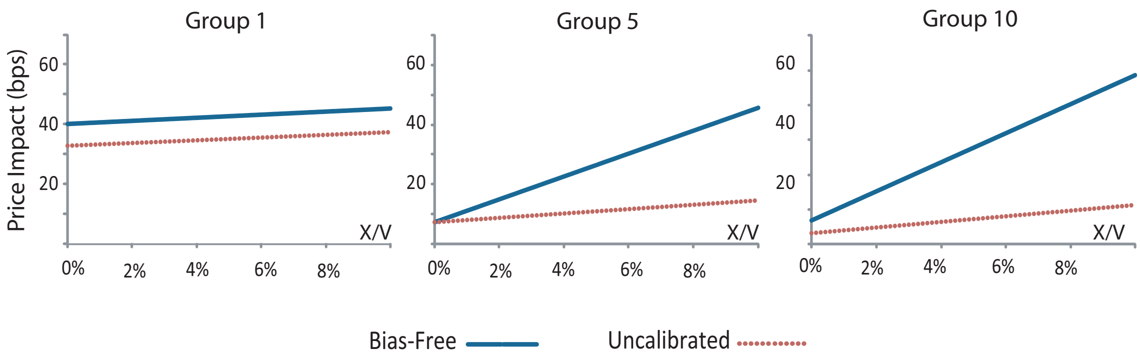

Figure shows the expected trading costs (in basis points) for volume group 1, 5, and 10. The expected cost $C_j(X)$ for volume group j is defined as,

$$C_j(X) = \frac{1}{2}\bar{\lambda}_j \cdot \frac{\sigma^e_{r,j}}{0.02} \cdot \frac{X}{(0.01)V} + \frac{1}{2}\bar{k}_j \cdot \frac{\sigma^e_{r,j}}{0.02},$$

and plotted as a linear function of a fraction of trade size X to daily volume V. The coefficient $\frac{0.02}{\sigma^e_{r,j}}$ is the adjustment for differences in volatility, where $\sigma^e_{r,j}$ is the average volatility for jth volume group from Table 1. The solid line corresponds to the estimates based on the bias-free coefficients of the market impact $\bar{\lambda}^{bf}_j$ and the spread $\bar{k}^{bf}_j$ from Table 2. The dashed line corresponds to the estimates based on the uncalibrated coefficients of the market impact $\bar{\lambda}^{crsp}$ and the spread $\bar{k}^{crsp}$ from Table 3. The thresholds for the ten volume groups are 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume.

Figure 3: The Non-Linear Price Impact Functions.

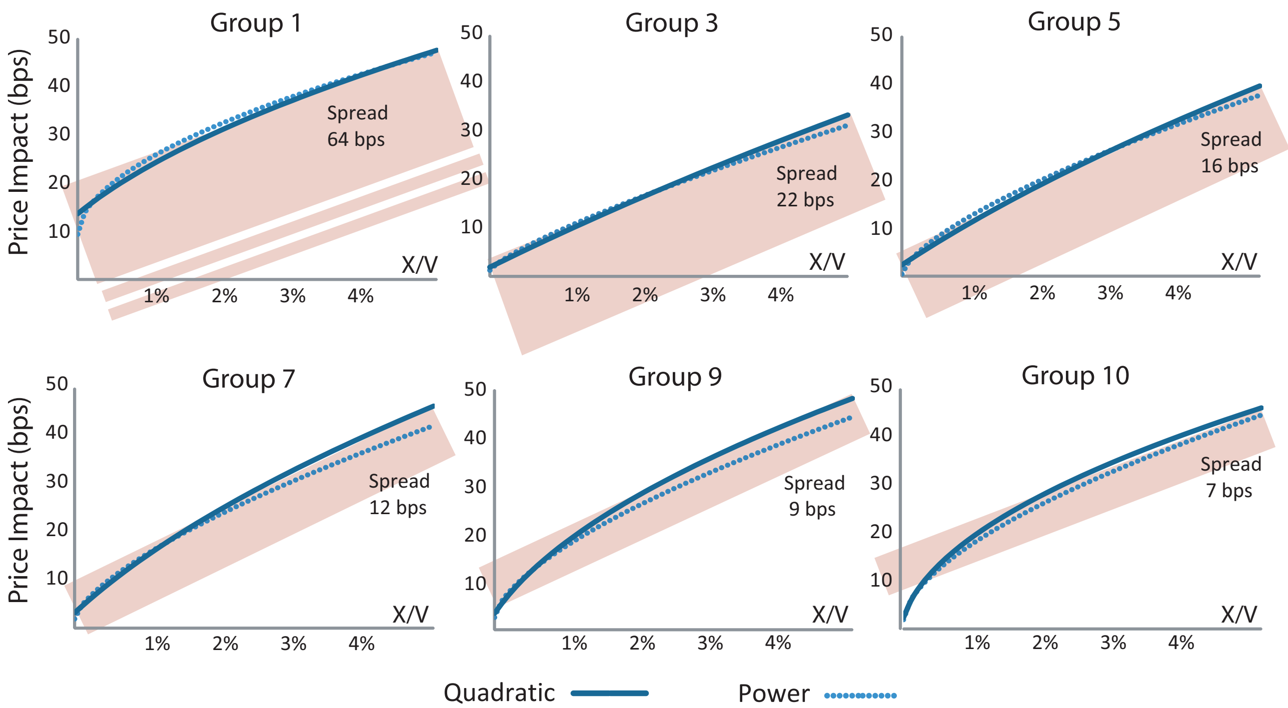

Figure shows the expected trading costs (in basis points) for volume group 1, 3, 5, 7, 9, and 10. The expected cost $C_j(X)$ for volume group j is defined as,

$$C_j(X) = \frac{1}{2}\bar{\lambda}_j \cdot \frac{\sigma^e_{r,j}}{0.02} \cdot \frac{X}{(0.01)V} + \frac{1}{2}\bar{k}_j \cdot \frac{\sigma^e_{r,j}}{0.02},$$

and plotted as a non-linear function of a fraction of trade size X to daily volume V. The coefficient $\frac{0.02}{\sigma^e_{r,j}}$ is the adjustment for differences in volatility, where $\sigma^e_{r,j}$ is the average volatility for jth volume group from Table 1. The solid line corresponds to the estimates based on the inverse quadratic specification of the price impact functions and the coefficients of the market impact $\bar{\lambda}^{bf}_j$ and the spread $\bar{k}^{bf}_j$ from Table 2. The dashed line corresponds to the estimates based on the power specification of the price impact functions and the coefficients of the market impact $\bar{\lambda}^{bf}_j$ and the spread $\bar{k}^{bf}_j$ from Table 2. The thresholds for the ten volume groups are 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, and 95th percentiles of the dollar volume for common NYSE-listed stocks. Group 1 (Group 10) contains stocks with the lowest (highest) trading volume. The shaded area shows the average quoted spread from Table 1 for each volume group.