



Estimating the Profits from Trading Strategies

Peter J. Knez, Mark J. Ready

Review of Financial Studies, Volume 9, Issue 4 (Winter, 1996), 1121-1163.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at <http://uk.jstor.org/about/terms.html>, by contacting JSTOR at jstor@mimas.ac.uk, or by calling JSTOR at 0161 275 7919 or (FAX) 0161 275 6040. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Review of Financial Studies is published by Oxford University Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://uk.jstor.org/journals/oup.html>.

Review of Financial Studies
©1996 Society for Financial Studies

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor@mimas.ac.uk.

©2001 JSTOR

Estimating the Profits from Trading Strategies

Peter J. Knez

University of Wisconsin-Madison
and Northwestern University

Mark J. Ready

University of Wisconsin-Madison

Price improvement is the difference between the execution price of an order and the quoted bid or ask when the order was submitted. We show that expected price improvement falls off dramatically as the size of the order approaches the quoted depth, and becomes negative for larger orders. This is particularly important for small firms because the quoted depths are low. Using quoted spreads and depths and our estimate of expected price improvement, we show that trading strategies that attempt to exploit the weekly predictability of small-firm returns would be swamped by transaction costs.

Lo and MacKinlay (1990) showed that the return on a portfolio of small-firm stocks is strongly correlated with its own previous week's return ($\rho = .33$) and with the previous week's return on a portfolio of large-firm stocks ($\rho = .28$). Given this predictability, we show that a trading strategy that switches between portfolios of small and large firms based on the prev-

The authors would like to thank Zhiwu Chen, Jianqing Fan, Lars Hansen, Joel Hasbrouck, Robert Korajczyk (the editor), Josef Lakonishok, Tim Mech, Walter Prah, Cyrus Ramezani, Jay Ritter, Jose Scheinkman, and an anonymous referee for their comments and suggestions. In addition, the authors would like to thank the seminar participants at the 1995 NBER Summer Institute and in workshops at the University of Illinois, Michigan State University, University of Wisconsin, and the Securities and Exchange Commission. Finally, the authors would like to thank Walter Prah and Jun Ren for their excellent research assistance. Mark Ready acknowledges financial support from the Wisconsin Alumni Research Foundation. Address correspondence and requests for reprints to Mark Ready, University of Wisconsin-Madison, 975 University Avenue, Madison, WI 53706-1323.

ious week's returns could generate excess annual returns of 15% or more if investors could always buy or sell at either the most recent trade price or the current quote midpoint. We also show, however, that effective spreads average 5% of the stock price for \$3,000 orders to buy or sell the smallest 20% of the firms traded on the NYSE. Once this effective spread is considered, we show that the realized return from the switching strategy will underperform a simple buy and hold strategy.

Lo and MacKinlay's primary purpose was to show that portfolio return autocorrelation implies cross-predictability of stocks within the portfolio and that this cross-predictability is a major source of what other researchers had labeled "contrarian" profits. They recognized that transaction costs may make these contrarian strategies unprofitable; their point was that if the profits are genuine, the mechanism generating the profits had been misunderstood. We provide a direct test of investors' ability to exploit the cross-predictability of returns by evaluating strategies that switch between portfolios of small and large firm stocks. We estimate conditional expected execution price as a function of quoted spread, quoted depth, and order size, and we use this relationship along with intraday price and quote data to estimate profits from simulated trading strategies.

The switching strategies that we investigate form equally weighted portfolios of either small- or large-firm stocks each time a switch is made. Accordingly, the returns on simple, equally weighted portfolios of the small- and large-firm stocks would provide natural benchmarks for judging the profitability of the switching strategies. In making such comparisons, however, one must realize that these equally weighted portfolios are themselves dynamic strategies requiring weekly rebalancing to maintain equal weights. As pointed out by Blume and Stambaugh (1983) and Roll (1983), bid-ask spreads and serial dependence of returns can cause rebalanced portfolio returns to be quite different from buy-and-hold returns. In addition, the rebalancing is essentially a "contrarian" strategy, selling stocks that increased in price and buying stocks that fell, so it will also benefit from cross-predictability. Assuming trades occur at the last transaction price and ignoring transaction costs, we show that the small-firm equally weighted portfolio has substantially higher returns than the buy-and-hold portfolio. Further, these returns are almost as large when transactions are assumed to occur at the current quote midpoint, indicating that cross-predictability is an important source of the difference. For the large-firm portfolio, the equally weighted and buy-and-hold returns are essentially the same. We show that the transaction costs associated with weekly rebalancing have a negligible effect on the portfolio of large firms, but they reduce the annual return on the portfolio of small firms by more

than 20%, from a 14% average annual profit to an 8% average annual loss.

In the design of our trading strategies, the lagged returns on both the small- and large-firm portfolios provide the signals of when to trade. Since small-firm portfolio returns are positively correlated with the lagged values of both small- and large-portfolio returns, and since large-firm portfolio returns are relatively unpredictable, it would be natural to consider switching investment from the large-firm portfolio to the small-firm portfolio following positive results on both and switching back following negative returns. However, it is not clear what action should be taken when the lagged returns give different signals, except that it would seem that the lagged small-firm return should be given more weight based on the higher correlation coefficient. In fact, as Boudoukh et al. (1994) point out, it could be that the small-firm return is the *only* important conditioning variable. That is, since the large- and small-firm portfolios exhibit high contemporaneous correlation ($\rho = .75$), it could be that the lagged large-firm return merely serves as a noisy proxy for the lagged small-firm return. They argue that the observed correlation structure is roughly consistent with this model, that is, $.75$ (the correlation between small and large) times $.33$ (the correlation between small and lagged small) is approximately equal to $.28$ (the correlation between small and lagged large). We use our nonparametric regression techniques to investigate the joint dependence of the small-firm portfolio returns on the lagged values of both the small- and large-firm portfolio returns. We use this joint relationship in designing our strategy, but we show that much of the information is provided by the lagged small-firm portfolio returns.

While we investigate whether the predictability of small-firm portfolio returns can be profitably exploited, we don't investigate what causes this predictability. Mech (1993) argues that this predictability can be *caused* by transaction costs. He develops a model in which some traders realize that some stocks respond slowly to market-wide information, but these traders only choose to trade when the mispricing exceeds the transaction costs. He shows that the stocks that have a large spread (relative to the volatility of their return) tend to react more slowly to innovations in a market-wide index. By providing support for his model, Mech's findings are indirect evidence that a trading strategy based on portfolio autocorrelations will not be profitable. In this article we provide direct evidence that the strategy is unprofitable. We also show that, excluding transaction costs, the estimated profits from the switching strategies are roughly the same whether we assume trades occur at the midpoint of the current quote or at the last transaction price. Since quotes are updated continuously, this provides additional confirmation of Mech's assertion that

the cross-predictability of returns is not an artifact of nonsynchronous trading.

Researchers have always recognized that transaction costs could have an important impact on investors' decisions [e.g., see Amihud and Mendelson (1986)], but it was difficult to estimate them reliably. The availability of intraday quote data has made this estimation easier, but it has not completely solved the problem because orders often execute at prices that are more favorable than the prevailing quotes. The actual execution price depends on the size and form of the order, as well as the prevailing conditions of the limit book, the trading crowd, the specialist's position, and the distribution of information among the various participants. For example, both Harris and Hasbrouck (1992) and Petersen and Fialkowski (1994) show that the average execution price of a market order is from 2 to 11 cents better than the prevailing quote point, depending on the width of the prevailing spread and the size of the incoming order. We also investigate this difference between execution price and the prevailing quote (termed "price improvement" following Petersen and Fialkowski), but we feel that our article extends the results from these previous articles in two important ways. First, we use nonparametric regression techniques to investigate the conditional expected price improvement, and we show that these techniques are superior to OLS because some of the relationships are highly nonlinear. Second, we show how these relationships impact the profitability of actual trading strategies.¹

We find that the expected price improvement for a market order depends on the quoted spread and depth and the size of the order. Although we estimate two-dimensional surfaces for each quoted spread that give expected price improvement jointly conditioned on depth and order size, we find that most of the important features of these joint relationships are captured using just the difference between quoted depth and order size. We call this difference "excess depth," and we show that the expected price improvement approaches zero as the order size approaches the quoted depth, becoming negative as the order size exceeds the depth. Interestingly, we find that the spread, depth, and order size variables capture the important features of the conditional expected price improvement. Once the depth and order size variables are "normalized" by the average quoted depth for each firm, we show that there is little additional benefit from using other conditioning variables, including firm-specific characteristics such as market value and average trading volume, or measures of re-

¹ Chan and Lakonishok (1993) investigate the execution of orders for a sample of money managers, but they lack data on when the orders were submitted.

cent market activity such as price volatility and order imbalance. Note that this does not mean that these other variables are not related to effective spreads. Rather, it means that the impacts of the variables are largely captured through their impacts on quoted spread and depth.

It should be noted that all of our results are based on strategies using only market orders. We do not consider limit orders, partly as a matter of convenience, but also because we believe that the estimation of profits from relatively short-term trading strategies is more problematic in the case of limit orders. Limit orders may be inappropriate for implementing a particular trading strategy because of the risk that the strategy's potential profit will evaporate while waiting for the limit order to be hit. In addition, Harris and Hasbrouck (1992) show that although it may be better to use limit orders in some circumstances, the difference is an order of magnitude smaller than the difference between the execution price and the prevailing quote for either type of order.

The organization of the remainder of this article is as follows. Section 1 reviews the mechanics of trading on the NYSE and describes the various data sources. In Section 2, we introduce the nonparametric estimation techniques and present the results from the estimation of conditional price improvement. Section 3 tests the profitability of trading strategies that attempt to exploit the predictability of small-firm returns and compares the realized returns on rebalanced (equally weighted) portfolios to the returns on buy-and-hold portfolios. Section 4 concludes.

1. NYSE Trading Procedures and Data

Trading on the NYSE (and the AMEX) is a continuous auction.² Each stock is assigned to a single specialist on the floor of the NYSE who is charged with the responsibility to maintain a "fair and orderly market" in the stock. In practice, this means that the specialist tries to avoid large price swings between trades and tries to quote narrow spreads and adequate depths. The specialist profits from the market-making service he provides, but faces competition in providing these services from specialists on the regional exchanges³ and from limit orders submitted to the NYSE.

Most of the customer orders submitted to the NYSE are either market orders, which demand immediate execution at the best available

² Trading at the open (and sometimes the close) is a call auction. See Hasbrouck, Sofianos, and Sosebee (1993) and Hasbrouck and Sosebee (1993) for a description of NYSE trading and reporting procedures.

³ See Lee (1992) for a discussion of intermarket competition.

price, or limit orders, which agree to execute only at the limit price or better. The specialist posts continuous quotes consisting of bid and ask prices and associated depths (the number of shares offered at the ask and the number of shares demanded at the bid). These quotes are essentially minimum guarantees of performance extended to market orders. For example, if a market sell order arrives for a quantity less than or equal to the quoted depth at the bid, the order will execute at or above the quoted bid price. Market orders frequently do better than the minimum guarantee contained in the NYSE specialist's quotes; sometimes execution prices improve on quoted prices and sometimes orders for more than the quoted depths execute at the quoted prices.⁴

The profitability of a particular trading strategy will depend not only on the quoted spread at the submission time of the order, but also on the likelihood that the order will execute at a more favorable price than is quoted. For example, if we knew that the specialist had stopped a market sell order, we would expect a market buy order would have a good chance of executing below the quoted ask, especially if the spread were greater than one-eighth. Clearly, the ability to observe all of the features of the trading floor would be a great help in estimating the probability of a price improvement. Unfortunately, most of this information is not available to traders away from the trading floor. In addition, most of the available intraday data include only the sequence of quotes and trades. Not only does this limit the ability to distinguish between different market conditions, it makes it difficult to observe the performance of orders that result in trades. Even though every market order eventually triggers a trade, it is difficult to tell submission time (it may have been stopped). It is also difficult to distinguish trades that result from market orders, and whether the market order was a buy or a sell.

The TORQ database (Trades, Orders, Reports, and Quotes) is newly available from the NYSE. It covers a period of 3 months for a sample of 144 NYSE listed firms. These firms were selected by partitioning all NYSE firms into size deciles and then randomly selecting 15 firms from each decile. Some data problems encountered after the initial selection was made caused some firms to be dropped from the final

⁴ There are at least four reasons for price improvements compared to the current NYSE specialist's quote: (1) the market order may cross with another market order, possibly because one of the market orders was "stopped"; (2) there may be a limit order in the book inside the quote points and the specialist does not reflect the order in the quotes because it is small and he prefers to quote "in size"; (3) a floor trader (or specialist) may decide to better the quoted price in order to step ahead of a limit order on the book and take the other side of the trade; (4) a better quote may be available from another exchange. Orders exceeding the depth may still execute at the quoted prices if floor traders (or the specialist) decide to step ahead of the limit orders entered at the next price tick on the book.

sample. In addition to intraday consolidated trade and quote information, the database contains information on the parties on each side of a trade and a record of all orders submitted to the floor by way of the NYSE's SuperDOT system. Importantly, the SuperDOT order information includes the time that the order was submitted, as well as when it was executed. This allows us to measure market conditions such as spread and depth at the time of the decision to submit the order. The database also allows us to unambiguously determine the order direction and to examine the overall performance of an order that results in more than one reported trade. Harris and Hasbrouck (1992) used the TORQ database to compare the execution of SuperDOT market and limit orders, whereas here, as in Petersen and Fialkowski (1994), we use the database to develop conditional estimates of price improvements for SuperDOT market orders.

The ISSM database contains intraday consolidated trade and quote information for all NYSE and AMEX listed securities covering a period of several years. However, it does not contain any orders or information on the parties to each trade. Due to the limited coverage of TORQ, it is desirable to use the ISSM database to evaluate the profitability of trading strategies. In order to accomplish this, the conditioning variables used to develop price improvement estimates for market orders must be common between the two databases. This will allow the application of estimates developed from the TORQ database to the ISSM database.⁵ Table 1 contains the information variables we consider in developing the conditional estimate of price improvement. All of the current market information variables are measured as of the time that the order was submitted, whereas the historical market data variables include the 6.5 hours (one full trading day) up to that time. Explicit definitions of the individual variables are contained in the Appendix.

In general, the particular trading strategy under consideration will specify which stocks to buy or sell, as well as when to trade and in what quantity. Thus, the conditioning variables in columns 1 and 2 of Table 1 are directly determined by the trading strategy. Strategies that attempt to exploit market information may also directly determine some of the variables in columns 3 and 4. In addition, strategies that are not explicitly based on market information may be correlated with those variables.

Ultimately, the set of conditioning variables represents a trade-off between a desire to incorporate relevant information and a need to

⁵ In evaluating our trading strategies, we use only the NYSE firms from the ISSM database to be consistent with the coverage of TORQ.

Table 1
Conditioning information

Order characteristics	Firm characteristics	Current market information	Historical market data
Direction	Market capitalization	Spread midpoint	Volume
Size	Average daily volume	Spread width	Order imbalance
Time of day		Depths at bid and ask	Transaction price volatility Percent of trades inside spread

The table lists the set of conditioning information variables considered in developing the conditional estimate of expected price improvement. We use all market orders in the TORQ database submitted during normal trading hours and measure price improvement as the difference between the average execution price for the order and the bid (or ask) price as of the time the order was entered in the SuperDOT system. The current market information variables are also based on the quote in effect when the order was entered, and the historical market data variables cover the 6.5-hour period (one full trading day) up to that time. Explicit definitions of the variables are included in the Appendix.

constrain the dimensionality of the relationships to be estimated. For example, for the historical market data variables in Table 1, we first examine aggregate statistics for a full trading day. To the extent that some of these variables are found to contain information important for price improvement, we can always run additional regressions to investigate alternative specifications. Another technique for reducing the dimensionality of the relationships is normalization by firm-specific averages. We show in Section 2 that the expected price improvement conditioned on depth and order size seems to have a similar shape across firms if we divide both conditioning variables by the firm's average quoted depth. This normalization seems to work reasonably well in our data, in the sense that we find no economically important relationships when we regress the residuals from the full sample regressions on average volume, depth, and other firm-specific characteristics. These results indicate that little additional information would be gained from estimating more general relationships using higher-order regressions.

2. Estimating Conditional Price Improvement

In this section we first present the statistical methodology. We then apply the technique to the sample of market orders in the TORQ database to nonparametrically estimate the functional relationship between the price improvement and various conditioning variables. In Section 3 we use these estimated curves and surfaces to assign execution prices to orders generated by trading strategies that attempt

to exploit small-firm predictability. Of course, there will always be a danger of omitting some economic variables that are correlated with the trading strategy and important to price improvement. While we cannot eliminate this possibility, we try to minimize it by a careful consideration of a broad set of variables.

2.1 Statistical methodology

This section describes the nonparametric regression techniques we use to estimate the mean regression function of price improvement given the conditioning variable. We choose to estimate this regression function nonparametrically for two reasons. First, there is currently little theoretical guidance as to what parametric form the regression function should take (e.g., linear or nonlinear). Second, it allows us to avoid making strong distributional assumptions about the error terms.⁶

Let $\{Y_i, X_i\}_{i=1}^n$ be a finite record of observations of price improvements and a conditioning variable. The mean regression function of Y_i on X_i is given by

$$Y_i = m(X_i) + \varepsilon_i \quad i = 1, \dots, n,$$

where m is the unknown regression function such that $m(x) = E[Y | X = x]$ and ε_i is the observation error. The mission of nonparametric regression analysis is to find an approximation to the unknown response function m that captures the crucial features of the mean dependence of Y on X . The general approach of nonparametric regression is to estimate the mean response of Y for a given value of X by computing a *local* average of the points in a small neighborhood around the given value of X . Formally, this local averaging or smoothing can be defined as

$$\hat{m}(x_o) = \sum_{i=1}^n w_i(x_o; X_i, \dots, X_n) \cdot Y_i,$$

where $\{w_i(\cdot)\}_{i=1}^n$ denotes a sequence of weights that depends on the entire sequence of X 's and x_o is an arbitrary grid point. Typically the shape of the weight function w_i is described by a density function. This density function is adjusted by a *scale* parameter that controls the size and the specific form of the weights near a given X_i . The function is referred to as a *kernel* since it is continuous, bounded, symmetric, and integrates to one.

⁶ The drawback of this class of estimators is that they often require a large number of observations and can be computationally intensive. However, the market microstructure data sets are in general very large so that the requisite number of observations is available.

We consider two closely related and complimentary nonparametric regression techniques. The first is a variable span local linear smoother known as *loess* [see Cleveland (1979, 1994)]; the second is a variable bandwidth local linear smoother [see Fan and Gijbels (1992, 1994)].⁷ These techniques share two very important features. First, they are both calculated using a linear (or higher order) weighted least squares estimate for a neighborhood around x_o . This reduces the bias that can result from using only the local mean if most of the points in the neighborhood are on one side of x_o , which is an important problem at the boundaries of the support of the data.⁸ The second property shared by these techniques is that the size of the neighborhood is increased around grid points in the relatively sparse regions of the support of the data.

The loess smoother is defined by

$$\hat{m}(x_o) = \hat{\alpha}(x_o) + \hat{\beta}(x_o)x_o,$$

where $\hat{\alpha}(x_o)$ and $\hat{\beta}(x_o)$ are the weighted least squares estimates for the data contained in the neighborhood $N(x_o)$. That is, $\hat{\alpha}$ and $\hat{\beta}$ are found by minimizing

$$\sum_{i=1}^n w_i(x_o)(Y_i - \alpha - \beta x_i)^2.$$

The weights are assigned to each point in $N(x_o)$ using a *tricube* kernel defined by

$$W\left(\frac{|x_o - x_i|}{\Delta(x_o)}\right),$$

where $W(u)$ is given by $(1 - u^3)^3$ if u is $0 \leq u < 1$ and 1 otherwise, and $\Delta(x_o)$ is the size of the neighborhood, which is determined by the *spanning* parameter. This spanning parameter is defined as the proportion of points in the sample to be included in each neighborhood.

Careful choice of the spanning parameter is important because as this parameter is increased the variance of the estimate is reduced but the bias of the estimate is increased. The idea is that since at a given grid point we are estimating the conditional mean by *averaging* the response variables associated with the predictor variables in a

⁷ Loess is sometimes referred to as a "locally weighted running line smoother." We use terminology similar to the Fan and Gijbels smoother to emphasize the close similarity between the two and to help focus the attention on the key difference, which is the way the bandwidth is selected.

⁸ See Prahl (1993) for a discussion of boundary effects in the context of a Gasser-Muller estimator, which is an alternative procedure to the ones described here.

given neighborhood, as we increase the span we increase both the number of observations in the neighborhood and the relative weights placed on the observations further from the grid point. The added observations potentially decrease the variance, but since we are now averaging response variables that are farther away from grid point, we are potentially increasing the bias. Although there are objective ways of choosing the span, such as “leave one out” cross-validation, these methods are computationally infeasible in large data sets, because for n observations they require n smooth calculations for *each* trial span.

The spanning parameter for the loess smoother is chosen by trial and error. This process begins with a smooth based on a trial value for the parameter. The residuals from this smooth are then smoothed on the same conditioning variable. If there is any structure detected in the residuals as a function of the conditioning variable, then the trial spanning parameter was too large. The curve is then reestimated with a smaller span and the process is repeated. If there appears to be no structure in the residuals, then a larger span is tried in order to reduce the variance of the estimate. The “optimal” span is the largest value that leaves no economically important structure in the residuals.

The variable bandwidth local linear smoother differs from the loess smoother in that the size of each neighborhood (also called the bandwidth) is chosen directly, rather than being indirectly determined by a single spanning parameter. The main advantage of the variable bandwidth local linear smoother is that it includes a computationally efficient algorithm for optimally selecting these bandwidths at each grid point. This algorithm minimizes the *integrated mean square error* over all of the grid points, where the mean square error is given by

$$M\hat{S}E(x_o; b) = \hat{b}^2(x_o) + \hat{V}(x_o),$$

where $\hat{b}(x_o)$ estimates the bias and $\hat{V}(x_o)$ the variance. The details of the variable bandwidth local linear smoother are included in Appendix B.

We use the loess nonparametric regression technique as our primary tool for analyzing the relationships in the following sections because it has several advantages over the variable bandwidth local linear smoother. It is less sensitive to clustered data, is easily extended to multiple conditioning variables, and allows for the use of robust techniques for attenuating the effect of outliers and for calculating standard errors in the presence of non-Gaussian residuals.⁹ The dis-

⁹ The implementation of the robust estimator begins with the standard weighted least squares estimate at each grid point, calculates the residuals using this estimate, and then reweights the observations, reducing the weight on those with large residuals. For a complete discussion see Chambers and Hastie (1993).

advantage of the loess technique is that the spanning parameter is chosen by examination of the residuals as opposed to the objective procedure based on the minimization of integrated mean square error. Accordingly, we use the variable bandwidth local linear smoother as a check of the univariate loess regressions to make sure we are not choosing an unreasonable span. In particular, this guards against possible oversmoothing, which might cause us to miss important features in the data.

With two exceptions we used a spanning parameter of .25 because it produced loess curves with similar features to those produced by the variable bandwidth local linear smoother and left little discernable pattern in the residuals. One exception is in the top half of Figure 1 in Section 2.2, where the firm (ASARCO) has relatively few observations. In this case we found that a spanning parameter of .5 was more appropriate. The second exception is the residual regressions reported in Table 3 in Section 2.3, where the independent variables are highly clustered. These regressions include a regression of residual price improvement on firm size, in which the independent variable has just 144 distinct values. With this high degree of clustering, we found that loess curves estimated using the .25 spanning parameter were very erratic and the variable bandwidth curves were even more so. Accordingly, in these regressions we also used a spanning parameter of .5.

2.2 Estimates of price improvement

In this section, we present the results of our nonparametric regressions of price improvement on various conditioning variables. Recall that price improvement is defined as the ask quote minus the trade price for a buy order, and as the trade price minus the bid quote for a sell order. Table 2 provides summary statistics for price improvement segregated by the size of the quoted spread for the entire TORQ sample of market orders.

Several features of the data deserve special note. The right-most column shows that most of the market orders in the sample were submitted when the spread was either 1/8 or 1/4. Price improvement is possible when the spread is 1/8, because orders are sometimes crossed at the opposite quote. The table shows that for 12% of the orders where the spread was 1/8, a market buy order executed at the bid or a market sell order executed at the ask, and these cases more than explain the average price improvement (12% of \$.125 is \$.015). When the spread was 1/4, more than half of the market orders executed at the midpoint, whereas just over one-third executed at the quote (at the ask for a buy or the bid for a sell). These midpoint orders explain the bulk of the average price improvement when the

spread is $1/4$ (52.6% of \$.125 is \$.066). Since price improvement is available to both buy and sell orders, the average effective spread is equal to the quoted spread less twice the average price improvement. The second to last column on the table shows that when the quoted spread is $1/8$, the average effective spread is 79% of the quoted spread (\$.099), whereas when the quoted spread is $1/4$, the average effective spread is only about 42% of the quoted spread (\$.105). As pointed out by Petersen and Fialkowski (1994), it appears that there is little difference in the absolute size of the average effective spread when comparing quoted spreads of $1/8$ and $1/4$. Interestingly, this effect does not continue for quoted spreads above $1/4$.

Although the average effective spreads are of similar magnitudes for quoted spreads of $1/8$ and $1/4$, our goal is to estimate conditional expected price improvement, which is evidently quite different. In addition, price improvement for $1/8$ spreads is primarily due to trades at the opposite side of the spread, whereas for $1/4$ spreads it is primarily due to midpoint trades, so the combinations of events that result in the price improvements may be fundamentally different. Accordingly, we feel it is important to estimate the conditional relationships separately. We also estimate separate relationships for spreads below $1/8$ and for spreads above $1/4$.

Our goal in investigating different conditioning variables is to come up with a parsimonious list that captures the economically important variation in the conditional expected price improvement. We have already decided to condition on spread size (by running separate smoothes for four different sizes), but the obvious question is where to begin the search for the remaining variables. Although it is possible that many, if not all, of the potential conditioning variables contain some information when considered individually, we would hope that certain of the variables would proxy for some of the information contained in others. For example, it would seem natural that the rate of information arrival could affect effective spreads, because market makers may become more anxious about trading with informed traders. However, if this anxiousness shows up in lower quoted depths, then if we condition on depth we may effectively include the information arrival effect, making it unnecessary to include other variables such as price volatility.

After examining scatter plots showing price improvement versus several conditioning variables for a collection of individual firms, we noticed that there is a strong positive relationship between price improvement and quoted depth and a strong negative relationship between price improvement and order size. These relationships are not surprising given the previous theoretical results of Dupont (1995), Easley and O'Hara (1987), and Kyle (1985), and the empirical results

Table 2
Distribution of price improvements

Quoted spread size	Mean price improvement	Price improvement frequencies (%)					Effective/quoted spread (%)	Number of orders
		-1/8	0	+1/8	+1/4	Multiple price Negative Positive		
1/8	.013	1.5	85.0	12.0	0.2	0.6	0.4	197,065
1/4	.073	1.1	39.9	52.6	3.2	0.6	1.9	117,227
> 1/4	.095	1.2	31.5	46.7	13.3	1.3	1.7	19,981
< 1/8	.011							1,831
								<u>336,106</u>

The table provides summary statistics for price improvements for market orders, segregated by the size of the quoted spread in effect when the order was submitted. The mean price improvement is in dollars per share. Multiple price executions occur because the order was split into multiple trades, frequently because the order was larger than the quoted depth. The effective spread is the quoted spread less twice the average price improvement. The < 1/8 category includes 1,057 orders where the spread size was $\geq 1/8$, but the stock price was less than \$2. For these orders, executions can occur in 1/16 increments. No frequencies are reported for the < 1/8 category because the minimum tick sizes vary from 1/32 to 1/8.

of Lee, Mucklow, and Ready (1993). Summarizing the ideas from these articles, larger trades are more likely to be information based, so the specialist and other market participants would be less likely to step ahead of the posted quotes in order to take the other side. In addition, quoted depths are likely to be low when the specialist and other market participants sense an increased risk of informed trades, which would also imply that they would compete less aggressively for order flow.

Although the directions of the individual relationships are as expected, extant theoretical and empirical results give little guidance as to the likely form of the joint dependence of price improvement on both depth and order size. In examining the results for several firms in our sample, we noticed that most of the important features of these joint relationships could be captured in a univariate relationship using the *difference* between quoted depth and order size as the independent variable. We call this difference *excess depth*, and we found that price improvement drops off dramatically as the excess depth approaches zero and becomes negative, that is, as the order size becomes larger than the quoted depth. Based on these results, we chose depth and order size (along with spread) as conditioning variables. The benefit from adding additional variables is then assessed by smoothing them against the residuals from the regression of price improvement on depth and order size.

The nonparametric methodology we employ lends itself to a graphical presentation of the results. However, it is infeasible to present the results for all 144 firms in the TORQ database. Figure 1 shows four regressions of price improvement on excess depth for two “typical” firms, ASARCO and Boeing. The two regressions on the left side of the figure are for 1/8 spreads, and those on the right are for 1/4 spreads. Based on total number of orders, Boeing is fifth largest in our sample, and ASARCO is thirty-fourth. As shown in Table 2, there were a total of 197,065 orders at 1/8 spreads and 117,227 orders at 1/4 spreads for the 144 firms in our sample. This works out to roughly 1,000 per firm for each spread size, but there are a few very actively traded firms in the sample, so the medians of the orders per firm at 1/8 and 1/4 spreads are both about 200. The fact that ASARCO and Boeing have roughly equal numbers of orders at 1/8 and 1/4 spreads is not unusual. The relative proportions of orders at 1/8 and 1/4 spreads lies between 25%/75% and 75%/25% for two-thirds of the firms in the sample.

The solid lines in each figure are the loess estimates of the conditional mean dependence of price improvement on excess depth. The lines with short dashes are plotted two standard errors above and below the estimated curves. Due to the high degree of clustering of

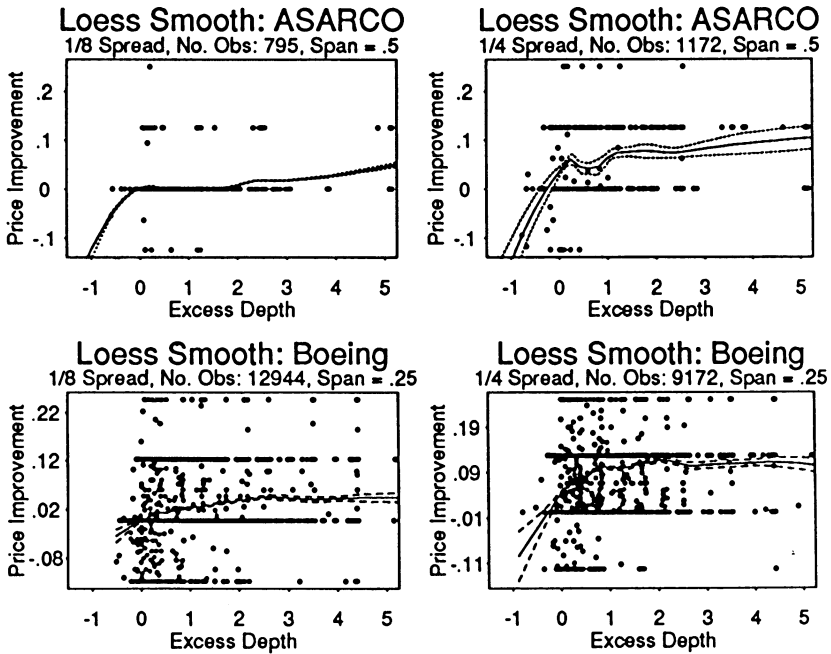


Figure 1
Loess smooths of price improvement regressed on excess depth—ASARCO and Boeing
 Price improvement is measured in dollars per share. Excess depth is the difference between quoted depth and the size of the order, where both are expressed as a fraction of the average quoted depth for the firm. The points represent individual market orders, the solid curves are loess smooths, and the dashed curves are two standard errors from the smooths.

price improvements at even eighths, the residuals from the regression are clearly not Gaussian. Accordingly, the curves and standard errors are calculated using the robust technique described in Section 2.1. The smooths estimated without the robust adjustment are very similar, with the only discernable differences near the endpoints. The standard error calculations assume independence of the residuals, so if there is important time-series dependence in price improvement, these standard errors may be understated. The regressions are shown for excess depths ranging from -1 to 5 times the average quoted depth for the firm (more on this below). In all four cases, there were orders where the excess depth was outside this range, but in no case did these orders constitute more than 1% of the orders observed for that spread size for that the firm. Although the smooths were estimated over the entire span of the data, graphing them over a wider range makes it difficult to see the details of the relationship around zero, where most of the orders are concentrated.

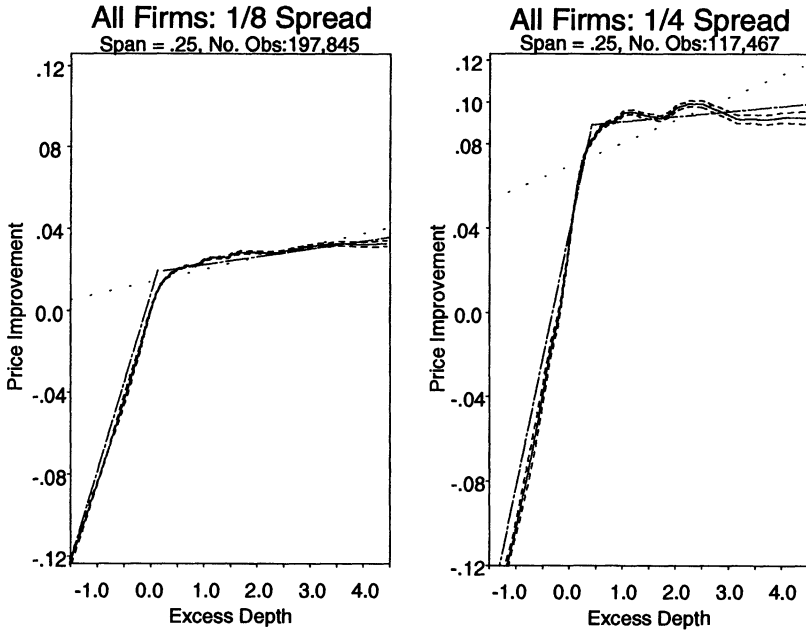


Figure 2
Loess smooths of price improvement regressed on excess depth—all firms
 Expected price improvement is measured in dollars per share. Excess depth is the difference between quoted depth and the size of the order, where both are expressed as a fraction of the average quoted depth for the firm. The dotted lines are linear OLS estimates, and the lines with alternating long and short dashes are least-squares piecewise-linear estimates. The solid curves are loess smooths, and the dashed curves are two standard errors from the smooths.

For each regression in Figure 1, the conditional expected price improvement is a concave function of excess depth that becomes zero at approximately zero excess depth (the point where the incoming order is equal to the specialist's quoted depth). We observe these same features for most of the firms in our sample, which gives hope that the relationship can be estimated for the aggregate sample. Consistency in concavity and intercept, however, is not enough. To see the problem, consider the following example. For firm A, the average quoted depth at the ask is 1,000 shares, the current depth is 2,000 shares, and a market order arrives to purchase 6,000 shares. For firm B, the average quoted depth at the ask is 20,000 shares, the current depth is 16,000 shares, and a market order arrives to purchase 20,000 shares. The excess depth is $-4,000$ shares in both cases, but the order is clearly much more unusual for firm A. To adjust for this effect, we normalize all order sizes, depths, and excess depths by the average quoted depth for the firm. In the example, this would mean that the order for 6,000 shares of firm A represents a normalized excess depth of -4 , and

the order for 20,000 shares of firm B represents a normalized excess depth of -0.2 . We check whether this normalization is appropriate by regressing the residuals from the regressions of price improvement on depth and order size on the average depth for the firm.

Figure 2 shows the results of the regressions for all orders at $1/8$ spread and all orders at $1/4$ spread, where the excess depths have been normalized as described above. As in Figure 1, the dashed curves are two standard errors above and below the smooths.¹⁰ There are three important features illustrated by the graphs in Figure 2. First, the general shape of the smooths for orders at $1/8$ spread and orders at $1/4$ spread are similar, both showing a zero expected price improvement for orders equal in size to the quoted depth. Note that even if the normalization used is imperfect, it has little effect for orders with excess depth near zero. Second, as we discuss more fully below, the shape of the curves exhibits substantial nonlinearity. Finally, the impact of excess depth on price improvement is economically important. For orders at $1/4$ spread, the conditional mean price improvement ranges from roughly 8 cents for (normalized) excess depth greater than 1 to roughly -9 cents for orders with excess depth of -1 . Thus, in any application where the size of the spread is important, the size of expected price improvement and its relationship to excess depth is also likely to be important.

The nonlinearity of the relationships in Figure 2 bear further discussion because it is at the heart of the question of whether the non-parametric regression techniques are an important improvement over simpler approaches, particularly OLS. We have added additional lines to Figure 2 depicting two sets of least-squares estimates. The dotted lines show simple linear OLS estimates of the relationship between price improvement and excess depth. The data are concentrated in the region with positive excess depths, so the OLS estimates primarily reflect this region (where the relationships are fairly flat). Consequently, the OLS relationships incorrectly predict positive price improvement for large orders. Estimating a dummy variable for negative excess depth (not shown), as in Petersen and Fialkowski (1994), does result in negative estimates for price improvement in this region, but this approach can't capture the drop in price improvement as excess depth approaches zero (as the order size approaches the excess depth). If both linear and dummy variable specifications were included in the

¹⁰ The regression curves and standard errors were calculated using the same robust technique as in Figure 1 (see also Section 2.1). Unlike the single firm regressions in Figure 1 where there were some slight differences, the smooths estimated without the robust adjustments are virtually identical to those shown in Figure 2. Since the robust reweighting seems to have little or not effect in full sample regressions and it is more complex, it is not used in the remainder of this article.

same regression, the experienced econometrician would likely recognize that the relationship has important nonlinearities and would move on to examining other functional forms. It is possible that a well-executed search would ultimately lead to estimated relationships similar to those shown in Figure 2. To reach this point would take some care, however, because while the region with negative excess depths is clearly economically interesting, it contains relatively few data points. Consequently, estimated parametric relationships could fit poorly in this region and still have reasonable mean squared error. The advantage of the nonparametric regressions is they provide an immediate picture of the relationship that makes the nonlinearity obvious. Although their implementation is more complex than linear OLS, it is considerably simpler than the search for arbitrary parametric functional forms.

After seeing the shape of the smooths in Figure 2, it appears that the relationship between price improvement and excess depth could be closely approximated by piecewise linear functions with the knot points located where the smooths bend abruptly downward. Of course, it is important to remember that this observation is made after seeing the nonparametric regressions, so it can't be an a priori argument for using only this technique. Once the smooths have been estimated, however, it might be desirable to have characterizations of the relationships that are more convenient than the list of values at grid points that defines each smooth. The lines in Figure 3 with alternating long and short dashes show least-squares estimates of piece-wise linear functions, where the locations of the knot points are estimated along with the slopes and intercepts. These functions follow the smooths quite closely. For both 1/8 and 1/4 spreads, the lines to the right of the knot points are quite flat, and the knot points both occur at positive excess depths (excess depth of .24 for 1/8 spreads and .42 for 1/4 spreads). The slopes of the lines to the left of the knot points are .09 for 1/8 spreads and .12 for 1/4 spreads.

In order to check whether the excess depth variable fully captures all of the interactions between depth and order size in the joint determination of expected price improvement, we examine loess surfaces that jointly condition on depth and order size. Figure 3 shows contour plots of these surfaces for all orders at 1/8 and 1/4 spreads. To allow aggregation across all firms, we continue to normalize depth and order size by dividing by the mean quoted depth for each firm. The contour plots are shown over the range of order sizes and quoted depths up to three times the average quoted depth. This region was selected because it contains over 95% of the orders used to estimate the surface for each smooth. Note that if excess depth perfectly characterized these joint relationships, all of the contours would be straight

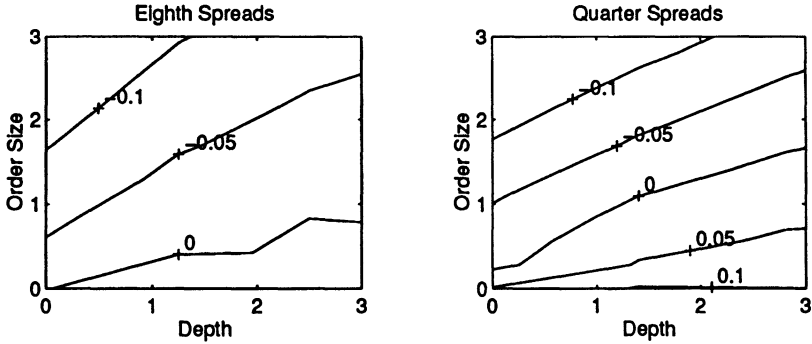


Figure 3
Contour plots of loess surfaces—price improvement regressed on depth and order size
 The curves show constant contours of the expected price improvement surfaces, measured in dollars per share, estimated using a two-dimensional loess regression with a spanning parameter of .25. The surfaces are estimated over the entire range of the data, but the plots are restricted to the region where depth and order size are less than three times the average depth for the firm. In each case, this region contains more than 95% of the observations.

lines with 45-degree slopes. This characterization does seem to be a fair description of the surfaces, confirming that excess depth appears to provide a good one-dimensional parameterization of these relationships. However, the contour lines appear to be flatter than 45 degrees in the lower halves of the figures. This seems to indicate that for very small orders, there is a limit to the amount of price improvement. That is, the expected price improvement for a very small order doesn't increase very much with larger and larger depth. This feature of the relationship is not captured in a one-dimensional excess depth smooth. Accordingly, we use the surfaces shown in Figure 3 to estimate the profits from trading strategies in the next section.¹¹

2.3 Other conditioning variables

Although the above results suggest depth and order size are important predictor variables, it remains to be seen if other variables such as volume are important, given the functional relationship already captured by spread, depth, and order size. To assess the potential importance of the other conditioning variables shown in Table 1, we regressed the residuals from the surfaces shown in Figure 3 on each variable. Table 3

¹¹ Orders at spreads of more than 1/4 were grouped together and price improvement was measured as a percent of spread. A single surface was estimated for these orders. The same procedure was used for orders at spreads of less than 1/8, and/or when the stock price was less than or equal to \$2.

provides summary statistics from these regressions.¹² The maximum and minimum columns reflect the extremes of the conditional estimated residual price improvement over the range of the conditioning variable that excluded the top and bottom 1%. We make this exclusion because the regression estimates in these extreme regions are very noisy due to sparse data. The minimum and maximum statistics give some indication of the potential additional improvement in fit that could be obtained by adding the variable to the estimation. The columns labeled “Slope” give the difference in the simple mean of the residuals for the top and bottom decile of the conditioning variables. This statistic gives a rough indication of whether there is a monotonic relationship between price improvement and the conditioning variable (after controlling for spread, depth, and order size). Monte Carlo estimates of these parameters obtained by randomly assigning the residuals to the values of the conditioning variable indicate that almost all of the values in the table are statistically significant at traditional levels. This isn’t too surprising, given the extremely large sample size. Accordingly, the issue is whether the regressions in Figure 3 have captured most of the economically interesting effects.

In the next section, we use only spread, depth, and order size to estimate price improvements. Table 3 seems to indicate that by omitting the remaining variables we are missing somewhat larger price improvements associated with large firms (which will also have high volume, high price, and high quoted depths) and larger price improvements for sell orders. Virtually any strategy will ultimately make equal numbers of buy and sell orders, so the difference in price improvements between buys and sells should cancel out. Also, the trading strategies evaluated in the next section switch the entire portfolio investment back and forth between small and large firms, so we will underestimate the costs of trading in small firms and overestimate the cost of trading in large firms by similar amounts. It should be noted, however, that it could be important to include some of these other variables when evaluating strategies that concentrate trading in either small or large firms.

3. The Predictability of Small-Firm Returns

In this section we use the fitted relationship developed in the last section to estimate the profit from a trading strategy that attempts to exploit the predictability of small-firm returns using both lagged small-

¹² These regressions use a spanning parameter of .5, as opposed to the .25 used in the depth and order size regressions, because several of the conditioning variables, particularly firm size and mean depth, have a high degree of clustering.

Table 3
Summary statistics from residual regressions

Conditioning variable	1/8 spreads			1/4 spreads		
	Minimum	Maximum	Slope	Minimum	Maximum	Slope
Mean depth	-.009	.010	.004	-.016	.018	.028
Firm size	-.009	.016	.022	-.008	.018	.026
Share price	-.008	.026	.024	-.013	.024	.022
Volume	-.007	.026	.023	-.013	.015	.023
Order imbalance	-.004	.009	-.003	-.007	.024	-.011
Volatility	-.012	.017	.021	-.010	.018	.020
Inside spread (%)	-.009	.009	.004	-.023	.016	.029
Order direction	-.002	.002	.004	-.006	.006	.012

This table shows the results of nonparametric regressions of the residuals from the bivariate surfaces shown in Figure 3 on other conditioning variables. Minimum and maximum are the minimum and maximum values of the estimated conditional expected residual price improvement. These statistics are measured over the region of the conditioning variable that excludes the 1% largest and smallest values. The slope statistic is the difference between the mean residuals for the 10% largest and 10% smallest values of the conditioning variable. Mean price improvements are larger for sells than for buys. Accordingly, for the last line in the table, the minimum columns show the mean residual for buys, the maximum columns show the mean for sells, and the slope column shows the difference.

firm returns and lagged large-firm returns. We use a nonparametric regression to show how the expected return on small firms depends on various combinations of lagged small- and large-firm returns.¹³ This leads to the specification of the particular trading rules that we test. The returns from these strategies must be compared to some benchmark, and this leads to the investigation of the realized returns from equally weighted and buy-and-hold portfolios.

It will be necessary to apply the strategies to a sample of ISSM data, which are available for 1988 to 1992. Accordingly, the first step is to calculate small- and large-firm portfolio returns using most recent transaction prices for the 1988 to 1992 period and then to show that this data set exhibits cross- and autocorrelations similar to those found by Lo and MacKinlay (1990). The first column of Table 4 reproduces the correlation coefficients found by Lo and MacKinlay for weekly returns from 1962 to 1987. They measured returns from Wednesday to Wednesday and selected all CRSP NYSE, and AMEX firms with no missing weekly returns for the entire sample period. These firms were then divided into five size portfolios, based on their market capital-

¹³ Note that the cross- and autocorrelations merely suggest the existence of profitable strategies. To find them, we must see which combinations of lagged small- and large-firm returns lead to unusually high (or low) current small-firm returns.

Table 4
Correlations between small- and large-firm portfolio returns

	1962–1987		1988–1992		
	Lo and MacKinlay	NYSE only/	CRSP	ISSM	
	results	overlapping		Trades	Quotes
Contemporaneous	.75	.80	.55	.56	.58
Autocorrelations					
Small	.33	.27	.39	.40	.43
Large	.04	.06	-.09	-.08	-.08
Cross-autocorrelations					
Small to lag large	.28	.25	.21	.22	.22
Large to lag small	.02	.02	-.04	-.04	-.04
Individual return autocorrelations					
Small	-.079	-.043	-.070	-.074	-.046
Large	-.013	-.034	-.078	-.076	-.070

The table shows simple sample correlation coefficients for weekly returns on a portfolio comprised of large firms and a portfolio comprised of small firms. The last two rows of the table show the average autocorrelations for the individual firms contained in each portfolio. The first column reproduces the results from Lo and MacKinlay (1990), which are based on CRSP closing prices. The second column shows the results for the same time period as in Lo and MacKinlay, also using CRSP closing prices, but using only NYSE firms, using overlapping observations, and using only the information available at the time of portfolio formation. The final three columns all use our portfolio formation methodology, and compare the results from calculating returns using CRSP closing prices, the last trade as of 11:00 A.M., and the current quote as of 11:00 A.M.

ization in the middle of the period. The table shows the results for the smallest and largest of the five size portfolios. The final two entries in the first column are the averages for the individual firm weekly return autocorrelations for the smallest and largest size quintiles of NYSE and AMEX stocks that had at least 52 nonmissing weekly returns.

Lo and MacKinlay's (1990) data selection procedures introduce "look ahead" biases because firms with return data for the entire sample period neither went bankrupt nor were acquired. Also, firms that are in the smallest market capitalization quintile as of the middle of the sample period are more likely to have had poor performance up to that point, whereas the opposite is true for the firms in the largest quintile. Lo and MacKinlay investigated the potential effect of this survivorship bias by splitting the sample period in half and performing the analysis on all firms that existed for each subperiod. Based on the fact that the correlation and autocorrelation results for these subperiods were very similar to those for the full sample, they concluded that this bias did not have an important effect. In this article we are focusing primarily on profits from trading strategies, so we felt that it was important to construct a data set that is free of these biases.

We simulate the profits from our trading strategies over the period from 1988 through 1992. We measure firm market capitalization as of June 30 of each year (from 1987 to 1992). NYSE-listed firms with June 30 market capitalization above (below) the top (bottom) quintile breakpoints for all NYSE-listed firms, except ADRs, are included in our sample for the following 12 months (6 months for the first half of 1988 and six months for the second half of 1992).¹⁴ For the firms that were delisted, we use the delisting value from CRSP to calculate the final return and liquidation value. If this value is missing we use the price immediately before delisting, effectively assuming the security is sold on the last day of NYSE trading. We use only NYSE firms to be consistent with the coverage of the TORQ database.

The CRSP database was used to select the firms and obtain closing price, dividend, split, and delisting information. The ticker symbols for these firms were then used to extract intraday information from ISSM. This process required some manual intervention, because CRSP data contain only the "base" three letter ticker symbol and the share class. The actual exchange ticker symbol may or may not include a one letter extension that is equal to the class. For example, although both are designated as class "A" securities, the ticker symbol for Nova Corp is NVA and the ticker symbol for First Republic Bank Corp is FRB.A.¹⁵ Another matching problem occurred because a firm's ticker symbol changes periodically, and in a few cases the dates of the changeover recorded in CRSP did not agree with the dates when the old symbol disappeared and the new symbol appeared in ISSM. In these cases, the ISSM changeover dates were used.

The ISSM database was used to extract the last trade price and the existing NYSE specialist's quote as of 11:00 A.M. Accordingly, when simulating the profits from trading strategies, all orders are assumed to be submitted as of 11:00 A.M. This time of day was chosen in order to maximize the probability of there being a good quote. Also, quoted liquidity seems to be highest in the middle of the day [see Lee, Mucklow and Ready (1993)]. By comparing the ISSM quote prices to the closing prices from CRSP, we were able to confirm that our matching procedure described above had been successful. We were also able to implement some data screens to eliminate potential reporting errors in ISSM. The most common of these errors are transpositions,

¹⁴ The one exception to this selection rule is that we exclude Berkshire Hathaway. Due to the extremely large share prices and spreads, we were not confident that we could use the results from the nonparametric estimates of price improvement.

¹⁵ In most cases, the exchange adds an extension to the base ticker symbol to distinguish the stock from another with the same base ticker symbol, usually issued by the same firm. In the case of FRB, however, there was another issue with the same ticker symbol that traded through the middle of 1988.

for example an ask price of \$53 instead of \$35, and dropping digits, for example \$4 instead of \$40. Errors of this type result in large (or negative spreads) and a midpoint of the quote that is very different from the CRSP closing prices. ISSM already employs some screens to eliminate potential errors by looking at adjacent prices. In addition to these screens we considered a quote price to be missing if the spread was zero or negative or if it was more than four times the median spread level for the firm.¹⁶ A quote was also considered to be missing if the midpoint was outside of the range defined by the closing price of the previous and current days by a percentage amount that equaled more than three standard deviations of the daily return. These screens eliminated 207 of the more than 800,000 quote observations in the database.

In order to increase the information gleaned from the sample period, we use overlapping observations, as in Boudoukh, Richardson, and Whitelaw (1994). Thus, in contrast to the Wednesday-to-Wednesday returns used by Lo and MacKinlay (1990), we construct five sequences of returns, one for each day of the week. The second column of Table 4 shows the correlation coefficients for the period covered by Lo and MacKinlay's data using our sampling methodology. The data include overlapping observations for portfolios formed from all small and large NYSE-listed stocks existing as of June 30, including those subsequently delisted. As in Lo and MacKinlay, the portfolios are equally weighted and rebalanced each week. As can be seen from Table 4, using our sampling methodology yields a higher contemporaneous correlation between the small- and large-firm portfolios and a lower autocorrelation for the small-firm portfolio. For the most part, both of these differences are probably explained by the fact that we exclude AMEX firms, many of which have very low market capitalizations. The final two entries are the averages of the individual firm autocorrelations using overlapping return intervals. The averages are calculated using all firms with at least 250 adjacent pairs of weekly returns (overlapping observations are used, so this is roughly equivalent to the 52-week minimum imposed by Lo and MacKinlay).

The final three columns of Table 4 show autocorrelations for the 1988 to 1992 period. The first of these three columns uses CRSP closing prices, the second uses the last trade price as of 11:00 A.M., and the third uses the quote midpoint as of 11:00 A.M. Although the correlations are remarkably different for the more recent sample period, with

¹⁶ The median was calculated using only quote observations with midpoints from 25% lower than to 50% higher than the midpoint of the quote in question. Without such an adjustment, a few observations either before or after a large split or reverse split could be incorrectly classified as outliers.

lower contemporaneous correlation between small and large firms and larger autocorrelation, there is little difference between the three different data sources for the period. It is interesting that the correlations are not sensitive to the time of day used to measure prices, which can be seen by comparing the results from CRSP data to those from ISSM data. Further, the fact that the small-firm portfolio autocorrelation is largely unchanged using quote-to-quote returns indicates that the autocorrelation is not a result of nonsynchronous trading. This finding is consistent with Mech (1993), who finds the same result using portfolios of NASDAQ firms. In fact, Mech goes further and shows that 4-day return intervals separated by an intervening day (where there was at least one trade) have essentially the same autocorrelation as is observed for adjacent 5-day periods. He argues that this means the autocorrelation does not result from "stale quotes." Of course, for our purposes it is not important whether stale quotes are part of the reason for the autocorrelation in quote-to-quote returns. Whatever the cause of the autocorrelation, our primary focus is determining whether it can be profitably exploited.

The averages of the individual firm autocorrelations for the 1988 to 1992 period are shown in the last three columns of the final two rows of Table 4. We used all firms with at least 50 adjacent weekly returns, including overlapping intervals. These average autocorrelations are all negative as in the earlier period, but unlike the earlier period, the large-firm autocorrelations are similar in magnitude to those of small firms. This result is probably an artifact of the realized sample path for the entire market over the 1988 to 1992 period, as revealed in the autocorrelation for the large-firm portfolio. As shown in the final column, using quote midpoints eliminates the effect of bid-ask bounce, substantially reducing the average small-firm autocorrelation, but leaving the average for the large firms unchanged.

3.1 Designing a trading strategy

The profitability of a trading strategy will ultimately depend on the execution prices of the trades, which will depend in turn on the size of the transactions. Thus, when defining a trading strategy, it is necessary to specify the size of the investment as well as the trades that will be made. Since larger trades will tend to get poorer execution, a larger investment strategy will tend to have poorer performance as measured in rate of return. However, institutions that might consider a strategy incur fixed costs for development and implementation, so a high potential return on a small investment is economically uninteresting. With this in mind we consider a strategy with an initial investment of \$5 million.

Suppose we begin 1988 with \$5 million and we decide to invest

\$1 million in each of five dynamic portfolios. For the first of these five portfolios, changes in holdings will be made on Monday at 11:00 A.M., changes for the second will be made on Tuesday, and so on. At each decision point we will decide whether to invest all of the funds either in a portfolio of small firms or a portfolio of large firms, and our decision will be based on the past week's returns for these two "size" portfolios. When we switch, we take equal positions in all of the firms. If we do not decide to switch portfolios, no changes are made in our holdings (we do not rebalance to reestablish equal weights). We will assume for the moment that our objective is to maximize expected return; later we will assess risk by comparing the return to a strategy of buy and hold.

If there are no transaction costs, then the assets should always be invested in whichever size portfolio has the higher conditional expected return. To calculate the difference in conditional expected returns, we use the same nonparametric regression techniques developed in Section 2. To avoid bias that could result from overfitting in the sample, we use the 1962 to 1987 data to estimate the conditional expected returns and determine the trading rules, and we use the 1988 to 1992 ISSM data to evaluate the profitability.¹⁷ We also show that the strategy can be improved by requiring the conditional expected difference in returns to be above some minimum amount before deciding to switch between the two portfolios. We consider both 0.5% and 1% as potential minimum cutoffs. Finally, we consider "optimizing" the strategy by trading only in the subset of firms with lower transaction costs.

To get a preliminary feel for the importance of transaction costs in implementing the proposed trading strategy, Table 5 shows the average spreads and depths for the small and large firms in our 1988 to 1992 sample. For reference, we also report the data for the firms in the Dow Jones Industrial Average. In addition to providing indicative transaction costs for the particular strategy presented here, the data in Table 5 can be used for analysis of other trading strategies.

The numbers in Table 5 are the means of the individual firm means, and the numbers in parenthesis are the medians of the individual firm medians. The percentage effective spread was calculated using the quoted spread and depth as of 11:00 A.M., subtracting the sum of the estimated price improvements for both a buy and a sell order, and then dividing by the quote midpoint. The price improvement was estimated using the smooths from Section 2. We estimated the effective spreads

¹⁷ It turns out that the profits are not much higher if we use the 1988 to 1992 data to estimate the conditional returns. Just as in the earlier period, the conditional expected difference between small- and large-firm portfolio returns is positive following positive small-firm portfolio returns and negative otherwise.

Table 5
Estimated effective spreads by order size

	DJIA	Large NYSE	Small NYSE
Percentage effective spreads			
\$3,000 order	0.22 (0.18)	0.29 (0.25)	5.41 (2.34)
\$10,000 order	0.23 (0.19)	0.32 (0.27)	7.22 (3.29)
\$30,000 order	0.26 (0.22)	0.39 (0.33)	10.71 (5.02)
Quoted spreads			
Dollar	\$.19 (\$.125)	\$.23 (\$.25)	\$.20 (\$.25)
Percent	0.41 (0.34)	0.58 (0.52)	5.45 (3.23)
Dollar quoted depth	\$541,850 (\$392,330)	\$314,430 (\$191,530)	\$32,552 (\$13,594)

The table shows the average effective spread using the ISSM spread and depth as of 11:00 A.M. along with the estimated price improvement relationships estimated in Section 2 and depicted in Figure 3. All firms are listed on the NYSE. Large and small refer to the top and bottom NYSE size quintiles. The numbers are calculated by taking the mean of the individual firm means, and the numbers in parenthesis are the median of the individual firm medians.

for orders of \$3,000, \$10,000, and \$30,000. Note that the simulated trading strategies will initially divide \$1 million equally across about 280 firms, so the average trade sizes will be in the \$3,000 to \$4,000 range.

Table 5 shows that both quoted and effective spreads are quite small for large firms, and that one can expect price improvement on fairly large orders. In fact, for firms in the Dow Jones Industrial Average, even orders of \$30,000 can expect substantial price improvement. In contrast, the quoted and effective spreads for small firms are very large as a percent of price. This is important for any trading strategy that makes frequent trades in small firms. For example, we show that a strategy that attempts to exploit *any* difference in expected returns between small and large firms will switch portfolios roughly 20 times per year, which means 10 round-trip transactions for both small and large firms. Given that average trade prices start in the neighborhood of \$3,000, this implies cumulative transaction costs of roughly 52% for the small-firm trades as compared to about 5% for the large-firm

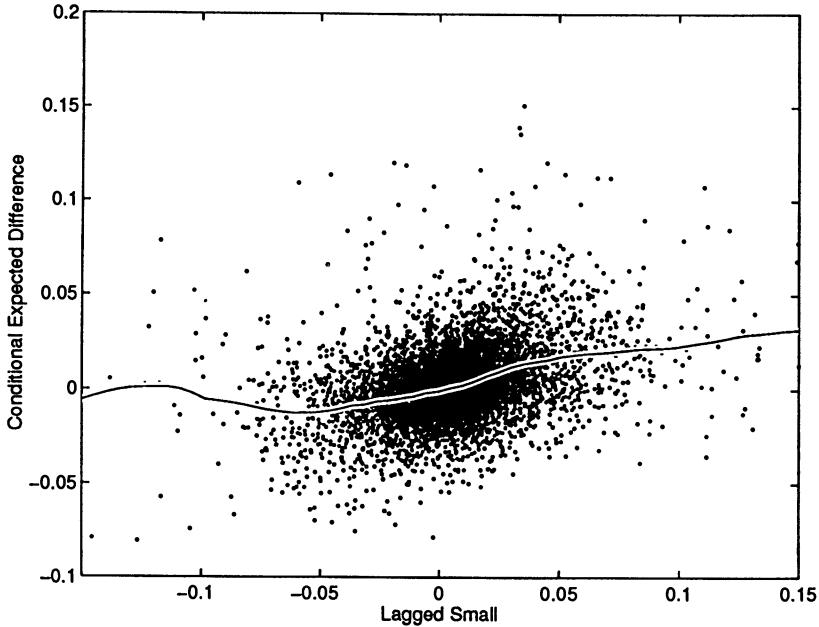


Figure 4
Loess smooth of the difference between small- and large-firm returns regressed on lagged small-firm returns

The vertical axis measures the contemporaneous difference in weekly return between a portfolio of small firms and a portfolio of large firms. The horizontal axis measures the previous week's return on the small-firm portfolio. The points are weekly observations from 1962 to 1987. To avoid obscuring the smooth from view, nearby points have been deleted.

trades.¹⁸ Of course, since some of the strategies will quickly lose large amounts of money, the trade sizes in later years will be smaller. This, in turn, will result in somewhat lower average transaction costs.

We begin the design of the trading strategy by examining the information in lagged small-firm portfolio returns. Figure 4 shows a scatter plot of the difference between the weekly returns on the small- and large-firm portfolios against the lagged value of the small-firm portfolio return for the 1962 to 1987 period. The curve is the non-parametric regression estimate of the conditional expected difference between the small- and large-firm portfolio returns. Separate non-parametric regressions (not shown) of the small- and large-firm portfolio returns on lagged small returns show that the predictability of the dif-

¹⁸ The magnitude of the transaction cost effect depends on the pretransaction cost returns on the portfolio. For example, if prices increase by 30% and there are 10 round-trip transactions each costing 5%, the cumulative return is $1.3 \times .95^{10} = .78$. This is about 52 percentage points lower than the return measured before transaction costs.

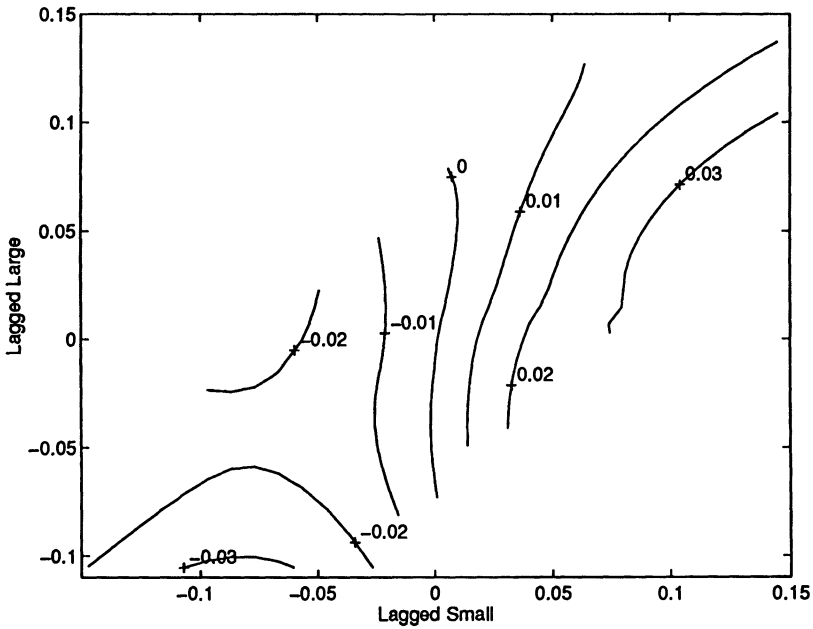


Figure 5
Contour plot of loess surface—difference between small- and large-firm returns regressed on both lagged small- and large-firm returns

The curves show constant contours of the expected difference in return between a portfolio of small firms and a portfolio of large firms, conditioned on the previous week's return for both portfolio's. The surface is estimated over the entire range of the data, but the plot is restricted to the region where the previous week's returns on the two portfolios were within 8% of each other. This region contains more than 99% of the observations.

ference in returns primarily stems from the predictability of small-firm returns. This is consistent with the data reported in Table 4, in that the autocorrelation in small-firm returns for the 1962 to 1987 sample is .25, whereas the cross-autocorrelation between large-firm returns and lagged small-firm returns is only .02. Based on the figure, it appears that a good strategy would be to invest in the small-firm portfolio whenever its previous week's return is positive.

The next step in developing the trading strategy is to incorporate the information about lagged large-firm portfolio returns. The residuals from the regression in Figure 4 are essentially uncorrelated with the lagged returns on the large-firm portfolio, but this does not imply that there is no information in lagged large-firm returns because correlation is a linear measure. To put it another way, zero correlation does not imply statistical independence, and there is no *ex ante* reason to restrict our attention to linear relationships. Figure 5 shows a contour plot of the loess two-dimensional regression of the differ-

ence between small- and large-firm portfolio returns on both lagged small- and large-firm portfolio returns. The surface is shown over a region where the difference between the lagged portfolio returns is $\leq 8\%$, and this region contains over 99% of data points. Although the loess smoother does produce estimates for the relationship over a much larger region, it is difficult to draw reliable inferences beyond the support of the data.

It is clear from Figure 5 that if one were to ignore transaction costs, the trading strategy need only consider lagged small-firm return. This is because the expected difference is positive in the right half of the figure (this is the region where lagged small-firm return is positive). However, the relationship is not generally independent of lagged large-firm portfolio returns. (If it were, all of the contour lines in Figure 5 would be vertical.) Accordingly, when we consider trading strategies that only switch if the expected difference in returns is larger than either 0.5% or 1%, we use the surface shown in Figure 5.

Comparing Table 5 to Figure 5 yields a preview to the potential profitability from the trading strategies. Approximate one-way transaction costs are half of the effective spread, or about 3% for small firms. Accordingly, trading rules that switch between the two size portfolios when the difference in weekly expected returns is 0%, 0.5%, or 1% seem bound to lose money. We could select higher cutoffs for the switch point, but the problem is that conditional differences in expected return larger than 1% are very rare events. Of course, it is theoretically possible for the rules to succeed in spite of the results in Table 5, for two reasons. First, a positive expected difference in next week's return may be followed by continued excess returns in later weeks [see Badrinath et al. (1995)]. Second, it could be the case that the switch points chosen by the strategy happen to be times of lower-than-average spreads or higher-than-average depths. In any case, we feel it is important to complete the simulations to drive home our main point, which is that transaction costs must be considered before drawing any conclusions about the importance of predictability of returns. In addition, the results below include an interesting comparison between equally weighted and buy-and-hold portfolio returns.

The calculation of the profits from the trading strategies are described below for the Monday portfolio; the process is repeated for each day of the week. The \$1 million starting value is invested in either small or large firms on the first Monday of 1988 depending on whether the previous week's small-firm portfolio return was positive or negative (this meant four of the five portfolios began with an investment in small firms). The investment was made only in firms that had a good quote on that day (over 99% of the firms/days have good quotes). The investment was split equally among the qualifying firms, and the

number of shares acquired was calculated three different ways: using the last trade price; using the quote midpoint; and using the estimated transaction price based on the quoted spread, depth, dollar amount invested, and the estimated price improvement interpolated from the smooth in Section 2. Accordingly, for each day of the week there are three portfolios for each trading rule under consideration.

On the next trading Monday, the strategy determines whether a switch is appropriate. For example, if the portfolio currently holds small firms and the return for the small-firm portfolio over the previous week was negative, then the strategy that switches to capture any difference in expected returns will dictate a changeover to large firms. A switch is made by selling all of the current shares at the last trade price, the quote midpoint, or the estimated execution price. If there is no good quote on the sale day, all three portfolios assume a costless sale of that firm's shares at the most recent transaction price. The share amounts are adjusted for any splits since the stocks were purchased and all dividends are assumed to be costlessly reinvested at the closing prices. The net proceeds from any delisted firms are calculated using the CRSP delisting amount if available, or by assuming a sale on the last trading day at the most recent trade price. The total proceeds from the sale of the securities and delistings are then split evenly among all firms in the new size portfolio that have good quotes and are reinvested. The positions are liquidated on the final trading Monday of 1992. The average annual return for a particular strategy and trade price assumption is calculated by taking the total liquidation proceeds from all five days, dividing by the \$5 million initial investment, taking the fifth root, and subtracting one.

The results of using the trading strategies over the 1988 to 1992 period are shown in Table 6. As a reference point for the performance of the dynamic strategies, we note that the 5-year average annual return for a buy-and-hold strategy was 12.9% for large firms and 9.3% for small firms. To calculate the profits from these buy-and-hold strategies, only firms with good quotes at the beginning 1988 are used. Any proceeds from delisted firms are equally invested among the remaining firms. Accordingly, the return measured using estimated execution prices differs from the return estimated using quote midpoints because of transaction costs incurred at initial investment, final liquidation, and reinvestment of delisting proceeds.

The strategy that switches portfolios to try to capture any return difference generates large excess returns ignoring transaction costs, but substantial losses when transaction costs are included. These losses are due to the large spreads for small firms, combined with the fact that the strategy makes about 10 round-trip transactions per year. The strategies that wait for larger expected return differences do much bet-

Table 6
5-Year average annual returns from various trading strategies

Strategy	Last trade price	Quote midpoint	Estimated execution price	Annual turnover (%)
Switch if expected return difference is				
0.0%	30.3	29.0	-52.8	2,036
0.5%	20.6	20.4	-8.2	480
1.0%	13.7	13.6	5.0	131
Invest only in lower spread stocks,				
Switch if expected return difference is				
0.0%	23.9	24.1	-21.3	2,036
0.5%	20.0	20.0	6.0	480
1.0%	15.6	15.7	11.6	131
Buy and hold				
Small	10.0	10.0	9.3	0
large	13.0	13.0	12.9	0
Equally weighted				
Small	14.6	14.2	-8.1	146
Large	14.1	14.0	13.2	62

The table summarizes the profitability of trading strategies that switch between a portfolio of large firms and a portfolio of small firms. All switches are made at 11:00 A.M., based on the expected return difference depicted in Figure 5. There are five portfolios, each with initial value of \$1 million, one for each day of the week. Profits are simulated using three different assumptions for transaction prices: the last trade price, the midpoint of the current quote, and the estimated transaction price including estimated price improvement. The annual turnover is the percent of the portfolio value sold, so annual turnover of 2,036% means that 20 switches are made each year (10 round-trips). The returns are calculated by taking the total ending value of the five portfolios, dividing by \$5 million, taking the fifth root, and subtracting one. The table also shows the results of restricting the strategy to the half of the stocks with the lowest estimated effective spread as a percent of price.

ter after transaction costs but do not match the simple buy-and-hold strategy. The profitability of the strategies using quote midpoints and the losses using expected transaction costs are both consistent with Mech's (1993) hypothesis that small firms may have delayed reactions to information because the delay can't be profitably exploited.

The switching strategy forms equally weighted portfolios at each switching point, so it might be natural to use equally weighted returns as a benchmark for comparison to the profits from the size-based portfolio trading strategy. In so doing, however, it must be recognized that the equally weighted portfolio is itself a dynamic portfolio that entails weekly rebalancing. In order to keep a constant proportion of wealth invested in each stock, it is necessary to sell some of the shares that increased in value and buy more of the shares that declined in value.

In calculating the benchmark equally weighted returns in Table 6, we again assumed five different portfolios, each with \$1 million initial investment, rebalanced once a week. On average, the weekly rebalancing trades were 1.2% of total portfolio value for the large-firm portfolio and 2.8% for the small-firm portfolio. Accordingly, average annual turnovers were 62% and 146%, respectively. The table indicates that for large firms, it doesn't matter much whether returns are measured using equally weighted or buy-and-hold portfolios, and the effect of transaction costs is relatively small. In contrast, comparison of the equally weighted and buy-and-hold returns for small firms yields two important results. First, the equally weighted portfolio appears to outperform the buy-and-hold portfolio, whether transactions are assumed to occur at the last transaction price or at the midpoint of the current quote. Using the quote midpoints eliminates the effect of bid-ask bounce, so the results indicate that the difference between the equally weighted and buy-and-hold portfolios stems either from contrarian profits or cross-predictability. Second, after transaction costs, the equally weighted portfolio has a lower return than the buy-and-hold portfolio. These results confirm the assertions by Blume and Stambaugh (1983) and Roll (1983) that it is better to analyze small-firm portfolios using buy-and-hold strategies.

Although we have looked at improving the trading strategies by making fewer round-trips, it would also be natural to try to make them sensitive to transaction costs by adjusting the relative investments in the various securities. This task can be quite difficult. It may not be optimal to simply reduce the trading in the highest transaction cost securities, because these may be the stocks that most strongly exhibit the behavior that the strategy is trying to exploit. To provide an indication of whether it is possible to improve on the switching strategies in this manner, we simulated the strategies using only the half of the firms at each switch point that had the smallest estimated effective spread as a percent of price. The results in Table 6 show that this does indeed reduce the transaction costs incurred by the strategies, as evidenced by the smaller return difference between the quote midpoint and estimated transaction price columns. However, another impact from adjusting the strategy in this manner is a drop in the last transaction price and quote midpoint returns. This provides further evidence that predictability is primarily a feature of high transaction cost stocks that survives because it can't be profitably exploited.

4. Conclusion

In this article we use nonparametric regression techniques to show that conditional expected price improvement is strongly and nonlin-

early related to the difference between quoted depth and order size. This dependence is important for the estimation of profits from trading strategies because it means that the profits will depend critically on the size of the positions taken. This is particularly true for small firms, where the average effective spreads range from 5% to 11% of the stock price as order size ranges from \$3,000 to \$30,000.

We use the estimated relationship between price improvement and the conditioning variables to evaluate the profitability of trading strategies that a switch between small- and large-firm portfolios in an attempt to exploit the predictability of small-firm returns. We find that the naive strategy that attempts to exploit *any* difference in conditional returns, regardless of the size of the difference, is swamped by the transaction costs. Strategies that only trade on larger conditional differences do better, but they appear to underperform the returns from a simple buy-and-hold strategy.

We also consider a modified strategy that avoids investment in the stocks with wider spreads. Although this does improve the returns after transaction costs, the pretransaction cost returns fall. Moreover, the returns after transaction costs still fall short of a simple buy-and-hold strategy. In summary, it appears that predictability is primarily a feature of high transaction cost stocks that survives because it can't be profitably exploited.

Appendix A

In this appendix we give explicit definitions of the variables used in this article. The definitions are listed alphabetically.

Depth (at bid, at ask, average): based on NYSE quotes. When estimating price improvement, for a market buy order this is the depth at the ask, and for a sell order it is the depth at the bid. Average depth is the average of the two depths over all days for the firm, and is used as a normalization for depth and order size when estimating the conditional price improvement from the pooled sample of all firms. We also regress the residuals from this pooled regression against average depth to check that the normalization "fits."

Execution Price: the trade price for the market order, given by the variable EXECPR in the TORQ database. In cases where the order was executed in separate pieces at different prices, the TORQ database includes several records in a set, and the Execution Price is the volume-weighted average. These are the instances that result in price improvements that are not even multiples of 1/8. Volumes executed at each price are calculated by summing the TORQ variables CONQTY1 through CONQTY4 for each record.

Excess Depth: difference between order size and depth at the ask

(bid) at the time a buy (sell) order was submitted, normalized by average depth.

Firm Size (Market Capitalization): the market value of outstanding equity as of June 30, given by YRVAL(CAP,I) in the CRSP database.

Order Imbalance: difference between the numbers of shares classified as buy and sell over the last 6.5 hours using the Lee and Ready (1991) algorithm. Note that the TORQ database allows unambiguous identification of trades triggered by market orders when the orders are transmitted to the floor via SuperDOT. The Lee and Ready algorithm will misclassify some of these trades; however, it was desirable to develop a statistic that could be calculated for both TORQ and ISSM, and the market order information is unavailable in the ISSM database.

Order Direction: +1 for a buy order and -1 for a sell order. Direction is given for all market orders by the OSIDE variable in the SOD file in the TORQ database. When evaluating the profitability of a trading strategy, the strategy generates buy and sell orders.

Order Size: number of shares executed for the market order in the TORQ database. The data giving the execution of a market order can span several records in the TORQ database, particularly when the record is filled at multiple prices. The order quantity was calculated by summing the values of the TORQ variables CONQTY1 through CONQTY4. This kept the order size calculation consistent with the fill price, which was calculated as the average across the orders in the set, weighted by these volumes. In rare instances, the other variables giving order size (OSHRs and RSHRS) would give different answers. This variable is normalized by average depth. When evaluating the profitability of a trading strategy, the strategy dictates the order size.

Percent of Trades Inside the Spread: number of trades that occur inside the quoted spread divided by the number of trades that occur when the spread is greater than the minimum tick size. Both numbers are calculated over the last 6.5 trading hours.

Price Improvement: NYSE ask - execution price for a buy order and execution price - NYSE bid for a sell order. Quotes are measured at the time the order was submitted, given by the OTIME variable in TORQ.

Spread Midpoint: $(ask + bid)/2$, based on NYSE quotes when the order was submitted.

Spread (Width): ask - bid, based on NYSE quotes when the order was submitted.

Time of Day: minutes past 9:30 A.M. This is the time the order was submitted, given by the OTIME variable in TORQ. In estimating the conditional relationships, we only included orders submitted between 9:30 A.M. and 4:00 P.M. if there was a good quote in effect. We found that there was little pattern in residual price improvement (after con-

trolling for changes in the quoted spread and depth) over the day. In the evaluation of trading strategies, market quote data were sampled as of 11:00 A.M.

Volatility: difference between the highest and lowest transaction price over the last 6.5 hours of trading.

Volume: number of the firm's shares traded over the last 6.5 hours of trading.

Appendix B

In this appendix we discuss the local linear smoother in more detail, present some smooths using the variable bandwidth local linear smoother, and sketch the steps necessary to implement the variable bandwidth selection procedure. Recall that the span in the loess smoother is checked by smoothing the residuals on the same conditioning variables. We used the variable bandwidth local linear smoother's objective determination of bandwidth as an additional check of the spans used in our univariate loess regressions.¹⁹

Again let $\{Y_i, X_i\}_{i=1}^n$ be a finite record of observations of price improvements and a conditioning variable. We assume that the finite record is a sample from a stationary stochastic process that satisfies either *strong mixing* or ρ -mixing. The local linear smoother is a kernel-based smoother developed recently by Fan (1992) and Fan and Gijbels (1992). The local linear smoother at a grid point x_o is computed by minimizing

$$\sum_{i=1}^n \left(Y_i - \sum_{j=0}^1 \beta_j(x_o)(x_o - X_i)^j \right)^2 \alpha(X_i) K \left(\frac{X_i - x_o}{h_n} \alpha(X_i) \right),$$

where $K(\cdot)$ is the kernel function and h_n is the smoothing parameter or bandwidth.²⁰ K determines the *shape* of the weights in the weight function w_i , and h_n determines the *size* of the weights or neighborhood.²¹ In theory the function $\alpha(X_i)$ is an estimate of the density of the regressors at the point x_i . However, in practice the α function is defined over grid points, not observations. It is more convenient to write the local linear smoother as a solution to a weighted

¹⁹ For a comparison of the local linear smoother and the loess smoother see Fan and Gijbels (1994).

²⁰ Masry and Fan (1994) show how to conduct local polynomial estimation of regression functions (of which the local linear smoother is a special case) using dependent data.

²¹ The variable h_n is used to denote the dependence of the bandwidth h on the number of observations n .

least-squares problem. Define the following:

$$X = \begin{pmatrix} 1 & (X_1 - x_0) \\ \vdots & \vdots \\ 1 & (X_n - x_0) \end{pmatrix},$$

$$y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

$$W = \text{diag}(\alpha(X_i)K(((X_i - x_0)/b_n)\alpha(X_i))).$$

Then the solution to this problem is given by

$$\hat{\beta}(x_0) = (X'WX)^{-1}X'Wy,$$

where $\hat{\beta}(x_0)$ is given by

$$\hat{\beta}(x_0) = [\hat{\beta}_0(x_0), \hat{\beta}_1(x_0)],$$

and the estimated response function is given by

$$\hat{m}(x_0) = \hat{\beta}_0(x_0).$$

There are many possible choices for K . However, Gasser et al. (1985) and Härdle and Kelly (1987) have shown that on the basis of mean square error, the choice of various kernels is not of primary importance. They suggest that the selection of the kernel be based on computational efficiency. With this in mind, we use the kernel with a parabolic shape and finite support known as the *Epanechnikov kernel*, which is given by

$$K(u) = 0.75(1 - u^2)I(|u| \leq 1).$$

In addition, Fan and Gijbels (1992) show that the Epanechnikov kernel has larger minimax efficiency than the normal or uniform kernels.

While the choice of kernel is of secondary importance, the choice of bandwidth is of primary importance. This is because the bias variance trade-off is governed by the size of the bandwidth. That is, as the bandwidth is increased (increasing the size of the neighborhood or weights) this reduces the variance of the estimate but increases the bias of the estimate (see the discussion in Section 2). The advantage of the *data driven* bandwidth selection technique developed recently

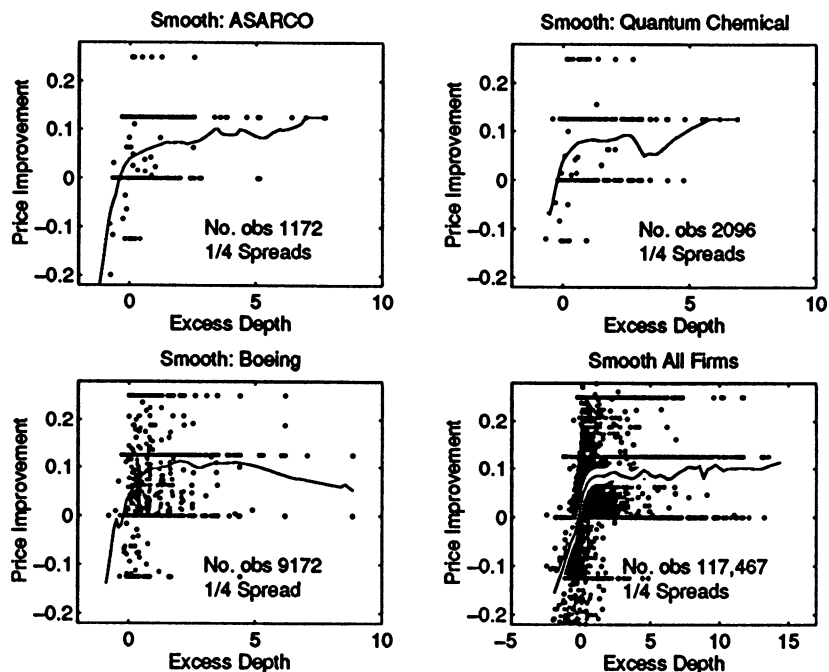


Figure 6
Local linear smooth of price improvement on excess depth for 1/4 spreads—variable bandwidth selected by minimizing the integrated mean square error
 Expected price improvement is measured in dollars per share. Excess depth is the difference between quoted depth and the size of the order, where both are expressed as a fraction of the average quoted depth for the firm. The points are individual observations from the TORQ sample. To avoid obscuring the smooth from view, nearby points have been deleted.

by Fan and Gijbels (1994)²² is that it allows the data to dictate the optimal size of the bandwidth taking into account the bias variance trade-off. In addition, the procedure allows one to select a variable bandwidth even when an extremely large number of observations are being used.²³

Figure 6 shows fitted smooths and scatterplots of price improvement as a function of excess depth for orders submitted at a 1/4 spread using the variable bandwidth local linear smoother. The four panels in the figure show the data and smooths for ASARCO, Boeing, and Quantum Chemical, and the aggregate across all firms. Figure 7 contains the same data for orders submitted at a 1/8 spread. Comparison

²² See Härdle and Kelly (1987) for a discussion of bandwidth selection techniques.

²³ Alternative procedures such as leave one cross-validation become computationally infeasible on data sets of the size considered here.

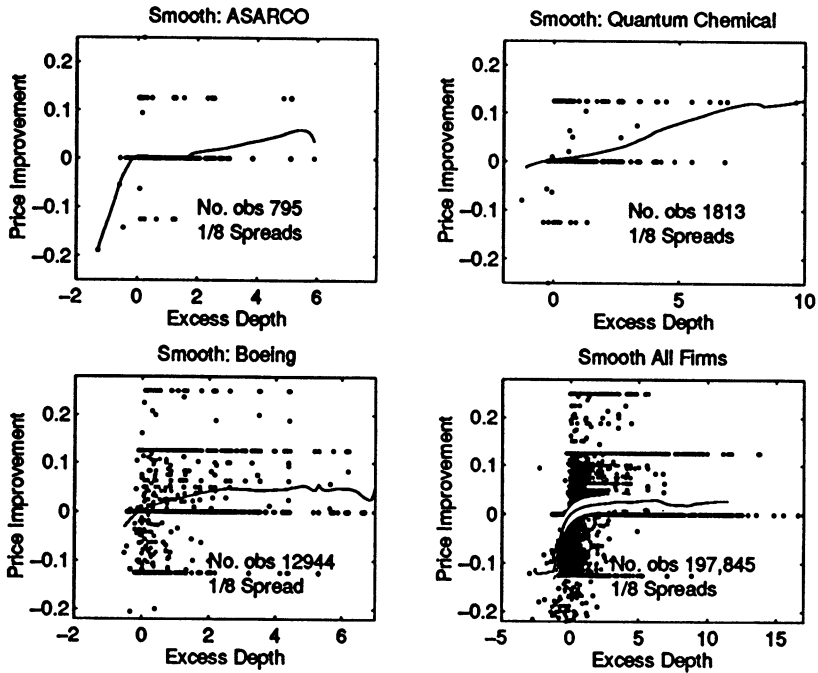


Figure 7
Local linear smooth of price improvement on excess depth for 1/8 spreads—variable bandwidth selected by minimizing the integrated mean square error
 Expected price improvement is measured in dollars per share. Excess depth is the difference between quoted depth and the size of the order, where both are expressed as a fraction of the average quoted depth for the firm. The points are individual observations from the TORQ sample. To avoid obscuring the smooth from view, nearby points have been deleted.

of these figures to Figures 1 and 2 suggest that the loess smooth and the variable bandwidth local linear smooth provide the same interpretation of the relationship between price improvement and excess depth.

The bandwidths for the smoothes in Figures 6 and 7 were computed using the following procedure.

Step 1. Begin by partitioning up the interval over which the estimation is to be conducted. Suppose the interval of estimation is denoted by $I = [I_0, I_N]$ and an arbitrary subinterval by I_k . The smoothes in Figures 6 and 7 had five subintervals.

Step 2. For each Interval I_k fit a third-order polynomial. Note that the local linear smoother is a first-order polynomial. However, a third-order polynomial is fit at this stage in order to extract curvature information from the data. The optimal bandwidth \hat{h} is selected at this

step by minimizing the *integrated residual squares criterion* given by

$$IRSC(b) = \int_{I_k} RSC(y; b) dy,$$

where RSC is defined as

$$RSC(x_o; b) = \hat{\sigma}^2(x_o)[1 + 2V],$$

and $\hat{\sigma}^2(x_o)$ is the normalized weighted residual sum of squares given by

$$\hat{\sigma}^2(x_o) = \frac{1}{tr\{W - (X^T W X)^{-1} X^T W^2 X\}} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 K\left(\frac{X_i - x_o}{b}\right).$$

V is the variance term and is given by the first diagonal element of the matrix:

$$\{(X^T W X x)^{-1} (X^T W^2 X) (X^T W X)^{-1}\}.$$

To understand the residual squares criterion note the following. The conditional variance of $\hat{\beta}$ is given by

$$\text{Var}(\hat{\beta}_{X_1}, \dots, \hat{\beta}_{X_n}) \approx \sigma^2(x_o) \{(X^T W X)^{-1} (X^T W^2 X) (X^T W X)^{-1}\}.$$

The above equation implies that if the bandwidth is too large (i.e., big b), then $\hat{\sigma}^2(x_o)$ is large, since the residual sum of squares is large. On the other hand, if b is too small the V is large. The residual squares criterion bandwidth is then given by

$$\hat{h}^{RSC} = adj\hat{h},$$

where adj is an adjustment parameter that depends on the order of polynomial being estimated and the choice of kernel.²⁴ For the Epanechnikov kernel and a third order polynomial this adjustment parameter is given by .7776.

Step 3. Since step 2 is conducted for each interval I_k , step 2 produces a bandwidth step function. A “smoothed” bandwidth step function is produced by locally averaging. With this smoothed bandwidth function, a third-order polynomial is used to estimate $\hat{\beta}_2(x_o)$, $\hat{\beta}_3(x_o)$, and $\hat{\sigma}^2(x_o)$. These are the “pilot” estimates used as inputs to estimate the *mean square error*.

Step 4. Given the estimates from step 3, a bandwidth is again selected for each interval by minimizing the integral of the estimated

²⁴ See Fan and Gijbels (1994).

mean square error defined by

$$M\hat{S}E(x_0, b) = \hat{b}^2(x_0) + \hat{V}(x_0).$$

$\hat{b}(x_0)$ is the estimated bias, which is given by $(X^T W X)^{-1}(\hat{\beta}_2 s_{n,2} + \hat{\beta}_3 s_{n,3})$, where $s_{n,j} = \sum_{i=1}^n (X_i - x_0)^j K\left(\frac{X_i - x_0}{b}\right)$. $\hat{V}(x_0)$ is the estimated variance term and is given by the first diagonal element of the matrix:

$$(X^T W X)^{-1} (X^T W^2 X) (X^T W X)^{-1} \hat{\sigma}(x_0).$$

Note that step 4 uses resulting estimates from step 3 to construct an estimate of the mean square error. It is in this sense that they are pilot estimates.

Step 5. Finally, smooth the resulting sequence of bandwidths (i.e., one for each interval) by local averaging. Then, using the smoothed bandwidth function, fit a first-order polynomial (i.e., a local linear smooth).

A constant bandwidth can be selected using this procedure by simply removing the actions related to partitioning up the interval. That is, for a constant bandwidth it is not necessary to partition up the interval of estimation since the bandwidth is the same over the range of the data.

References

- Amihud, Y., and H. Mendelson, 1986, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 17, 221–249.
- Badrinath, S., J. Kale, and T. Noe, 1995, "Of Shepherds, Sheep, and the Cross-autocorrelation in Equity Returns," *Review of Financial Studies*, 8, 401–430.
- Blume, M., and R. Stambaugh, 1983, "Biases in Computed Returns," *Journal of Financial Economics*, 12, 387–404.
- Boudoukh, J., M. Richardson, and R. Whitelaw, 1994, "A Tale of Three Schools: Insights on Autocorrelations of Short Horizon Stock Returns," *Review of Financial Studies*, 7, 539–574.
- Chambers, J., and T. Hastie, 1993, *Statistical Models in S*, Chapman and Hall, London.
- Chan, L., and J. Lakonishok, 1993, "Institutional Trades and Intraday Stock Price Behavior," *Journal of Financial Economics*, 33, 173–199.
- Cleveland, W., 1979, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–839.
- Cleveland, W., 1994, *Visualizing Data*, Chapman and Hall, London.
- Dupont, D., 1995, "Market Making, Prices and Quantity Limits," working paper, Federal Reserve Board.
- Easley, D., and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19, 69–90.

- Fan, J., 1992, "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.
- Fan, J., and I. Gijbels, 1992, "Variable Bandwidth and Local Linear Regression," *Annals of Statistics*, 20, 2008–2036.
- Fan, J., and I. Gijbels, 1994, "Data Driven Bandwidth Selection for Variable Bandwidths," manuscript.
- Gasser, T., H. Muller, and V. Mammitzsch, 1985, "Kernels for Nonparametric Curve Estimation," *Journal of Royal Statistical Society Series B*, 47, 238–252.
- Härdle, W., and G. Kelly, 1987, "Nonparametric Kernel Regression Estimation—Optimal Choice of Bandwidth," *Statistics*, 18, 21–35.
- Harris, L., and J. Hasbrouck, 1992, "Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy," working paper, University of Southern California and New York University.
- Hasbrouck, J., G. Sofianos, and D. Sosebee, 1993, "New York Stock Exchange Systems and Trading Procedures," working paper, New York Stock Exchange.
- Hasbrouck, J., and D. Sosebee, 1993, "Orders, Trades, Reports and Quotes at the New York Stock Exchange," working paper, New York Stock Exchange.
- Kyle, A., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1336.
- Lee, C., 1992, "Market Integration and Price Execution for NYSE-Listed Securities," *Journal of Finance*, 48, 1009–1038.
- Lee, C., B. Mucklow, and M. Ready, 1993, "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis," *Review of Financial Studies*, 6, 345–374.
- Lee, C., and M. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733–746.
- Lo, A., and A. MacKinlay, 1990, "When are Contrarian Profits Due to Stock Market Overreaction?" *Review of Financial Studies*, 3, 175–205.
- Masry, E., and J. Fan, 1994, "Local Polynomial Estimation of Regression Functions for Mixing Processes," working paper, University of North Carolina.
- Mech, T., 1993, "Portfolio Return Autocorrelation," *Journal of Financial Economics*, 34, 307–344.
- Petersen, M., and D. Fialkowski, 1994, "Posted Verses Effective Spreads: Good Prices or Bad Quotes?" *Journal of Financial Economics*, 35, 269–292.
- Prahl, W., 1993, "Market Market Price Discrimination and Stock Return Seasonality," working paper, University of Wisconsin-Madison.
- Roll, R., 1983, "On Computing Mean Returns and the Small Firm Premium," *Journal of Financial Economics*, 12, 371–386.