Evidence on the Speed of Convergence to Market Efficiency

by

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam

June 22, 2004

Abstract

Daily returns for stocks listed on the New York Exchange (NYSE) are not serially dependent. In contrast, order imbalances on the same stocks are highly persistent from day to day. These two empirical facts can be reconciled if sophisticated investors react to order imbalances within the trading day by engaging in countervailing trades sufficient to remove serial dependence over the daily horizon. How long does this actually take? The pattern of intra-day serial dependence, over intervals ranging from five minutes to one hour, reveals traces of efficiency-creating actions. For the actively traded NYSE stocks in our sample, it takes longer than five minutes for astute investors to begin such activities. By thirty minutes, they are well along on their daily quest.

Contacts

	Chordia	Roll	Subrahmanyam
Voice:	1-404-727-1620	1-310-825-6118	1-310-825-5355
Fax:	1-404-727-5238	1-310-206-8404	1-310-206-5455
E-mail:	Tarun_Chordia@bus.emory.edu	Rroll@anderson.ucla.edu	asubrahm@anderson.ucla.edu
Address:	Goizueta Business School Emory University Atlanta, GA 30322	Anderson School UCLA Los Angeles, CA 90095-1481	Anderson School UCLA Los Angeles, CA 90095-1481

We thank an anonymous referee, Peter Bossaerts, Michael Brennan, Jeff Busse, Eugene Fama, Laura Frieder, Will Goetzmann, Clifton Green, Andrew Karolyi, Pete Kyle, Francis Longstaff, Elizabeth Odders-White, Mike Piwowar, Stephen Ross, Ross Valkanov, Kumar Venkatraman, Ingrid Werner, and seminar participants at Bocconi University, Princeton University, Southern Methodist University, Texas A&M University, the University of Texas, the New York Stock Exchange, the 2002 Western Finance Association Conference, and the 2002 University of Maryland Accounting and Finance Conference for valuable comments and suggestions. We are grateful for financial support from the Q-Group.

Evidence on the Speed of Convergence to Market Efficiency

Abstract

Daily returns for stocks listed on the New York Exchange (NYSE) are not serially dependent. In contrast, order imbalances on the same stocks are highly persistent from day to day. These two empirical facts can be reconciled if sophisticated investors react to order imbalances within the trading day by engaging in countervailing trades sufficient to remove serial dependence over the daily horizon. How long does this actually take? The pattern of intra-day serial dependence, over intervals ranging from five minutes to one hour, reveals traces of efficiency-creating actions. For the actively traded NYSE stocks in our sample, it takes longer than five minutes for astute investors to begin such activities. By thirty minutes, they are well along on their daily quest.

Evidence on the Speed of Convergence to Market Efficiency

I. The Issue.

For most of its scientific life, the field of finance has debated the question of market efficiency. Despite a long list of empirical anomalies and extensive indications of psychological quirks among investors, most financial economists and professionals still profess that asset prices are difficult to predict. Schwert (2001) reviews a number of well-documented anomalies and finds that some of them have disappeared, perhaps revealing ephemeral market <u>in</u>efficiencies. But he argues also that other anomalies appear to have been "discovered" even though they did not exist.

There is a growing literature about the irrationalities of individual investors. Odean (1999), for instance, finds that small investors have a perverse ability to forecast future returns; their stock purchases perform worse than their sales. Barber and Odean (2000) find that the more individuals trade, the worse their returns. Benartzi and Thaler (2001) document bizarre portfolio choices among individuals allocating pension assets to various classes.

Despite their reluctance to forecast prices, most scholars admit also that some individuals behave foolishly all the time and all individuals behave foolishly some of the time. When reconciling these conflicting views, we usually resort to flurry of hand waving and invoke the mantra of aggregation. Somehow, from within the blizzard of behavioral proclivities, the "market" becomes efficient, or, at least efficient enough that professors and money managers have a very difficult time beating passive investment strategies. But exactly how does this happen and how long does it take?

The concepts of market efficiency as defined by Fama (1970) in his seminal review, weak, semi-strong, or strong form efficiency, represent a road map for statistical tests. They are silent, however, about market processes that might deliver the hypothesized phenomena. Clearly, efficiency does not just congeal after spontaneous combustion. It depends, somehow, on individual actions.

This idea was formalized by Grossman (1976) and Grossman and Stiglitz (1980). In their model, the market price does not fully incorporate all knowable information because informed investors make (infra-marginal) returns from exploiting deviation of prices from fundamental values. Further, Cornell and Roll (1981) borrowed a model from evolutionary biology to show that efficient markets must be inhabited by both passive investors, who take prices as correct forecasts of future value, and by active investors who expend resources in an effort to detect errors in prices. Market efficiency is the state in which neither the marginal active nor the marginal passive investor has an incentive to alter his or her respective approach. Infra-marginal active investors pay to become better informed and somehow move prices enough that passive investors can enjoy a free ride without sacrificing much return (indeed, <u>any</u> return at the margin). Of course, the extent to which prices do not incorporate all information depends in both settings on the cost of information production. The smaller are these costs, the more efficient is the market.

Many investors still follow technical trading strategies that appear to generate little revenue and much cost; these strategies have long been the subject of much critique by finance professors. Recently, Chordia, Roll, and Subrahmanyam (2002) document a seemingly related and intriguing phenomenon during a study of market-wide order imbalances on the New York Stock Exchange. Market order imbalance, defined as the aggregated daily market purchase orders less sell orders for stocks in the S&P500 index, is highly predictable from day to day. A day with a high imbalance on the buy side will likely be followed by several additional days of aggregate buy side imbalance; and similarly for an imbalance on the sell side. This implies that investors continue buying or selling for quite a long time, either because they are herding or because they are splitting large orders across days, or both. More than fifty percent of tomorrow's imbalance among S&P500 stocks can be forecast by past returns and past imbalances.

Yet the S&P500 index is virtually a random walk over a horizon of one day. During the 1996-2002 sample period, it had a first order autocorrelation coefficient of -0.0015 (p-value=0.95) and insignificant autocorrelations at all longer daily lags. This suggests, of course, that some astute investors must be correctly forecasting continuing price pressure from order imbalances and conducting countervailing trades within the very first day, trades sufficient to remove all serial dependence in returns, which would otherwise be induced by the continuing procession of order imbalances.

There are at least two puzzles here: First, why do some naïve investors persist in their orders for days on end when it does them no good (because there is no inter-day return dependence)? Second, how long within the day does pressure from order imbalances continue to move prices? When thinking about this second and more important question, it seems rather obvious that some finite time period, albeit perhaps quite a short period, is required for sophisticated investors to counteract a sudden and unexpected preponderance of orders on the same side of the market.

It should not be true that returns are independent from trade to trade or even from minute to minute. It must take at least some time for astute investors to figure out what is happening to orders, to ascertain whether there is new pertinent information about values, and to expunge any serial dependence remaining after prices adjust to their new equilibrium levels. The horizon over which this activity takes place is the object of our study. We propose to investigate how long it takes the market to achieve weak-form efficiency; i.e., how long it takes to remove return dependence.

Other researchers have investigated questions similar to the one we address, but in very specific contexts. In early work, Patell and Wolfson (1984) show that dividend and earnings announcements "interrupt" the usual pattern of return serial dependence for at least fifteen minutes and that prices do not revert completely to their normal serial correlation pattern for up to ninety minutes. Although they make no explicit statement about how this happens, they clearly have in mind the activities of arbitrageurs who offset the impulsive reactions to company announcements of naïve investors.[1]

Garbade and Lieber (1977) formulate a model of independent changes in equilibrium price coupled with random orders to buy or to sell at quoted ask and bid prices. They use data on two stocks for a single month and find that this model does not describe price moves for short time

[1] More recently, Busse and Green (2002) find that news reports about individual stocks on the financial television network CNBC are incorporated into stock prices within one to two minutes

intervals (a few minutes) while it is consistent with price moves over longer horizons.[2] In concluding, they recognize that "…investors who monitor the market continually during the day…" might be instrumental in bringing about the observed pattern.

Epps (1979) studies price adjustments for a group of firms in the same industry (automobiles). He finds rapid but not instantaneous adjustments across firms to common news relevant for all industry firms. Correlations among the returns increase with the time interval, which suggests cross-firm variation in the speed of adjustment to new information. Epps' overall conclusion is that "…the predictive value of a price change in one stock endures not much more than one hour…" but "…the average lag in the response of prices [to new information] is more then 10 minutes" (p. 298).

Related theoretical models were developed by Copeland (1976) and Hillmer and Yu (1979). Copeland's model predicts a positive correlation between trading volume and absolute price change and positive skewness in volume. However, it does not include a provision for the activities of arbitrageurs. Hillmer and Yu note that the incorporation of information into prices "cannot be completed instantaneously" because "…in practice an investor will not react…unless he is convinced that it is economically advantageous." (p. 321). They develop various alternative statistical models involving price, volume, and volatility, all inspired by the idea that investor/arbitrageurs would be watching the market closely and reacting occasionally. Their tests, however, involve only a handful of anecdotal events.

[2] Unlike us, Garbade and Lieber (1977) do not have access to bid-ask quote mid-points and hence are unable to separate bid-ask bounce in transaction prices from true serial correlation.

Much later, Chakrabarti and Roll (1999) formulate a model populated by Bayesian traders/arbitrageurs who attempt, through observing the trading of others, to deduce the quality of their information. Simulations of the model show that the market usually converges more rapidly to an equilibrium price that is a better predictor of true value when arbitrageurs react to one another as opposed to trading solely on their own information.

Section II below describes the data. Section III presents our analysis of how quickly prices of highly liquid stocks become efficient. Section IV concludes and suggests further investigations.

II. The Data

Since we already know that serial dependence in returns is close to zero for active stocks over a daily horizon, our investigation of the efficiency-creating process must focus on intra-day trading. We would like to measure the timing of efficiency creation as precisely as possible, so it seems sensible to examine frequently-traded stocks for which very short term serial dependence can actually be observed. This suggests that very small stocks should be excluded owing to the difficulty inherent in measuring serial dependence when trading is infrequent.

Because transactions data are so voluminous (e.g., IBM alone has several million transactions a year), this study uses a limited sample of stocks and time. Our calculations here cover 150 large stocks listed on the New York Exchange for three recent years, 1996, 1999 and 2002.[3] These years were chosen because (a) transactions data are available from the TAQ (Trade and Automated Quotations) database recorded by the Exchange, and (b) they bracket significant changes in the minimum tick size, which was reduced from $1/8 to $1/16 during 1997 and was

[3] Filters applied to the transactions data are described in Chordia, Roll, and Subrahmanyam (2000).

reduced further to one cent in by January 2001.[4] We hoped to discern changes in the price formation process during years preceding and following these events. Future investigations should extend the investigation to smaller firms, and other years, exchanges, and countries.

The first fifty firms in our sample are the largest listed firms at the beginning of each sample year. Their market capitalizations (across the three sample years) range from \$398 billion to \$18.4 billion. The mid-cap group (i.e., mid-cap within our sample) consists of the next largest 50 stocks by market capitalization at the beginning of each year. Their sizes range from \$41.0 billion to \$9.72 billion. Finally, the smaller size group (referred to as "small" in our study) is comprised of, in turn, the next 50 stocks by market capitalization with sizes ranging from \$18.8 billion to \$6.69 billion. Market caps within our sample vary by a factor of about 60 to 1 but even the smallest firms are large to be actively traded. We present combined results for the 150 stocks throughout most of the paper, and present size-stratified results towards the end.

Each transaction for each of the 150 stocks during the three years is recovered from the TAQ database, which provides not only trade prices, but also bid and ask quotes associated with each transaction. This allows us to use the Lee/Ready (1991) trade assignment algorithm to estimate whether a particular trade was buyer- or seller-initiated.[5] Order imbalance for each stock over any time interval can then be calculated variously as the number of buyer- less the number of seller-initiated trades (OIB#), the number of buyer-initiated shares purchased less the number of

[4] Ball and Chordia (2001) show that for the largest stocks, more than half the bid-ask spread in 1996 was due to the impact of rounding onto the tick grid. Also, Chordia and Subrahmanyam (1995) argue that a reduction in the tick size would result in more competition and less payment-for-order-flow, thus, causing orders to flow to the least cost providers of market making services.

[5] The Lee/Ready algorithm classifies a trade is as buyer- (seller-) initiated if it is closer to the ask (bid) of the prevailing quote. If the trade is exactly at the mid-point of the quote, a "tick test" is used whereby the trade is classified as buyer- (seller-) initiated if the last price change prior to the trade is positive (negative). Note that a limit order is most often the passive side of the trade; i.e., the non-initiator.

seller-initiated shares sold (OIBSh), or the dollars paid by buyer-initiators less the dollars received by seller-initiators (OIB$).

The first of these order imbalance measures disregards the size of the trade, counting small orders equally with large orders. The second and third measures weight large orders more heavily. The distinction is important here because we hope to shed light on how arbitrageurs make the market more efficient over very short horizons and presume that arbitrageurs, in an effort to quickly exploit deviations of prices from fundamentals, usually undertake larger trades than naïve investors.

III. The Evidence.

III.A. Evidence of efficiency at a daily horizon.

Using Center for Research in Securities Prices (CRSP) returns data, we first set out to ascertain whether our sample of stocks conformed to semistrong-form efficiency over a daily horizon; i.e., whether future returns could be predicted by either past returns or past order imbalances. Table 1 documents the daily return serial correlations and shows that the average first-order daily autocorrelation coefficient during 1996 was 0.002; the t-statistic, 0.38, was calculated from the cross-section of sample autocorrelation coefficients assuming independence.[6]

This positive (though insignificant) coefficient is somewhat surprising because the bid-ask bounce is known to induce negative first-order autocorrelation in trade-to-trade returns.[7] During

---

[6]It seems likely that the assumption of cross-sectional independence actually results in an overstatement of statistical significance because returns, and hence sample correlation coefficients, are mostly positively correlated. This implies that the estimated standard error of the sample mean is too small since it omits the mostly positive covariance terms that would be in the true standard error.

[7] Roll (1984) suggests that the negative autcorrelation due to bid-ask bounce could be attenuated by positive serial dependence because of sluggish price adjustment to news.

2002, stocks did exhibit such negative (and significant) autocorrelation. No serial correlation coefficient is positively significant in any year

To avoid contamination of return serial correlations by bid-ask bounce, we compute returns from quote midpoints as well as from transaction prices. For each transaction during each day, the prevailing quote before the trade was used to compute a bid-ask midpoint.[8] Returns are then computed from these midpoints. For example, the daily midpoint returns in Table 1 are computed from the bid and ask quotes just prior to the last transaction of the day. Again, no positive coefficient is significant, and the only significant coefficient is in 2002.

Table 1 also reports simple correlations between returns and the three measures of order imbalance, both contemporaneous correlations and correlations with OIB lagged by one day. As could be expected, there is a very strong positive contemporaneous correlation between either measure of return (trade or midpoint) and any of the OIB measures. Not surprising also, the share and dollar measures, OIBSh and OIB$, are considerably more highly correlated with contemporaneous returns, particularly for larger firms.

Lagged OIB# (in number of trades) is significantly and positively correlated with returns in all years. The magnitude of the correlation is 0.06 or less, so the economic value of the implied prediction would be relatively small. Moreover, the magnitude is lower in 1999 and 2002 relative to 1996. This is consistent with a small improvement in market efficiency perhaps brought about by the minimum tick size reduction. The correlation between returns and lagged

[8] A five-second delay rule was used in 1996, and, based on feedback from microstructure scholars, who indicated that reported errors dramatically declined in the 1999-2002 period, this delay was not imposed for the latter two years.

OIB in either shares or dollars is insignificant in 1996 and 2002 and negatively significant in 1999.

Notice that the order imbalance measures themselves are strongly and positively autocorrelated from day to day, a feature particularly striking for OIB# (which weights all trades equally regardless of size). Indeed, the autocorrelation coefficient for OIB# exceeds 0.3 in both 1996 and 1999. It is 0.249 in 2002. In an earlier paper, Chordia, Roll, and Subrahmanyam (2001) show that even aggregate market order imbalances persist for several days.

III.B. Evidence about efficiency over short horizons within the trading day.

We compute short-horizon returns from prices closest to the end of various time intervals within the trading day. For example, ten-minute returns are computed for each stock by finding the transaction closest to 9:40 a.m., 9:50 a.m., etc.[9] Because some calculations involve lagged values, the first interval of each trading day is discarded since it would have been correlated with a lagged interval from the previous trading day.[10]

There is some imprecision in very short-term returns because trades do not necessarily occur at the exact end of time each interval. If the closest price to the end of an interval is more than 150 seconds away, either before or after, the return for that interval is not used in our calculations. Within the large stock sample, the average time between transactions is 19 seconds (averaged across the three years). Over intervals longer than five minutes, this problem obviously becomes progressively less material.

[9]New York Stock Exchange trading hours are 9:30 a.m. to 4:00 p.m. except for rare exceptions (e.g., 9/11).

[10] Intervals of sixty minutes were set backward from the end of the trading day. For example, each day has five one-hour intervals (11-12, 12-1,…,3-4) included in the calculations; the interval from 10 to 11 a.m. provides lagged observations only and data from 9:30 to 10 a.m. are not used at all.

Order imbalances are computed over all trades within each time interval. For example, contemporaneous OIB# during the ten-minutes ending at 9:50 a.m. consists of the number of buyer-initiated trades less the number of seller-initiated trades between 9:40:01 a.m. and 9:50:00 a.m. The lagged ten-minute OIB# is the corresponding accumulation between 9:30:01 a.m. and 9:40:00 a.m.

We use two measures of significance. The first is the cross-sectional average of the t-statistic from individual stock regressions. The second is computed from the cross-section of coefficients after accounting for cross-correlation in the individual stock regression residuals. The specific formula for the correction that we employ, also mentioned in footnote 8 of Chordia, Roll, and Subrahmanyam (CRS) (2000), assumes that the residual cross-correlation and the residual variance are homogeneous across stocks. Under these assumptions, the standard error is inflated by $[1+(N-1)\rho]^{1/2}$, where N is the number of regressions[11] and ρ is the common cross-correlation in the residuals, which is proxied by the average residual cross-correlation across the 150 adjacent regressions for stocks (i.e., the average of the 11,175 pairwise residual correlations for the 150 regressions).

The first t-statistic can be viewed as the expectation of the distribution of individual regression t-statistics. The second t-statistic is calculated in a manner similar to the Fama/Macbeth (1973) (FM) t-statistic, except that here we are cross-sectionally averaging coefficients from individual time-series regressions, whereas the FM method does the opposite. In the FM method, residual returns can be viewed as serially independent so it is acceptable to simply calculate the simple t-

[11] The CRS formula contains an typographical error in the form of an extraneous digit, 2.

statistic for testing that the mean coefficient is different from zero. In our case, residual returns are cross-sectionally dependent, so we have to adjust the analog of the FM t-statistic for residual cross-correlation.

Our first results are in Table 2, which reports serial regressions for returns and univariate regressions of returns on lagged order imbalances. Turning first to returns, there is little evidence of unconditional serial dependence. We base this conclusion on the second of the two reported t-statistics. Across all years, no t-statistic exceeds 2.0 in absolute value; 13 of the 15 are less than 1.0 in absolute value. This suggests that these stocks conform well to weak-form efficiency; i.e., using only the past history of returns, there is little, if any, predictability of future returns even over intervals as short as five minutes.

The story does not stop there, however, because Table 2 also shows that lagged order imbalances are often significant predictors of future returns over short intervals.[12] There is significance at up to thirty minutes during 1996, up to ten minutes in 1999, and up to five minutes in 2002 (based on the second of the two t-statistics, which we believe is more reliable). The obvious pattern in all regressions where OIB predicts returns is the declining predictive ability as the return interval lengthens. The coefficients and t-statistics are much larger for the shorter intervals. This suggests that the market is not strong-form efficient over very short periods. Strong-form efficiency is the appropriate criterion here because agents off the exchange cannot observe order imbalances easily; only the NYSE specialist and perhaps astute floor traders observe an imbalance immediately. However, traders seem able to deduce and accommodate the

[12] Since the computation of the imbalance measure requires the use of prices, it may appear as though the predictability of returns from imbalances is somehow proxying for serial dependence in returns. However, returns themselves show little serial dependence, as is evident from Table 2, and we use quote mid-points to rule out serial

impact of an imbalance quickly and their accommodative capacities appear to have increased over time. In no year does it take more than 30 minutes to eliminate the predictive content of OIB. This is all the more impressive when one considers the serial regressions of OIB itself in Table 3. We saw earlier that OIB is strongly autocorrelated over a daily horizon (Table 1) and Table 3 confirms that a similar pattern is present even over intervals as short as five minutes. Notice that the coefficients generally grow larger with interval length, so that the serial dependence in OIB actually increases as the interval lengthens from five minutes to one hour. The explanatory power (R-square) also is modestly larger for longer intervals.

In the absence of countervailing trading activity by arbitrageurs and the specialist, this persistence in order imbalances would have induced strong return predictive ability for OIB over longer intervals, which, as we have just seen, does not obtain. Consequently, the results are consistent with the notion that agents are acting not only to countervail imbalances concurrently but also to predict and forestall the influence of future imbalances.

Overall, we interpret these results to reveal the actions of three distinct groups. Order imbalances in the first instance arise from traders who demand immediacy for liquidity or informational needs. Order imbalances are positively autocorrelated, which suggests that traders are herding (Hirshleifer, Subrahmanyam, and Titman, 1994), or spreading their orders out over time (Kyle, 1985), or both. Second, NYSE specialists react to initial order imbalances by altering quotes away from fundamental value in an effort to control inventory. Finally, outside arbitrageurs (by way of market or limit orders) intervene to add market-making capacity by

dependence due to bid-ask bounce. Thus the predictability results of Table 2 indeed appear to be capturing the notion that it takes time for the market to respond to short-term buying and selling pressures.

conducting countervailing trades in the direction opposite to the initial order imbalances. This arbitrage activity takes at least a few minutes.[13]

Evidence supportive of outsider intervention is provided in Table 4, which correlates initial orders with future orders on the other side of the market. In other words, buyer-initiated orders are related to seller-initiated orders in the next time interval, and vice versa. Notice that every single coefficient is positive and strongly significant and that the coefficients increase both in size and in significance with interval length. The t-statistics also are generally much larger in this table than they were for order imbalances in Table 3.

To understand the combination of these results, notice that the serial covariance in order imbalances can be written and decomposed as follows:

$$\begin{aligned} \mathrm{Cov}(OIB_t, OIB_{t-1}) &= \mathrm{Cov}(B_t - S_t, B_{t-1} - S_{t-1}) \\ &= \mathrm{Cov}(B_t, B_{t-1}) + \mathrm{Cov}(S_t, S_{t-1}) - \mathrm{Cov}(B_t, S_{t-1}) - \mathrm{Cov}(S_t, B_{t-1}) \qquad (1) \end{aligned}$$

where B and S denote buyer- and seller-initiated orders, respectively. Table 4 shows that the last two covariances in (1) are always positive and grow with interval length (unreported tests show that all reported correlations at intervals longer than five minutes are statistically greater than the corresponding five minute quantity).[14] However, this growth is not sufficient to overcome the

[13] Note that a competitive but risk-neutral market making sector with minimal costs of maintaining a market would compete away all expected profits and thus allow no predictive relation between imbalances and returns, otherwise money would be "left on the table" (which would be inconsistent with equilibrium). For example, in Kyle (1985), order flows have no predictability for returns. Since we find minimal evidence of predictability and only at short horizons, the evidence indicates that entry costs to the market making sector are low and that risk aversion and/or inventory control play a material role only at intervals of thirty minutes or less.

[14] One might think that the results in Table 4 represent a mechanical consequence of the persistence in total (unsigned) volume. However, positive serial dependence in volume does not necessarily imply that these two terms are positive. Instead, since volume = B+S, volume persistence implies $\mathrm{Cov}(B_t+S_t, B_{t-1}+S_{t-1})>0$, so the unsigned sum of all four covariances in (1) is positive. This can happen even if the last two terms are negative, so long as the first two terms are sufficiently large. The fact that the last two covariances in (1) turn out to be positive actually provides a partial explanation for volume persistence rather than vice versa; viz., countervailing arbitrage activity, which causes the last two terms in equation (1) to be positive, increases the serial correlation in total volume.

even stronger and growing persistence in (naïve) orders of the same sign, which are captured in the first two covariances in (1). Indeed, OIB is still positively dependent out through sixty minutes and even over an entire day. Countervailing arbitrage never offsets it completely. Hence, to render the market informationally efficient, arbitrage trading must be augmented by assistance from two other phenomena, limit orders and specialist actions.

To trace these influences, Table 5 presents a series of multiple regressions with both lagged order imbalances and lagged returns as predictors. In all regressions, the future midpoint return is the dependent variable. Regressions are computed for individual stocks and the table reports the average coefficients. Again, two t-statistics are provided. The first is simply the individual coefficient's average t-statistic across all the regressions. The second is calculated from the cross-sectional array of estimated individual coefficients, correcting for residual cross-correlation. For a given intra-day return interval, all returns over the entire year are included in the regression except for the first interval return on each trading day. (It appears only as a lagged value.) For each return interval, Table 5 reports two regressions which differ by the measure of order imbalance used as a regressor, either OIB# for the number of trades or OIB$ for the dollar amount traded. Both regressions include the lagged return.

Focusing first on 1996, lagged returns have significant negative coefficients in all regressions involving OIB#, for five- ten- and fifteen-minute intervals. Meanwhile, lagged OIB# (OIB$) is significantly positive up to thirty (fifteen) minutes. At sixty minutes, nothing is significant. The magnitudes of the OIB coefficients generally decline monotonically as the regression interval increases.

In 1999, there is a marked reduction in the magnitude as well as the significance of the imbalance coefficients. Lagged OIB# is significant only through fifteen minutes. For 2002, order imbalances are significant at five minutes but not at longer horizons. Lagged returns are not significant over any interval for 1999 and 2002. Again, the magnitudes of the OIB coefficients are generally decreasing as the time interval increases. Overall, the coefficients tend to be smaller for later years, suggesting a more efficient market after the reduction of the minimum tick size.

Note, however, that the Table 5 regressions are not evidence against weak-form market efficiency for very short intervals such as five minutes because outsiders cannot easily observe order imbalances. A trading rule based on the short-period results in Table 5 is therefore feasible only for the specialist and perhaps sophisticated floor traders.

We now turn to regressions stratified by the market capitalization. Our aim is to ascertain if there are any size-related differences in the predictability of imbalances within our sample of the 150 largest stocks in each of the three years. The results are reported in Table 6. Focusing first on 1996, lagged OIB has significant and positive coefficients in all regressions for five-, ten- and fifteen-minute intervals. At the thirty minute interval, OIB# is significant for the two smaller-sized groups, but not for the largest-sized stocks. At sixty minutes, nothing is significant. The magnitudes of the coefficients are smaller for larger firms supporting the notion that the market for larger stocks is more efficient. In addition, as before, the coefficients continue to decrease with the regression horizon.

In 1999, lagged OIB is significant only through fifteen minutes for the mid-size and smaller-size stocks. At five-minute intervals, OIB is a significant positive predictor of future returns only for the two smaller size groups. For 2002, order imbalances are significant at five minutes for all size groups. There is some significance over 10 and 15 minutes as well but mainly for the mid-size and the smaller-size groups. Lagged returns are significant at five- and ten-minute intervals and even fifteen-minute intervals for larger stocks. At longer intervals nothing is significant. In general (though not in every case), for both years, the magnitudes of the lagged OIB coefficients are smaller for larger firms.

All of these results are consistent with the dual notions that (i) imbalances in larger stocks are offset more efficaciously by market makers, so that predictability of imbalances is removed sooner, and that (ii) the speed of convergence to efficiency has increased after the tick size reduction in that imbalances lose predictability at shorter horizons in 1999 and 2002 than in 1996.

Are the results above economically significant? To answer this question by illustration, consider the five-minute interval during the year 2002; the coefficients on $OIB\#_{t-1}$ (scaled by 10^5) are respectively 1.07, 1.91, and 1.97 for large, medium and small stocks. The cross-sectional averages of the standard deviations of $OIB\#_{t-1}$ (not reported in the tables) are 14.88, 11.73, and 8.83, respectively. Assuming (1) there is a one standard deviation shock every day for 250 trading days during a year, and (2) an agent with knowledge of the imbalance trades once a day on the one standard deviation shock, the annualized gross returns are, respectively, 3.98%, 5.60%, and 4.44%. Such a trader has an open position for only five minutes during the day, so the returns are very large per unit of time. However, the returns may not be substantial after accounting for transaction costs, given the frequency of trading that would accompany any practical trading rule designed to take advantage of the predictability.

An important issue is whether the smaller number of observations at the sixty-minute interval explains the lack of significance of lagged OIB at this horizon. We do not think this is the entire explanation for two reasons. First, the coefficients of lagged OIB at longer horizons are generally smaller than those at shorter horizons, suggesting that the market making sector does eventually offset short-term imbalances. As another check, we perform a back-of-the-envelope estimation of the coefficient on lagged OIB under the supposition that market makers to not offset any part of the imbalance. We do this by taking the contemporaneous regression coefficient of return on OIB and multiplying it by the coefficient obtained from the serial regression of OIB on its lag. We then compare this benchmark coefficient with the coefficient reported in Table 6. In all cases, the Table 6 coefficient is smaller than the benchmark by magnitudes ranging from 82%-90%, suggesting that market makers do indeed alleviate price pressure.[15]

The predictability of returns using imbalances at intervals of up to thirty minutes indicates that it takes time for floor traders and other arbitrageurs to compute and react to imbalances. Since imbalances are not public information, outside traders in effect have to apply a rule akin to the Lee and Ready (1991) algorithm to deduce imbalance information, which is a possible reason for the modest predictability of returns out to thirty minutes. Our results suggest that wider dissemination of order imbalances by the exchange could bring faster convergence to efficiency

[15] We thank the referee for suggesting this calculation.

Overall, the results suggest that outsider arbitrage activity (as documented in Table 4), along with specialists' actions and limit orders, increase market-making capacity and remove the influence of order imbalances within a half-hour and usually within a much shorter time. This rapid convergence to efficiency is all the more impressive when one considers that order imbalances persist strongly over much longer intervals, even over days. Evidently, astute agents have little difficulty in forecasting the persistence in order imbalances and conducting trades sufficient to eliminate its impact on prices after just a few minutes.

The only remaining puzzle is why lagged returns are significantly negatively related to future returns in the multiple regressions of Table 6 when they are virtually unrelated to future returns in the univariate regressions of Table 3. At least two explanations seem possible and they are not mutually exclusive. First, prices might overreact to the initial order imbalance in t-1 and then rebound from the overreaction in t. Consequently, the return in t-1 is negatively related to the return in t (conditional on knowing OIB_{t-1}). It would appear, however, that this explanation implies that the specialist is not trading optimally on his privileged information about OIB.

A second explanation is that the specialist responds to an initial order imbalance by <u>intentionally</u> adjusting the bid and ask prices by more than he knows is appropriate. This might help maintain control of inventory risk. By moving prices away from equilibrium, the specialist attracts outside arbitrageurs who enter and help "lean against the wind" of the highly persistent order imbalances. Notice that the negative coefficients of lagged returns, though significant at short intervals, are rather small in absolute magnitude. This suggests that the specialist is not leaving all that much money on the table. It might be well worth giving up a modest profit to attract the attention and aid of astute outsiders.

IV. Conclusions

The long and continuing debate about financial market efficiency has been relatively silent about the behavior of actual traders. Somehow, perhaps unwittingly, they act collectively to push markets toward efficiency. Except in an idealized theoretical world, this cannot happen instantaneously. There must be some time interval, albeit very short, over which the actions of efficiency-creating traders remain incomplete. A central goal of this paper is to present evidence about this important issue, the speed of convergence to market efficiency.

We first examine weak-form efficiency (Fama, 1970), which is concerned only with serial dependence in returns. Of course, even weak-form efficiency might not be attained immediately. But using intra-day returns for 150 NYSE stocks during calendar years 1996, 1999, and 2002, we find that weak-form efficiency <u>does</u> appear to prevail over intervals from five minutes to one day. There is evidence, however, that the market is not strong-form efficient over short intervals of a few minutes. Order imbalances are highly positively dependent over both short and long intervals and imbalances predict future returns over very short intervals. But order imbalances are known unambiguously only by the NYSE specialist.

Conditional on knowing the current order imbalance, returns are negatively serially dependent over intervals up to ten minutes. This conditional negative dependence in returns is consistent with NYSE specialists altering quotes away from fundamentals for the purpose of inventory risk control, while awaiting help from countervailing traders. Indeed, there is strong evidence that outside traders soon become aware of price-moving order imbalances and undertake countervailing trades. In no more than thirty minutes, order imbalances lose their predictive ability and returns are no longer negatively dependent.

These results make one wonder about the existence of market anomalies and inefficiencies in general. There is no significant evidence of inefficiency at intervals of thirty minutes, yet the extensive literature on long-term anomalies[16] documents momentum at intervals of six months and beyond. How markets deteriorate from weak-form efficiency at very short horizons to predictability at long horizons seems a worthwhile area for further research.

[16]E.g., see Jegadeesh and Titman, 1993. Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998, 2001), and Hong and Stein (1999) attempt to explain momentum and other inefficiencies using models with irrational investors.

References

Ball, Clifford A., and Tarun Chordia, 2001, True spreads and equilibrium prices, *Journal of Finance* 56, 1801-1835.

Barber, Brad M., and Terence Odean, 2000, Trading is hazardous to your health: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.

Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49 307-343.

Benartzi, Shlomo, and Richard Thaler, 2001, Naïve diversification strategies in retirement saving plans, *American Economic Review* 91, 79-98.

Busse, Jeffrey, and Clifton Green, 2002, Market efficiency in real time, *Journal of Financial Economics* 65, 415-437.

Chakrabarti, Rajesh, and Richard Roll, 1999, Learning from others, reacting, and market quality, *Journal of Financial Markets* 2, 153-178.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of Financial Economics*, 56, 3-28.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, Order imbalance, liquidity and market returns, *Journal of Financial Economics* 65, 111-130.

Chordia, Tarun, and Avanidhar Subrahmanyam, 1995, Market making, the tick size, and payment-for-order flow: Theory and evidence, *Journal of Business* 68, 543-575.

Copeland, Thomas E., 1976, A model of asset trading under the assumption of sequential information arrival, *Journal of Finance* 31, 1149-1168.

Cornell, Bradford and Richard Roll, 1981, Strategies for pairwise competitions in markets and organizations, *Bell Journal of Economics* 12, 201-213.

Cramér, Harald, 1954, *Mathematical Methods of Statistics*, (Princeton: Princeton University Press).

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839-1885.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921-965.

Epps, Thomas W., 1979, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* 74, 291-298.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383-417.

Fama, Eugene F, and French, Kenneth R., 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

Garbade, Kenneth D., and Zvi Lieber, 1977, On the independence of transactions on the New York Stock Exchange, *Journal of Banking and Finance* 1, 151-172.

Grossman, Sanford J., 1976, On the efficiency of competitive stock-markets where traders have diverse information, *Journal of Finance* 31, 573-585.

Grossman, Sanford J. and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Hillmer, S.C., and P. L. Yu, 1979, The market speed of adjustment to new information, *Journal of Financial Economics* 7, 321-345.

Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman, 1994, Security analysis and trading patterns when some investors receive information before others, *Journal of Finance* 49, 1665-1698.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in assets markets, *Journal of Finance* 54, 2143-2184.

Jegadeesh, Narasimhan, and Sheridan Titman, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Jones, Charles, Gautam Kaul, and Marc Lipson, 1994, Transactions, volume, and volatility, *Review of Financial Studies* 7, 631-651.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.

Lee, Charles, and M. Ready, 1991, Inferring trade direction from intra-day data, *Journal of Finance* 46, 733-747.

Marriott, F. H. C., and J. A. Pope, 1954, Bias in the estimation of autocorrelations, *Biometrika* 41, 390-402.

Odean, Terrance, 1999, Do investors trade too much? *American Economic Review* 89, 1279-1298.

Patell, James M., and Mark A. Wolfson, 1984, The intra-day speed of adjustment of stock prices to earnings and dividend announcements, *Journal of Financial Economics* 13, 223-252.

Roll, Richard W., 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127-1139.

Schwert, G. William, 2001, Anomalies and market efficiency, Chapter 17 in George Constantinides, Milton Harris, and René Stulz, eds., *Handbook of the Economics of Finance*, (North-Holland: Amsterdam).

Table 1

Correlation Coefficients at a Daily Horizon
for Returns and Order Imbalances

Trade returns are computed from the last transaction price of each day and midpoint returns are computed from the average of the bid-ask quotes associated with the last transaction of each day, for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. Trade returns are from CRSP. Bid-Ask quotes and order imbalances (OIB) are from the NYSE TAQ data base. OIB# is the number of buyer-initiated less the number of seller-initiated trades during the same day as the return; OIBSh is the number of buyer-initiated shares purchased less the number of seller-initiated shares sold that day; OIB$ is the total dollars paid by buyer-initiators less the total dollars received by seller-initiators that day. The product-moment correlation coefficient is reported along with a t-statistic computed from the cross-sectional distribution of correlation coefficients.

	Trade Return$_t$	Midpoint Return$_t$	$\mathrm{OIB\#}_t$	OIBSh_t	$\mathrm{OIB\$}_t$	Trade Return$_t$	Midpoint Return$_t$	$\mathrm{OIB\#}_t$	OIBSh_t	$\mathrm{OIB\$}_t$	Trade Return$_t$	Midpoint Return$_t$	$\mathrm{OIB\#}_t$	OIBSh_t	$\mathrm{OIB\$}_t$
	1996					1999					2002				
Return$_{t-1}$[17]	0.002 (0.38)	0.007 (1.15)				-0.000 (-0.03)	0.010 (1.30)				-0.030 (-4.98)	-0.027 (-4.56)			
$\mathrm{OIB\#}_t$	0.315 (22.32)	0.314 (22.22)				0.215 (12.18)	0.214 (11.98)				0.232 (13.75)	0.230 (13.52)			
$\mathrm{OIB\#}_{t-1}$	0.055 (9.64)	0.056 (9.79)	0.336 (19.77)			0.019 (3.07)	0.022 (3.68)	0.371 (27.65)			0.022 (3.53)	0.022 (3.43)	0.249 (18.05)		
OIBSh_t	0.454 (40.80)	0.460 (41.12)	0.305 (21.91)			0.462 (37.21)	0.468 (37.28)	0.279 (16.32)			0.327 (25.82)	0.326 (25.56)	0.503 (42.82)		
OIBSh_{t-1}	0.002 (0.31)	-0.002 (-0.38)	-0.055 (-5.20)	0.125 (14.09)		-0.016 (-2.53)	-0.014 (-2.16)	-0.018 (-1.53)	0.204 (18.46)		0.006 (0.93)	0.006 (1.00)	0.088 (8.54)	0.179 (16.84)	
$\mathrm{OIB\$}_t$	0.452 (41.34)	0.458 (41.57)	0.303 (22.04)	0.991 (609.3)		0.457 (37.13)	0.462 (37.14)	0.280 (16.64)	0.981 (437.8)		0.321 (27.67)	0.321 (27.43)	0.488 (44.18)	0.965 (201.4)	
$\mathrm{OIB\$}_{t-1}$	0.000 (0.08)	-0.003 (-0.63)	-0.057 (-5.49)	0.124 (14.09)	0.128 (14.22)	-0.020 (-3.18)	-0.018 (-2.85)	-0.013 (-1.09)	0.197 (18.37)	0.209 (18.39)	-0.007 (-1.20)	-0.006 (-1.11)	0.075 (8.27)	0.161 (16.63)	0.180 (17.22)

[17] Trade (Midpoint) Return$_{t-1}$ in the Trade Return$_t$ (Midpoint Return$_t$) column.

Table 2

**Univariate Regressions Predicting Returns
for Intervals from Five to Sixty Minutes**

The dependent variable is the next period's mid-point return. Daily returns and order imbalances are obtained from the NYSE TAQ data base for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. The return is computed from the midpoint of the bid-ask spread associated with the transaction nearest the end of an intra-day time interval of fixed length. OIB# is the number of buyer-initiated less the number of seller-initiated trades during the same time interval as the return. OIB$ is the total dollar amount expended by buyer-initiators less the total dollar amount received by seller-initiators during that interval. The first interval of each day is excluded and all other interval observations during each calendar year, (either 1996, 1999 or 2002), are included in the same regression. A separate regression is estimated for each individual stock. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The second number (in parentheses) is the average t-statistic from the individual regressions. The third number (also in parentheses) is a t-statistic computed from the cross-sectional distribution of the estimated coefficients adjusting for cross-correlation in the residuals. The fourth number is the cross-sectional average adjusted R-square in percent. To adjust the units for presentation, the coefficients for OIB# and OIB$ have been multiplied by 10^5.

Table 2, (Continued)

	Return interval (minutes)				
Explanatory variable	Five	Ten	Fifteen	Thirty	Sixty
1996					
Midpoint $Return_{t-1}$	-0.003 (-0.50) (-0.12) 0.45	-0.025 (-2.52) (-0.98) 0.39	-0.027 (-2.22) (-1.03) 0.41	0.010 (0.40) (0.42) 0.23	0.022 (0.60) (0.61) 0.33
$OIB\#_{t-1}$	7.73 (18.04) (3.91) 2.61	3.72 (6.02) (2.58) 0.79	2.50 (3.36) (2.17) 0.44	1.81 (2.27) (2.23) 0.28	0.94 (1.17) (0.83) 0.22
$OIB\$_{t-1}$	21.40 (11.91) (3.54) 1.04	12.03 (4.64) (2.40) 0.38	9.04 (2.88) (2.36) 0.21	8.90 (1.90) (0.84) 0.22	4.07 (1.20) (0.29) 0.18
1999					
Midpoint $Return_{t-1}$	0.016 (2.15) (0.99) 0.19	0.010 (1.01) (0.70) 0.14	0.008 (0.66) (0.58) 0.13	0.009 (0.49) (0.56) 0.14	0.021 (0.76) (1.07) 0.19
$OIB\#_{t-1}$	2.83 (8.56) (2.78) 0.56	1.42 (3.26) (2.13) 0.20	0.94 (2.04) (1.88) 0.12	0.50 (0.94) (0.98) 0.10	0.53 (0.81) (0.97) 0.15
$OIB\$_{t-1}$	7.51 (4.91) (2.54) 0.18	4.15 (2.01) (1.99) 0.07	3.18 (1.45) (2.02) 0.05	2.43 (0.92) (1.25) 0.07	2.65 (0.79) (1.02) 0.12
2002					
Midpoint $Return_{t-1}$	-0.013 (-1.74) (-0.59) 0.21	-0.023 (-2.24) (-0.96) 0.25	-0.016 (-1.23) (-0.68) 0.21	-0.004 (-0.21) (-0.19) 0.12	0.001 (0.03) (0.04) 0.12
$OIB\#_{t-1}$	1.29 (5.93) (2.02) 0.28	0.41 (1.34) (0.95) 0.08	0.31 (0.89) (0.76) 0.08	0.16 (0.38) (0.37) 0.07	0.18 (0.39) (0.42) 0.06
$OIB\$_{t-1}$	7.11 (3.63) (1.83) 0.11	3.27 (1.18) (1.12) 0.04	2.93 (0.93) (0.98) 0.05	1.88 (0.45) (0.48) 0.05	1.10 (0.27) (0.24) 0.06

Table 3
Univariate Regressions Predicting Order Imbalances for Intervals from Five to Sixty Minutes

The dependent variable is order imbalance in numbers of transactions for predictor $OIB\#_{t-1}$ and in dollars for predictor $OIB\$_{t-1}$. OIB# is the number of buyer-initiated less the number of seller-initiated trades during the same time interval as the return. OIB$ is the total dollar amount expended by buyer-initiators less the total dollar amount received by seller-initiators during that interval. These imbalance measures are computed for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. The first interval of each day is excluded and all other interval observations during each calendar year, (either 1996, 1999 or 2002), are included in the same regression. A separate regression is estimated for each individual stock. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The second number (in parentheses) is the average t-statistic from the individual regressions. The third number (also in parentheses) is a t-statistic computed from the cross-sectional distribution of the estimated coefficients adjusting for cross-correlation in the residuals. The fourth number is the cross-sectional average adjusted R-square in percent.

	Imbalance interval (minutes)				
Explanatory variable	Five	Ten	Fifteen	Thirty	Sixty
1996					
$OIB\#_{t-1}$	0.219 (28.90) (11.68) 5.45	0.209 (19.49) (6.14) 5.52	0.216 (16.58) (5.06) 6.26	0.256 (14.11) (4.79) 8.93	0.298 (11.39) (4.12) 17.38
$OIB\$_{t-1}$	0.067 (8.39) (14.04) 0.54	0.067 (5.80) (10.39) 0.56	0.065 (4.57) (8.15) 0.58	0.077 (3.59) (5.73) 0.85	0.087 (2.85) (4.98) 1.18
1999					
$OIB\#_{t-1}$	0.180 (26.68) (7.50) 4.28	0.217 (22.93) (5.88) 6.45	0.238 (20.63) (5.60) 7.80	0.282 (17.29) (5.29) 10.91	0.353 (14.70) (5.63) 15.91
$OIB\$_{t-1}$	0.072 (9.87) (11.23) 0.68	0.089 (8.61) (10.51) 1.04	0.101 (8.01) (9.38) 1.41	0.122 (6.64) (8.72) 2.00	0.154 (5.60) (7.47) 3.27
2002					
$OIB\#_{t-1}$	0.164 (23.03) (6.57) 3.21	0.199 (19.70) (6.02) 4.71	0.229 (18.60) (6.08) 6.20	0.281 (16.14) (6.15) 9.33	0.343 (13.22) (6.73) 13.28
$OIB\$_{t-1}$	0.068 (9.13) (14.27) 0.51	0.087 (8.06) (9.93) 0.85	0.105 (7.93) (10.14) 1.23	0.133 (6.86) (8.86) 1.98	0.170 (5.74) (7.66) 3.19

Table 4

Cross-Autocorrelation Coefficients between Buying and Selling over Intra-Day Horizons

Buys and sells are estimated from NYSE TAQ data for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. The first interval of each day is excluded. The reported product-moment autocorrelation coefficient is between orders on one side of the market during a trading interval and orders on the opposite side of the market during the subsequent interval. In other words, buy orders are correlated with subsequent sell orders, and vice versa. The t-statistic (in parentheses) is computed from the cross-sectional distribution of correlation coefficients.

Panel A: Number of Orders

	Initial Order type	Five	Ten	Fifteen	Thirty	Sixty
		Trading Interval (minutes)				
1996	Buy	0.137 (13.02)	0.206 (17.53)	0.251 (20.60)	0.316 (24.75)	0.343 (26.60)
	Sell	0.126 (11.37)	0.205 (17.08)	0.255 (20.61)	0.322 (25.36)	0.355 (28.68)
1999	Buy	0.292 (23.48)	0.384 (28.26)	0.434 (31.86)	0.493 (36.25)	0.475 (32.24)
	Sell	0.292 (22.74)	0.384 (27.97)	0.433 (31.73)	0.496 (38.26)	0.485 (35.11)
2002	Buy	0.322 (24.24)	0.417 (30.04)	0.465 (34.09)	0.523 (39.85)	0.499 (35.09)
	Sell	0.322 (24.36)	0.420 (30.06)	0.472 (34.04)	0.535 (39.95)	0.522 (36.72)

Panel B: Dollar Value of Orders

	Initial Order type	Five	Ten	Fifteen	Thirty	Sixty
		Trading Interval (minutes)				
1996	Buy	0.067 (11.66)	0.110 (14.21)	0.141 (16.07)	0.190 (18.42)	0.232 (20.55)
	Sell	0.059 (9.96)	0.106 (13.06)	0.138 (15.13)	0.184 (17.11)	0.230 (21.11)
1999	Buy	0.142 (15.44)	0.215 (18.55)	0.255 (20.35)	0.318 (23.91)	0.348 (28.36)
	Sell	0.138 (15.24)	0.209 (18.27)	0.251 (20.46)	0.314 (24.23)	0.339 (27.65)
2002	Buy	0.226 (27.11)	0.321 (31.24)	0.373 (34.72)	0.435 (40.22)	0.435 (37.88)
	Sell	0.225 (26.13)	0.320 (29.44)	0.371 (32.50)	0.440 (37.47)	0.440 (37.85)

Table 5

Multiple Regressions of Returns on Lagged Returns and Two Different Measures of Lagged Order Imbalance for Return Intervals from Five to Sixty Minutes

Daily returns and order imbalances are obtained from the NYSE TAQ data base for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. The return is computed from the midpoint of the bid-ask spread associated with the transaction nearest the end of an intra-day time interval of fixed length. OIB# is the number of buyer-initiated less the number of seller-initiated trades during the same time interval as the return. OIB$ is the total dollar amount expended by buyer-initiators less the total dollar amount received by seller-initiators during that interval. The first interval of each day is excluded and all other interval observations during each calendar year, (either 1996 or 1999 or 2002), are included in the same regression. A separate regression is estimated for each individual stock. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The second number (in parentheses) is the average t-statistic from the individual regressions. The third number (also in parentheses) is a t-statistic computed from the cross-sectional distribution of the estimated coefficients adjusting for cross-correlation in the residuals. The R^2 is the cross-sectional average adjusted R-square in percent. To adjust the units for presentation, the coefficients for OIB# and OIB$ have been multiplied by 10^5.

Table 5 (Continued)

Explanatory Variable	Return Interval (minutes)									
	Five		Ten		Fifteen		Thirty		Sixty	
Dependent Variable is the Midpoint $Return_t$, 1996										
Midpoint $Return_{t-1}$	-0.069 (-7.93) (-3.31)	-0.035 (-4.37) (-1.34)	-0.070 (-5.66) (-3.05)	-0.051 (-4.51) (-1.86)	-0.063 (-4.13) (-2.43)	-0.051 (-3.62) (-1.79)	-0.014 (-0.68) (-0.56)	-0.008 (-0.45) (-0.30)	0.007 (0.10) (0.50)	0.005 (0.04) (0.14)
$OIB\#_{t-1}$	8.92 (19.12) (3.75)		4.88 (7.53) (3.01)		3.49 (4.59) (2.70)		2.03 (2.24) (2.00)		0.84 (0.96) (1.60)	
$OIB\$_{t-1}$		23.48 (12.61) (3.64)		15.26 (5.97) (2.71)		12.51 (4.06) (2.77)		8.95 (1.84) (1.09)		5.01 (0.99) (0.49)
R^2	3.32	1.59	1.42	0.94	1.02	0.78	0.53	0.43	0.52	0.43
Dependent Variable is the Midpoint $Return_t$, 1999										
Midpoint $Return_{t-1}$	-0.011 (-1.44) (-0.76)	0.005 (0.68) (0.31)	-0.005 (-0.46) (-0.34)	0.004 (0.35) (0.26)	-0.004 (-0.26) (-0.25)	0.002 (0.13) (0.13)	0.001 (0.08) (0.09)	0.003 (0.12) (0.17)	0.013 (0.43) (0.63)	0.014 (0.45) (0.66)
$OIB\#_{t-1}$	3.03 (8.29) (2.72)		1.52 (3.08) (2.24)		1.07 (1.99) (2.12)		0.56 (0.86) (1.22)		0.46 (0.76) (0.86)	
$OIB\$_{t-1}$		7.20 (4.53) (2.48)		3.88 (1.75) (1.91)		3.11 (1.27) (1.87)		2.38 (0.75) (1.24)		1.94 (0.50) (0.69)
R^2	0.70	0.34	0.30	0.20	0.23	0.17	0.21	0.19	0.29	0.26
Dependent Variable is the Midpoint $Return_t$, 2002										
Midpoint $Return_{t-1}$	-0.037 (-4.57) (-1.71)	-0.021 (-2.75) (-0.97)	-0.035 (-3.08) (-1.43)	-0.029 (-2.67) (-1.69)	-0.025 (-1.77) (-1.03)	-0.021 (-1.58) (-0.90)	-0.008 (-0.38) (-0.34)	-0.007 (-0.37) (-0.35)	-0.004 (-0.13) (-0.14)	-0.002 (-0.06) (-0.08)
$OIB\#_{t-1}$	1.65 (7.28) (2.76)		0.75 (2.47) (1.73)		0.52 (1.51) (1.34)		0.21 (0.50) (0.49)		0.23 (0.43) (0.52)	
$OIB\$_{t-1}$		8.24 (4.26) (2.64)		5.15 (1.86) (1.20)		4.35 (1.32) (1.54)		2.27 (0.54) (0.58)		1.18 (0.28) (0.26)
R^2	0.54	0.32	0.35	0.30	0.28	0.93	0.18	0.16	0.17	0.15

Table 6

Multiple Regressions of Returns on Lagged Returns and Two Different Measures of Lagged Order Imbalance for Return Intervals from Five to Sixty Minutes, by Size Grouping

Daily returns and order imbalances are obtained from the NYSE TAQ data base for the 150 largest NYSE stocks (by market capitalization) at the beginning of each of the relevant years. The three groups consist of the top 50 (large stocks), the next 50 (mid-cap stocks), and the last 50 (smaller stocks), ranked by market capitalization. The return is computed from the midpoint of the bid-ask spread associated with the transaction nearest the end of an intra-day time interval of fixed length. OIB# is the number of buyer-initiated less the number of seller-initiated trades during the same time interval as the return. OIB$ is the total dollar amount expended by buyer-initiators less the total dollar amount received by seller-initiators during that interval. The first interval of each day is excluded and all other interval observations during each calendar year, (either 1996 or 1999 or 2002), are included in the same regression. A separate regression is estimated for each individual stock. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The first number in each cell is the cross-sectional mean of the estimated regression coefficient. The second number (in parentheses) is the average t-statistic from the individual regressions. The third number (also in parentheses) is a t-statistic computed from the cross-sectional distribution of the estimated coefficients adjusting for cross-correlation in the residuals. The R^2 is the cross-sectional average adjusted R-square in percent. To adjust the units for presentation, the coefficients for OIB# and OIB$ have been multiplied by 10^5.

Table 6 (Continued)

Explanatory Variable	Return Interval (minutes)									
	Five		Ten		Fifteen		Thirty		Sixty	
Dependent Variable is the Midpoint Return$_t$, Large Stocks, 1996										
Midpoint Return$_{t-1}$	-0.050 (-6.39) (-2.67)	-0.039 (-5.04) (-1.65)	-0.068 (-5.94) (-3.47)	-0.069 (-6.06) (-3.26)	-0.063 (-4.45) (-2.98)	-0.068 (-4.79) (-3.13)	0.001 (0.06) (0.05)	-0.000 (-0.04) (-0.01)	0.024 (0.75) (0.89)	0.021 (0.59) (0.69)
OIB#$_{t-1}$	3.87 (15.35) (3.61)		1.87 (5.46) (2.72)		1.33 (3.25) (2.43)		0.92 (1.91) (1.96)		0.56 (1.07) (1.25)	
OIB\$$_{t-1}$		13.42 (12.98) (4.05)		8.53 (5.93) (3.44)		7.28 (4.19) (3.07)		4.43 (1.81) (2.37)		3.70 (0.95) (1.38)
R^2	1.59	1.16	0.71	0.71	0.57	0.64	0.35	0.31	0.45	0.43
Dependent Variable is the Midpoint Return$_t$, Mid-cap Stocks, 1996										
Midpoint Return$_{t-1}$	-0.060 (-7.25) (-3.29)	-0.016 (-2.37) (-0.59)	-0.061 (-5.02) (-2.73)	-0.038 (-3.41) (-1.36)	-0.054 (-3.53) (-1.99)	-0.040 (-2.89) (-1.39)	-0.015 (-0.69) (-0.71)	-0.008 (-0.40) (-0.35)	-0.008 (0.76) (-0.26)	-0.012 (-0.34) (-0.37)
OIB#$_{t-1}$	9.11 (20.47) (6.26)		4.87 (7.98) (4.71)		3.30 (4.65) (3.74)		1.90 (2.01) (2.51)		0.98 (0.76) (1.15)	
OIB\$$_{t-1}$		21.88 (12.45) (4.71)		15.11 (6.02) (3.72)		12.13 (4.03) (3.31)		6.80 (1.65) (2.37)		6.31 (1.04) (1.76)
R^2	3.36	1.53	1.31	0.88	0.88		0.39	0.31	0.47	0.47
Dependent Variable is the Midpoint Return$_t$, Small Stocks, 1996										
Midpoint Return$_{t-1}$	-0.096 (-10.04) (-4.30)	-0.049 (-5.62) (-1.76)	-0.081 (-6.03) (-2.97)	-0.049 (-4.08) (-1.46)	-0.070 (-4.38) (-2.38)	-0.046 (-3.19) (-1.27)	-0.030 (-1.42) (-0.95)	-0.015 (-0.90) (-0.45)	-0.000 (-0.26) (-0.17)	-0.005 (-0.10) (0.17)
OIB#$_{t-1}$	13.81 (21.51) (5.52)		7.73 (9.13) (5.17)		5.91 (5.87 (4.44)		3.17 (2.82) (2.50)		1.27 (1.12) (0.58)	
OIB\$$_{t-1}$		35.01 (12.41) (4.74)		22.11 (5.97) (3.00)		18.04 (4.02) (3.26)		9.37 (2.02) (1.79)		3.13 (0.92) (0.14)
R^2	5.02	2.08	2.20	1.27	1.73	1.05	1.01	0.72	0.01	0.00

Table 6 (Continued)

Explanatory Variable	Return Interval (minutes)									
	Five		Ten		Fifteen		Thirty		Sixty	
Dependent Variable is the Midpoint $Return_t$, Large Stocks, 1999										
Midpoint $Return_{t-1}$	-0.012 (-1.57) (-0.72)	-0.010 (-1.21) (-0.53)	-0.004 (-0.79) (-0.26)	-0.003 (-0.25) (-0.19)	0.001 (0.11) (0.05)	0.002 (0.13) (0.10)	0.013 (0.66) (0.90)	0.010 (0.48) (0.67)	0.027 (0.89) (1.20)	0.026 (0.85) (1.07)
$OIB\#_{t-1}$	0.906 (4.44) (1.76)		0.490 (1.88) (1.45)		0.445 (1.70) (1.54)		0.285 (0.97) (1.17)		0.398 (0.91) (1.31)	
$OIB\$_{t-1}$		2.66 (9.22) (2.15)		1.47 (1.44) (1.74)		1.32 (1.14) (1.34)		1.78 (1.07) (1.39)		1.73 (0.67) (0.93)
R^2	0.29	0.22	0.15	0.13	0.15	0.12	0.31	0.17	0.34	0.32
Dependent Variable is the Midpoint $Return_t$, Mid-cap Stocks, 1999										
Midpoint $Return_{t-1}$	-0.007 (-0.89) (-0.54)	0.012 (1.57) (0.85)	-0.004 (-0.42) (-0.31)	0.006 (0.53) (0.42)	-0.004 (-0.33) (-0.27)	0.001 (0.08) (0.10)	-0.007 (-0.36) (-0.47)	-0.004 (-0.22) (-0.22)	0.006 (0.22) (0.30)	0.008 (0.26) (0.38)
$OIB\#_{t-1}$	2.88 (8.84) (4.29)		1.39 (3.26) (3.17)		0.940 (1.84) (2.36)		0.534 (0.76) (1.30)		0.410 (0.45) (0.67)	
$OIB\$_{t-1}$		6.66 (4.77) (3.72)		3.52 (1.77) (2.77)		3.18 (1.32) (2.82)		1.72 (0.50) (1.30)		1.71 (0.41) (0.84)
R^2	0.66	0.32	0.28	0.18	0.23	0.18	0.18	0.13	0.28	0.22
Dependent Variable is the Midpoint $Return_t$, Small Stocks, 1999										
Midpoint $Return_{t-1}$	-0.014 (-1.77) (-0.89)	0.014 (1.77) (0.85)	-0.007 (-0.68) (-0.45)	0.008 (0.71) (0.52)	-0.008 (-0.56) (-0.54)	0.003 (0.18) (0.20)	-0.001 (-0.06) (-0.06)	0.003 (0.10) (0.14)	0.006 (0.18) (0.28)	0.008 (0.23) (0.40)
$OIB\#_{t-1}$	5.29 (11.52) (4.72)		2.67 (4.11) (3.64)		1.81 (2.38) (3.26)		0.870 (0.84) (1.59)		0.586 (0.42) (0.89)	
$OIB\$_{t-1}$		12.25 (15.72) (3.67)		6.69 (2.08) (2.75)		4.82 (1.31) (2.21)		3.62 (0.69) (1.34)		2.42 (0.42) (0.58)
R^2	1.16	0.48	0.48	0.28	0.31	0.21	0.28	0.26	0.24	0.24

Table 6 (Continued)

Explanatory Variable	Return Interval (minutes)									
	Five		Ten		Fifteen		Thirty		Sixty	
Dependent Variable is the Midpoint Return$_t$, Large Stocks, 2002										
Midpoint Return$_{t-1}$	-0.050 (-6.28) (-3.39)	-0.041 (-5.37) (-2.34)	-0.042 (-3.81) (-2.21)	-0.040 (-3.78) (-2.09)	-0.031 (-2.26) (-1.90)	-0.031 (-2.33) (-1.85)	-0.011 (-0.60) (-0.65)	-0.013 (-0.69) (-0.75)	-0.002 (-0.05) (-0.09)	-0.005 (-0.15) (-0.20)
OIB#$_{t-1}$	1.07 (5.86) (2.21)		0.487 (2.05) (1.64)		0.322 (1.22) (1.08)		0.071 (0.37) (0.29)		0.061 (0.24) (0.24)	
OIB\$$_{t-1}$		5.58 (4.30) (2.42)		3.74 (2.15) (2.44)		2.96 (1.44) (1.91)		1.20 (0.56) (0.68)		0.861 (0.40) (0.36)
R^2	0.45	0.34	0.33	0.32	0.24	0.21	0.13	0.13	0.15	0.15
Dependent Variable is the Midpoint Return$_t$, Mid-cap Stocks, 2002										
Midpoint Return$_{t-1}$	0.026 (-3.30) (-1.86)	-0.007 (-0.91) (-0.51)	-0.026 (-2.33) (1.59)	-0.018 (-1.70) (-1.11)	-0.022 (-1.64) (-1.48)	-0.016 (-1.27) (-1.08)	-0.004 (-0.20) (-0.23)	-0.003 (-0.16) (-0.21)	-0.010 (-0.36) (-0.47)	-0.004 (-0.15) (-0.19)
OIB#$_{t-1}$	1.91 (8.25) (4.54)		0.822 (2.69) (2.62)		0.680 (1.87) (2.21)		0.230 (0.48) (0.70)		0.348 (0.57) (0.88)	
OIB\$$_{t-1}$		8.42 (4.31) (4.11)		6.70 (1.66) (2.52)		4.79 (1.40) (2.32)		2.30 (0.56 (0.96)		0.030 (0.15) (0.02)
R^2	0.53	0.25	0.30	0.23	0.25	0.21	0.17	0.14	0.25	0.00
Dependent Variable is the Midpoint Return$_t$, Small Stocks, 2002										
Midpoint Return$_{t-1}$	-0.035 (-4.08) (-2.11)	-0.016 (-1.93) (-1.01)	-0.036 (-3.07) (-2.00)	-0.028 (-2.50) (-1.61)	-0.021 (-1.36) (-1.00)	-0.016 (-1.11) (-0.86)	-0.008 (-0.31) (-0.45)	-0.006 (-0.23) (-0.35)	0.000 (0.02) (0.02)	0.003 (0.12) (0.18)
OIB#$_{t-1}$	1.97 (7.69) (5.07)		0.322 (1.22) (1.07)		0.569 (1.43) (2.24)		0.301 (0.60) (0.94)		0.273 (0.49) (1.02)	
OIB\$$_{t-1}$		10.73 (4.17) (4.72)		7.08 (1.77) (2.62)		5.29 (1.12) (2.17)		3.24 (0.47) (0.87)		2.64 (0.30) (0.65)
R^2	0.64	0.37	0.24	0.35	0.36	0.34	0.25	0.21	0.13	0.15