JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS VOL. 40, NO. 1, MARCH 2005
COPYRIGHT 2005, SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

Correlated Order Flow: Pervasiveness, Sources, and Pricing Effects

Jarrad Harford and Aditya Kaul*

Abstract

We examine the importance of indexing, industry, and broad market forces in driving common effects in order flow, returns, and trading costs. Common effects are strong for order flow and returns in a sample of S&P 500 stocks, but are weak in a sample of non-index stocks and for trading costs in both samples. Industry and broad market effects exist in order flow for both samples, but indexing effects are dominant. Correlated order flow drives common effects in returns and, to a lesser extent, those in trading costs. An event study of the effect of index addition on order flow and return comovement reinforces these conclusions. Our results show that common effects are not pervasive and have implications for diversification strategies and price formation models.

I. Introduction

Recent research has uncovered common effects in trading activity, returns, and trading costs for individual stocks. Hasbrouck and Seppi (2001) use principal components to reveal comovement in intraday order flow, returns, and trading costs across the 30 Dow Jones constituents, and canonical correlations to establish that the common factors are related.[1] Chordia, Roll, and Subrahmanyam (2000) document the existence of market and industry factors in average daily changes in five measures of liquidity for a sample of NYSE stocks (see also Huberman and Halka (2001)). Common effects in longer-horizon trading volume are identified by Tkac (1999) via market model type regressions, and Lo and Wang (2000) via factor analysis.

*Harford, jarrad@u.washington.edu, School of Business Administration, Box 353200, University of Washington, Seattle, WA 98195, and Kaul, akaul@ualberta.ca, School of Business, University of Alberta, Edmonton, AB T6G 2R6, Canada. Part of this work was completed while Harford was at the University of Oregon and Kaul was visiting at the University of Oregon. The authors are grateful to the Q-Group for financial support, and Kaul also acknowledges financial support from a Xerox fellowship at the University of Alberta. We thank John Chalmers, Diane Del Guercio, Jonathan Karpoff (the editor), Jennifer Koski, Wayne Mikkelson, Wayne Wagner, seminar participants at the 2001 Q-Group Fall Meetings, the Universita Della Svizzera Italiana, and the University of Oregon, and especially Sunil Wahal (the referee) for helpful comments. We also thank Lipper and the Investment Company Institute and Brian Reid for providing S&P 500 index fund figures, Jeff Wurgler for providing part of the S&P 500 addition data, and Tim Poole and Rahul Ravi for research assistance. We are responsible for any errors.

[1]Order flow is the difference between buy and sell volume or number of transactions.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

Our study extends this literature by exploring the following unanswered questions about common effects or *commonality* (following Chordia, Roll, and Subrahmanyam (2000)) in order flow, returns, and trading costs. First, there are the related questions of the pervasiveness and sources of these common effects. Because the extant evidence on commonality is based on samples dominated by stocks in major indexes, we hypothesize that trading by index funds or arbitrageurs may be the major source of commonality. This implies that samples of non-index stocks will exhibit little in the way of common effects. Second, is commonality in returns and trading costs produced by commonality in order flow? Unless commonality in order flow affects returns or costs, it is of little importance. O'Hara ((2003), p. 1339) points out that the importance and sources of commonality in trading costs are unresolved issues. We revisit the question of commonality in trading costs and explore the extent to which commonality is driven by correlated order flow.

Using intraday data, we study common effects for stocks in the S&P 500 index and a large sample of non-S&P stocks for 1986 and 1996. In each year, we examine both the end of the day (when institutions are most active) and the rest of the day.

First, we establish a baseline level of commonality in the two samples in 1986 and 1996 by calculating the average correlation in intraday (15-minute) order flow, returns, and trading costs for all possible within-sample pairs of stocks in each of the S&P and non-S&P samples. To identify the factors driving order flow commonality, we hypothesize a hierarchy of non-mutually exclusive sources of common effects, comprising industry, index, and broad market order flow and returns. We sequentially add these sources of common effects to regressions that explain an individual stock's order flow, and examine the increase in R^2 and the decrease in the pairwise correlations of the residual order flows at each step. The advantage of our approach is that we identify candidates for common factors and examine their explanatory power.

Consistent with the hypothesis that indexing is driving common effects, common effects in order flow are larger in 1996 than in 1986 and are generally strongest at the end of the day. The strong common effects in order flow for S&P constituents drop to almost zero once indexing effects are removed, and the incremental effects of industry or broad market factors, while statistically significant, are economically negligible. When we examine order flow in non-index stocks, we find statistically discernible common effects, driven by industry and broad market factors. However, the effects are only one-quarter to one-half as large as the corresponding effects for index constituents.

Next, we estimate models that link an individual stock's returns or trading costs to order flow coming from the hypothesized sources of commonality. Our specification includes terms to capture the interaction between stock and common order flow, which allows us to measure how common order flow changes the impact of a stock's own order flow on its price and trading costs. We find that, as with order flow, common effects in returns are appreciably stronger for index stocks and increase between our sample years and at the end of the day. In both samples, controlling for own order flow leads to a large decline (of approximately 50%) in the return correlation. Controlling for aggregate index order flow reduces

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

index constituent return correlation by an additional 50% to 70%, and returns are slightly less sensitive to industry order flow. For non-index stocks, returns are sensitive to broad market and industry order flow, and the return correlation declines after either type of order flow is incorporated.

Additionally, we use the method of principal components to statistically extract common factors from the stock-level series, and examine the correlation between the principal components and our hypothesized sources of commonality. The first principal component for order flow and returns in each sample is dominant. In the case of order flow, the first principal component for S&P stocks is highly correlated with aggregate S&P order flow, while that for non-S&P stocks is highly correlated with aggregate non-S&P order flow. For returns, the first principal component in the two samples is strongly correlated with aggregate S&P and non-S&P order flow, respectively. Thus, we find correspondence between our economic variables and the statistically extracted factors.

Our last set of tests comprises an event study of index addition. We compare the comovement of stock order flow, returns, and trading costs with index, broad market, and industry variables pre- and post-addition. Post-addition, the comovement of a stock's order flow (returns) with index order flow (returns) increases six- to 10-fold, comovement with broad market order flow (returns) declines to zero or turns negative, and the sensitivity of returns to own and index order flow changes dramatically. This evidence emphasizes the importance of indexing in driving comovement in order flow and returns.

In both samples, common effects in trading costs, measured by the effective spread, are considerably weaker than those in order flow and returns. Moreover, order flow at the stock and aggregate levels does little to explain commonality in trading costs. The weak comovement in trading costs is consistent with other forces driving trading costs, as suggested by Domowitz and Wang (2002).

Establishing indexing as the primary source of common effects offers several insights. First, it shows that the common order flow effects identified by Hasbrouck and Seppi (2001) in a sample of exclusively index stocks are not pervasive. Second, our results identify tangible effects of indexing on the trading environment of stocks in the popular indexes. Third, the common effects in order flow induced by indexing appear to drive commonality in returns. We discuss the implications of these findings in the concluding section.

II. Hypothesis Development

A. The Sources of Order Flow Commonality

We explore three sources of common order flow in stocks. The first source of commonality is broad shifts in investor allocations, taking the form of increased flows to equity mutual funds, or asset allocation shifts from bonds to stocks, for example. Edelen (1999) and Edelen and Warner (2001) suggest that there is a common component to flows across mutual funds. Thus, the *broad market effects* hypothesis predicts the existence of commonality in order flow for all stocks, regardless of industry or index membership.

The second source of common effects is correlated industry events. Stocks in the same industry or in related industries will exhibit correlated order flow due to common information, or industry or sector mutual funds. Thus, the *industry effects* hypothesis predicts a strong common order flow factor for firms in the same industry. Note that an industry factor can explain common effects within an industry but not pervasive common effects across industries.

The third source of commonality is indexing. The dramatic growth in indexing, both by mutual funds and by pension funds, presents the possibility that the constituents of popular indexes are subject to strong common order flow effects. Since the S&P 500 is the index most widely followed by index funds and also by index arbitrageurs, firms in the S&P 500 index face a clearly delineated source of commonality. The *indexing* hypothesis predicts that order flow for S&P stocks will exhibit substantial common effects, and will be associated mainly with aggregate S&P order flow and returns. Since index funds are likely to be especially active at the end of the trading day, we also study time-of-day patterns in common effects.[2]

B. The Effects of Order Flow Commonality

We explore hypotheses concerning the effects of common order flow on prices and trading costs. Aggregate order flow need not affect either, as correlated returns or trading costs could be explained by (correlated) stock-specific order flow. This is the case, for instance, if each specialist sets prices using only order flow in his stock (e.g., Glosten and Milgrom (1985)). However, an incremental role for common order flow could plausibly arise if specialists learn from order flow in related stocks, as suggested in multi-asset microstructure models.[3] For instance, it is reasonable to assume that index fund trading is not motivated by private information (Subrahmanyam (1991)). Then, if specialists, by observing aggregate S&P order flow, can tell when index funds are trading, they can use that information to revise the probability that a given trade in an S&P constituent is uninformed. The prediction is that a trade in an S&P stock made in the same direction as aggregate S&P order flow should have a smaller price (and cost) effect than one in the opposite direction of aggregate order flow. Similarly, if market-

[2]Institutions, especially index funds, have a clear preference for trading near the close of the market. First, they are able to minimize tracking error in their returns relative to their benchmark (e.g., the S&P 500), whose value is computed using closing prices. Also, by waiting for a better estimate of net flow and thereby being able to offset individual inflows and outflows occurring during the day, funds can reduce the required amount of trading and their trading costs. Anecdotal evidence supports these arguments. In describing the daily activities of the Vanguard 500 Portfolio, *The Wall Street Journal* (January 28, 1997) reports that the fund "has been buying some stocks all day" but that "the bulk of its buying will occur in the next 45 minutes, right before the close of the New York Stock Exchange at 4:00 p.m." In the same newspaper (March 3, 1997), the fund's manager describes why one-third of his trades occur in the final hour: "[S]ince the investor is buying the fund at the closing price of the day, we want to pay a price as close as possible to that on close."

[3]Caballe and Krishnan (1994) assume that every trader and market maker can observe order flow in all assets, while Kumar and Seppi (1994) assume that this observability occurs with a lag. Easley, O'Hara, and Srinivas (1998) assume that stock and options market makers are able to view trading activity and prices in the related market. Admati (1985) considers price setting in a multi-asset rational expectations model.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

wide (or industry) trading is informative to specialists, broad market (industry) order flow will affect prices as well.

III. Data and Empirical Methods

Our tests analyze stocks in the S&P 500 and a random sample of non-S&P firms for two years, 1986 and 1996.[4] We choose two widely spaced years in order to capture possible effects associated with the growth in trading, especially institutional trading.

The list of firms in the S&P 500 is obtained from biannual issues of the S&P *Stock Price Record*. To obtain a sample of non-S&P stocks, we exclude ADRs, closed-end funds and REITs, and sort the remaining NYSE-listed common stocks into quintiles on the basis of annual trading volume for each year. We then select the first 75 firms per quintile in alphabetical order. This procedure is designed to produce a random sample of firms with broad representation in terms of trading activity. Before selecting stocks but after defining the volume quintiles, we eliminate stocks that in either year i) have an average price below $10; and ii) have fewer than 100 trades in total. The purpose of the first condition is to exclude stocks for which microstructure problems are likely to be severe, e.g., volatile returns. The second condition removes stocks that are very infrequently traded (have less than one trade every two days).

Transaction and quote data come from the Institute for the Study of Securities Markets (ISSM) tapes for 1986 and from the Trade and Quote (TAQ) CDs for 1996. Following Hasbrouck and Seppi (2001), we measure variables over 15-minute intervals. The presumption in our tests and those in Hasbrouck and Seppi, and indeed underlying theoretical models of market microstructure, is that causation runs from trading activity to prices. Accordingly, we would like to be able to attribute changes in prices to patterns in order flow. However, as the measurement interval is lengthened there is a heightened risk of causality being confounded by feedback from prices to trading activity. The use of 15-minute intervals represents a balance between minimizing such endogeneities and maximizing the number of intervals with non-missing observations. To examine time-of-day effects, we separately analyze the four 15-minute intervals corresponding to the last hour and the 22 intervals corresponding to the rest of the day. If correlated trading effects are driven by institutional transactions and institutions prefer to trade near the end of the day, these effects should be most apparent at the end of the day.

To calculate order flow, we sign each trade by comparing the transaction price with the midpoint of the prevailing quote, as long as the quote has been in effect for at least five seconds when the trade occurs (see Lee and Ready (1991)). Buys, given a positive sign, are transactions occurring at prices above the midpoint, while sells, given a negative sign, are transactions occurring at prices below the midpoint. We exclude trades occurring at the quote midpoint, and this causes us to lose a small proportion of transactions (e.g., 13.2% in 1986). Order flow can be defined in terms of the number of transactions (the difference between the

[4]The first full year for which the current trading hours on the NYSE (9:30AM to 4:00PM) were in effect is 1986.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

number of buys and the number of sells) or in terms of trading volume (which also incorporates the number of shares bought and sold). Prior research, e.g., Hasbrouck (1991), finds that order flow based on the number of transactions is better behaved. Hence, the results that we report use this measure of order flow, cumulated for each 15-minute interval in the day.

Intraday returns are calculated using the average of the bid and ask prices, so as to avoid bid-ask bounce. Returns are based on the most recently available midpoint, starting with the first quote of each day, and will automatically be zero if a quote has not been posted in any interval. We address the possible effects of stale quotes on our results in the next section.

When analyzing trading costs, we focus on the effective spread. This is defined as twice the absolute difference between the transaction price and the midpoint of the spread in effect at the time of the transaction, scaled by the quote midpoint. The effective spread is preferred to the relative spread (raw spread scaled by the quote midpoint) because it accounts for within-spread transactions. For each stock and each 15-minute interval, we calculate the average effective spread for buys and for sells. Analysis of the relative spread yields similar conclusions, and we do not report the results.

The factors that we use as explanatory variables consist of aggregate S&P 500, non-S&P, and industry order flow and returns. We construct these factors as follows. For each 15-minute interval, we calculate order flow and returns for all NYSE-listed stocks meeting our price and trade cutoffs in 1986 and 1996, and use these to calculate cumulative order flow and the equally weighted return for the S&P, non-S&P, and industry portfolios. Cumulative order flow for a 15-minute interval is the sum of the order flows for the constituent stocks. To prevent stale prices from affecting the portfolio return for an interval, this is computed using only stocks with actual quotes in the interval.

After excluding firms that i) fail to surpass the price and trade exclusionary hurdles, ii) are not ordinary domestic shares, and iii) are not NYSE firms, we are left with 1,196 firms in 1986 and 1,549 firms in 1996. Of these, 383 firms belong to the S&P 500 sample and 813 to the non-S&P sample in 1986; the corresponding sample sizes for 1996 are 454 and 1,095. We group stocks into 15 industries. The classification scheme follows Lewellen (2002) and is similar to Moskowitz and Grinblatt (1999), who employ 20 industries, and to the 17-industry classification at Kenneth French's Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/). There are two industries in 1986 and one in 1996 in which the number of firms is less than 10. For the remaining industries, order flow and returns are calculated using between 37 and 179 firms in 1986 and between 33 and 269 firms in 1996.

IV. Results

A. Trading Activity: Descriptive Statistics

Table 1 provides descriptive statistics for the S&P 500 and non-S&P samples for 1986 and 1996. The figures in this table are 15-minute averages, and are calculated separately for the first 5½ hours and the last hour of the trading

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

day. Starting with trading activity for 9:30AM to 3:00PM, total trading volume for S&P stocks has increased from 26,000 round lots to almost 70,000 round lots per 15-minute interval over this 10-year period. The increase in non-S&P trading volume, while smaller, is also substantial. Order flow (measured using net volume or net number of trades) is, for the most part, positive for both groups, and has grown substantially between 1986 and 1996. Per interval trading volume and order flow are larger during the final hour than during the rest of the day, in each year and for both samples. The mean effective spreads for S&P stocks are half as large as those for non-S&P stocks and both have decreased by more than 40% between 1986 and 1996.

TABLE 1

Summary Statistics for S&P and Non-S&P Stocks

	1986		1996		% Change between 1986 & 1996	
Summary Statistics	9:30AM–3:00PM	3:00–4:00PM	9:30AM–3:00PM	3:00–4:00PM	9:30AM–3:00PM	3:00–4:00PM
S&P Net Volume	938	1,209	2,408	4,894		
Non-S&P Net Volume	0	169	551	1,661		
S&P Net Trades	49	82	99	164		
Non-S&P Net Trades	12	30	56	73		
S&P Total Volume	26,225	28,623	69,575	87,434	165%	205%
Non-S&P Total Volume	15,208	16,603	35,366	44,306	133%	167%
S&P Total Trades	1,670	1,784	2,726	3,852	63%	116%
Non-S&P Total Trades	1,130	1,151	1,566	1,985	39%	72%
S&P Effective Spread						
Buys	0.0068	0.0069	0.0034	0.0034	−50%	−51%
Sells	0.0067	0.0067	0.0035	0.0035	−48%	−48%
Non-S&P Effective Spread						
Buys	0.0106	0.0106	0.0065	0.0064	−39%	−40%
Sells	0.0106	0.0106	0.0066	0.0065	−38%	−39%

Averages for net and total volume and number of trades and effective spreads are presented for the two sample years, 1986 and 1996. Net trades (net volume) are calculated by classifying transactions as buys or sells (using the Lee-Ready (1991) algorithm) and calculating the number of buys less sells (buy volume less sell volume) for each 15-minute interval. We use net trades as our order flow measure for the remaining tests in the study. Statistics are separately computed by S&P membership and by time of day. Volume numbers are round lots. The numbers presented are 15-minute averages; for example, there was an average total of 26,225 round lots traded in all S&P stocks every 15 minutes during the first part of the day in 1986. Note that the figures for the S&P 500 exclude data for NASDAQ stocks, and that the statistics for the non-S&P sample are calculated using all non-S&P stocks, not just the 375 stocks used in our tests.

The intraday patterns in trading activity are consistent with the tendency of institutions to place trades near the end of the day. While the jump in volume at the end of the day is consistent with the trading preference of index funds documented in footnote 2, index funds are not solely responsible for the higher volume. Unofficial indexing (by pension funds or individuals) along with the visibility of closing prices, and even informed traders drawn by the concentrated trading (Admati and Pfleiderer (1988)) undoubtedly account for some of the increase. [5]

These statistics show that trading activity has increased overall, especially at the end of the day, and with slightly larger increases occurring for stocks in the

[5]We should comment on the fact that order flow is positive in both years. As a percentage of total volume or number of trades, this order imbalance is small, less than 5% in each case. We also compare our numbers with descriptive statistics in Table 2 of Hasbrouck and Seppi. Their mean net volume is 16 round lots per firm per 15-minute period in 1994. Averaging net volume for the longer interval, 9:30AM to 3:00PM, across the number of S&P 500 firms in 1996 (454), we get approximately five round lots per firm. Hasbrouck and Seppi's figures probably are larger because they look at the 30 Dow Jones stocks.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

S&P 500. They suggest that we are most likely to find evidence of order flow commonality and of its effects at the end of the day.

B. The Magnitude of Common Effects: Correlations

Before addressing the sources of common effects, it is important to examine their strength and pervasiveness. To document these common effects, we calculate the correlation in 15-minute order flow (net number of trades), returns, and trading costs for every pair of stocks in the S&P and non-S&P samples and for 1986 and 1996. The correlations are calculated using the time series of observations for each 15-minute interval (e.g., 10:30–10:45AM) and then averaged across the two broad intervals corresponding to the first 5½ hours and the last hour of the trading day.[6] Order flow is cumulated within two trade size categories, 100–9,900 shares and 10,000 shares or more (large trade order flow).

Several general statements can be made about the order flow and return correlations in the top two panels of Table 2. The pairwise correlations for both are appreciably larger for S&P stocks than for non-S&P stocks, and are mainly in excess of 10% for S&P stocks. Given the high frequency nature of the data, these correlations are substantial and would be expected to increase as the interval was lengthened to a day or longer. For S&P stocks, the correlations increase at the end of the day and from 1986 to 1996. Finally, the correlation for large order flow is insignificant in both years for S&P stocks (and also for non-S&P stocks). This is not surprising since large (block) trades tend to be infrequent and often executed in time consuming ways. Therefore, in subsequent tests, we drop block trade order flow and focus on order flow cumulated over transactions of between 100 and 9,900 shares.[7]

These results provide initial support for the indexing hypothesis. The facts that the order flow correlations are large for S&P stocks, are higher at the end of the day, and are higher in 1996 are consistent with strong common order flow being induced by index funds that trade intensively at the end of the day. The positive correlations for S&P stock order flow during the rest of the day are consistent with the actions of index arbitrageurs (which need not be confined to the end of the day) or with funds spreading out their trades to avoid placing all of their trades at the end of the day. Given the growth in indexing through time, funds are more likely to trade throughout the day in 1996, which would explain the larger increase in the order flow correlation during the 9:30AM–3:00PM interval. However, the correlation remains below that at the very end of the day. The return correlations, which mirror the pattern of the order flow correlations, are consistent with order flow having a material price impact, and the elevated correlation at the end of the

[6] A detailed analysis of the correlations by 15-minute interval reveals relative homogeneity within the last hour and within the rest of the day, suggesting that the coarser partition is reasonable. In particular, the correlations during the opening hour are lower than those during the last hour, and are not different from those corresponding to the rest of the day.

[7] For each stock, we calculate the average fraction of 15-minute volume and number of trades accounted for by block trades. The cross-sectional averages (computed across samples and years) are 7% for volume and 3% for number of transactions. To check the sensitivity of our conclusions to the omission of block trades, we include block trades in order flow and repeat the main tests in the paper. Our conclusions are identical.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 2

Pairwise Order Flow, Return, and Trading Cost Correlations

	S&P		Non-S&P	
	9:30AM–3:00PM	3:00–4:00PM	9:30AM–3:00PM	3:00–4:00PM
Panel A. Median Pairwise Order Flow Correlation between any Two Stocks				
1986				
100-9,900	0.061	0.119	0.008	0.013
10,000+	0.000	0.000	0.000	0.000
1996				
100-9,900	0.105	0.159	0.011	0.013
10,000+	0.002	0.004	0.000	0.000
Panel B. Median Pairwise Return Correlation between Any Two Stocks				
1986	0.041	0.070	0.006	0.008
1996	0.117	0.142	0.017	0.017
Panel C. Median Pairwise Spread Correlation between Any Two Stocks				
1986				
Eff. Spread (buys)	0.005	0.009	0.007	0.011
Eff. Spread (sells)	0.006	0.009	0.015	0.012
1996				
Eff. Spread (buys)	0.021	0.024	0.008	0.009
Eff. Spread (sells)	0.017	0.017	0.009	0.007

The median pairwise correlations for order flow (net number of transactions), returns, and trading costs are tabulated. The numbers are presented for S&P and non-S&P stocks for different times in the day, for different trade sizes (in the case of order flow), and for the two sample years, 1986 and 1996. The trading cost measure is the mean effective spread (scaled by the quote midpoint) in a 15-minute interval, given separately for buys and sells. The correlations are computed at the 15-minute interval level and grand averages are presented for the first 5½ hours of the day and the last hour of the day. For example, the average correlation between order flow in trades of 100 to 9,900 shares for any two S&P stocks in a 15-minute period was 0.061 in the first part of the day during 1986.

day suggests that institutional transactions have particularly pronounced effects on prices at this time.

Table 2 also shows that the order flow and return correlations in the non-S&P sample are less than 2%. The small correlations imply that common effects are relatively unimportant for non-index stocks. Since previous evidence in Hasbrouck and Seppi (2001) is based on the widely tracked Dow stocks, this finding is significant because it indicates that commonality is basically an index-driven phenomenon. Nonetheless, we will evaluate the ability of broad market and industry factors to explain common effects in the non-S&P sample.[8]

As Panel C of Table 2 shows, the pairwise correlations for the effective spread are appreciably lower than those for order flow and returns. Not only are the correlations smaller, but the differences by time of day, between buys and sells, and across samples are also negligible. The weak comovement in trading costs documented in Table 2 is consistent with Hasbrouck and Seppi (2001), who analyze comovement in several measures of liquidity. Domowitz and Wang (2002) argue that common effects in returns are driven by correlated order flow while comovement in measures of liquidity is a function of the correlation in order type (i.e., market versus limit orders). Our result that the levels of the order flow

[8]We check that the results in Table 2 are not mechanically produced by differences in trading activity across the two samples or across the two years. For instance, if non-S&P stocks trade infrequently, many 15-minute intervals will have zero order flow and zero returns. This could lead to lower correlations for non-S&P stocks than for S&P stocks. As a test of this possibility, we ignore intervals with no trades or quotes and recompute the correlations using only the remaining "active" intervals. The results (not tabulated) continue to support all of the inferences from Table 2.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

and return correlations are similar and larger than the trading cost correlations is consistent with this argument.

C. Explaining Order Flow Commonality

Our next task is to explain the order flow correlations documented in Table 2. We proceed by specifying the following regression between order flow for each stock and broad market, index and industry order flow and returns, and own stock control variables,

$$\begin{aligned}(1)\quad OF_{it} \;=\; & a_0 + a_1 OF_{i,t-1} + a_2 RET_{i,t-1} + b_1 OF_{\text{Ind},t} + b_2 OF_{\text{Ind},t-1} \\ & + b_3 RET_{\text{Ind},t-1} + c_1 OF_{\text{SP500},t} + c_2 OF_{\text{SP500},t-1} + c_3 RET_{\text{SP500},t-1} \\ & + d_1 OF_{\text{Non-S\&P,t}} + d_2 OF_{\text{Non-S\&P},t-1} + d_3 RET_{\text{Non-S\&P},t-1} + e_{it}.\end{aligned}$$

Here, i represents an individual stock, OF stands for 15-minute order flow (net number of trades) consisting of trades between 100 and 9,900 shares, and t represents a particular 15-minute period; $OF_{\text{SP500},t}$, $OF_{\text{Non-S\&P},t}$, and $OF_{\text{Ind},t}$ are order flow for all S&P 500 stocks, non-S&P stocks, and the stock's industry in period t; $RET_{\text{SP500},t-1}$, $RET_{\text{Non-S\&P},t-1}$ and $RET_{\text{Ind},t-1}$ are S&P, non-S&P, and industry returns in the previous 15-minute period; and $OF_{i,t-1}$ and $RET_{i,t-1}$ are the previous period's order flow and return for stock i. When using stock i's order flow as the dependent variable, we remove stock i's order flow from all of the appropriate aggregate order flow regressors. For notational convenience, we exclude the firm subscript from the regression coefficients throughout the paper.

To allow comparisons across stocks and years, we standardize each stock-level and aggregate-level variable. For instance, we standardize stock i's order flow for 11:30–11:45AM on each day by the standard deviation of 11:30–11:45AM order flow for stock i, calculated over the entire year (and similarly for every regressor). The regression is estimated separately for each stock and in each year using the time series of standardized variables corresponding to the last hour of the day (i.e., four 15-minute periods for each day) and the remaining 5½ hours (i.e., 22 15-minute periods per day). We then average the regression coefficients separately for stocks in the S&P 500 index and those not in the index.

We report the final specification (1), but estimate the model sequentially, starting with the own stock variables and then adding the industry, indexing, and broad market return and order flow factors. The purpose is to assess the incremental contribution of the three sets of common factors by examining changes in both the R^2 and the residual order flow correlation. A residual correlation close to zero means that much of the commonality has been eliminated and is an indication of model adequacy. Since the incremental contribution of a regressor depends on the order in which it and the other regressors are added to the model, we estimate the model with different permutations of the common variables. For instance, in one permutation, we add the industry variables first, followed by the S&P variables, and then the broad market variables. In a second permutation, we add the index variables first, followed by the industry variables, and the broad market variables. The full specification is presented in Table 3 and the performance of the sequence of models is summarized in Table 4.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 3

Order Flow Regressions

		Own		Industry			S&P			Non-S&P			
	Intercept	OF_{-1}	Ret_{-1}	OF	OF_{-1}	Ret_{-1}	OF	OF_{-1}	Ret_{-1}	OF	OF_{-1}	Ret_{-1}	Adj. R^2
Panel A. S&P													
1986													
9:30AM–3:00PM													
Median	−0.029*	0.223*	−0.125*	0.018*	0.003	0.004*	0.253*	−0.047*	0.017*	0.014*	−0.015*	−0.005*	0.138*
p-val	0.000*	0.000*	0.000*	0.000*	0.155	0.001*	0.000*	0.000*	0.000*	0.000*	0.000*	0.024*	0.000*
3:00–4:00PM													
Median	−0.033*	0.216*	−0.129*	−0.003	0.006*	0.006	0.376*	−0.057*	0.017*	0.012*	−0.011*	−0.006	0.201*
p-val	0.000*	0.000*	0.000*	0.182	0.048*	0.246	0.000*	0.000*	0.000*	0.000*	0.011*	0.261	0.000*
1996													
9:30AM–3:00PM													
Median	−0.020*	0.167*	−0.094*	0.028*	0.002*	0.014*	0.308*	−0.056*	0.027*	0.000	−0.011*	−0.003	0.146*
p-val	0.000*	0.000*	0.000*	0.000*	0.060*	0.000*	0.000*	0.000*	0.000*	0.471	0.000*	0.113	0.000*
3:00–4:00PM													
Median	−0.019*	0.173*	−0.093*	0.035*	−0.005	0.011*	0.384*	−0.064*	0.025*	−0.010*	−0.007*	−0.003	0.207*
p-val	0.002*	0.000*	0.000*	0.000*	0.204	0.000*	0.000*	0.000*	0.000*	0.002*	0.014*	0.468	0.000*
Panel B. Non-S&P													
1986													
9:30AM–3:00PM													
Median	−0.011*	0.148*	−0.070*	0.010*	0.001	0.003*	0.000	−0.003	0.002	0.068*	−0.002*	0.004*	0.039*
p-val	0.009*	0.000*	0.000*	0.000*	0.417	0.000*	0.709	0.776	0.516	0.000*	0.047*	0.026*	0.000*
3:00–4:00PM													
Median	−0.013*	0.152*	−0.084*	−0.002	0.010*	−0.005	0.002	0.004	−0.006	0.100*	−0.020*	0.009*	0.055*
p-val	0.000*	0.000*	0.000*	0.971	0.016*	0.261	0.477	0.446	0.117	0.000*	0.002*	0.013*	0.000*
1996													
9:30AM–3:00PM													
Median	−0.009	0.147*	−0.080*	0.017*	0.003	0.009*	−0.010*	−0.009*	0.006*	0.083*	−0.009*	0.004*	0.034*
p-val	0.289	0.000*	0.000*	0.000*	0.252	0.000*	0.004*	0.000*	0.005*	0.000*	0.000*	0.018*	0.000*
3:00–4:00PM													
Median	0.000	0.151*	−0.083*	0.029*	0.003	0.009*	−0.030*	−0.006	0.002	0.099*	−0.015*	0.004	0.040*
p-val	0.827	0.000*	0.000*	0.000*	0.745	0.009*	0.000*	0.221	0.569	0.000*	0.000*	0.221	0.000*

Table 3 presents estimates of equation (1) in the text. The specifications are estimated separately for S&P and non-S&P stocks for two broad intraday periods and for the two sample years, 1986 and 1996. The specifications are estimated at the stock level for each year, separately using the time series of 15-minute observations covering the hours 9:30AM–3:00PM (22 observations per day) and the hour (3:00–4:00PM) at the end of the day (4 observations per day). For example, the dependent variable in the first row is order flow for an individual S&P stock for a given 15-minute interval between 9:30AM and 3:00PM in 1986. The cross-sectional medians of the regression coefficients and the adjusted R^2 are presented, along with their *p*-values, for each sample, year, and time of day. There are two main types of regressors: order flow and its first lag (OF and OF_{-1}) and lagged returns (Ret_{-1}). S&P, Non-S&P, Industry, and Own refer to aggregate S&P, aggregate non-S&P, aggregate industry, and the stock's own data. All variables are standardized by their time-series standard deviations. *indicates significance at the 10% level or better.

Panel A of Table 3 presents the median regression coefficients for the S&P 500 sample for the last hour of the day and the rest of the day, and Panel B presents the matching coefficients for non-S&P stocks. In discussing Table 3 (and Table 5), we will often make comparisons between coefficients. We have conducted formal tests of differences, but do not wish to complicate the presentation by tabulating the results of so many tests. Any difference we highlight is statistically significant and, in our opinion, economically significant.

1. S&P Stocks

The indexing hypothesis is supported by the results. The coefficient on contemporaneous aggregate S&P 500 order flow is always significantly positive and several times as large as the coefficient on the first lag of S&P order flow (which is negative). This means that there are more trades in an index constituent stock

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 4

Marginal Impact of Hypothesized Factors on Common Effects in Order Flow

	Corr.	Adj. R^2	Corr.	Adj. R^2	Corr.	Adj. R^2	Corr.	Adj. R^2
Panel A. S&P								
	1986				1996			
	Own Stock Controls		Own Stock Controls		Own Stock Controls		Own Stock Controls	
9:30AM–3:00PM	0.077	0.056	0.077	0.056	0.117	0.027	0.117	0.027
3:00–4:00PM	0.142	0.062	0.142	0.062	0.172	0.031	0.172	0.031
	Add Industry		Add S&P		Add Industry		Add S&P	
9:30AM–3:00PM	0.021	0.106	0.001	0.137	0.018	0.109	0.001	0.144
3:00–4:00PM	0.026	0.157	0.001	0.201	0.019	0.167	0.001	0.205
	Add S&P		Add Industry		Add S&P		Add Industry	
9:30AM–3:00PM	0.001	0.137	0.001	0.137	0.001	0.146	0.001	0.146
3:00–4:00PM	0.001	0.201	0.001	0.201	0.001	0.208	0.001	0.208
	Add Rest of Mkt.		Add Rest of Mkt.		Add Rest of Mkt.		Add Rest of Mkt.	
9:30AM–3:00PM	0.001	0.138	0.001	0.138	0.001	0.146	0.001	0.146
3:00–4:00PM	0.001	0.201	0.001	0.201	0.001	0.207	0.001	0.207
Panel B. Non-S&P								
	1986				1996			
	Own Stock Controls		Own Stock Controls		Own Stock Controls		Own Stock Controls	
9:30AM–3:00PM	0.008	0.021	0.008	0.021	0.010	0.019	0.010	0.019
3:00–4:00PM	0.013	0.025	0.013	0.025	0.013	0.020	0.013	0.020
	Add Industry		Add Rest of Mkt.		Add Industry		Add Rest of Mkt.	
9:30AM–3:00PM	0.002	0.033	0.000	0.039	0.002	0.031	−0.001	0.034
3:00–4:00PM	0.003	0.041	0.000	0.052	0.003	0.034	−0.001	0.038
	Add S&P		Add Industry		Add S&P		Add Industry	
9:30AM–3:00PM	0.002	0.033	0.000	0.039	0.002	0.032	−0.001	0.034
3:00–4:00PM	0.004	0.042	0.000	0.054	0.003	0.035	−0.001	0.040
	Add Rest of Mkt.		Add S&P		Add Rest of Mkt.		Add S&P	
9:30AM–3:00PM	0.000	0.039	0.000	0.039	−0.001	0.034	−0.001	0.034
3:00–4:00PM	0.000	0.055	0.000	0.055	−0.001	0.040	−0.001	0.040

Table 4 presents the average adjusted R^2 and the average correlation of the residuals from stepwise estimation of equation (1). Stock effects are added to the specification first, followed by the common factors in the indicated order. For example, the first set of numbers shows that accounting for only a stock's own effects (Own OF_{-1} and Own Ret_{-1}), the residual order flows of the S&P stocks in 1986 have an average pairwise correlation of 7.7% during the main part of the trading day. This number can be compared to the average pairwise correlation of raw order flows for the same stocks presented in Table 2 (6.1%). Proceeding down the column, once the industry factors (Ind OF_{-1}, Ind OF, Ind Ret_{-1}) are added to equation (1), the average correlation of the residual order flows drops to 2.1% and the fraction of an individual stock's order flow variation explained by the regressors increases from 5.6% to 10.6%. The second column shows the effect of adding S&P factors (S&P OF, S&P OF_{-1}, S&P Ret_{-1}) before the industry factors.

when other S&P 500 stocks are being traded, and these trades are of the same sign (buys or sells). Given the growth in indexing, the fact that the end-of-day coefficient is constant from 1986 to 1996 may appear to be at odds with indexing effects. However, the coefficient does not have to increase. Increased basket trading by index funds coupled with a constant per trade effect will still produce the stronger correlation in order flow for index constituents observed in Table 2. This is seen in the economic impact of trading in the index. Untabulated calculations show that a one standard deviation increase in 15-minute S&P order flow at the end of the day in 1986 leads to one more trade of the same sign in a constituent stock on average, compared to an average total of 2.2 trades per 15-minute interval. In 1996, the end-of-day effect is 1.5 additional trades relative to an average of three trades per interval. During the rest of the day in 1996, the effect is only

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

one additional trade on average, compared to an average total of 2.5 trades. The increase in the strength of this comovement at the end of the day, when index funds are most active, supports indexing as the underlying cause.

The performance of the index provides another source of commonality, with a lag. Two factors that motivate the inclusion of lagged index returns are index arbitrage strategies, determined in part by cash index returns, and feedback effects from returns to trades. While index arbitrage effects should lead to a negative coefficient on index returns, positive feedback trading effects should lead to a positive coefficient, ceteris paribus.[9] The positive coefficient on S&P returns from the previous 15-minute period is consistent with feedback effects. These effects are similar during the last hour and the rest of the day, which is reasonable since feedback trades need not exhibit time-of-day sensitivity.

The next hypothesized source of common effects is broad market shifts in allocations, preferences, or information. Such commonality would present itself in the form of significant coefficients on aggregate non-S&P order flow and returns. The small coefficients on non-S&P order flow and returns show that broad market factors are not important drivers of correlated order flow for index stocks. Lastly, the specification provides evidence regarding the industry effects hypothesis. By its nature, this source of commonality is more localized, since it will only explain common effects within an industry. Nonetheless, there is evidence of commonality induced by industry order flow, although its strength is no more than one-tenth that of the indexing effect. The increase in the coefficients between 1986 and 1996 shows that industry effects have become a more important source of order flow commonality in recent years.

Table 4 summarizes the changes in the residual order flow correlation when we sequentially add common variables. Note first that the residual correlations after firm-specific order flow and returns are added actually increase slightly in comparison to the raw correlations in Table 2. This is not surprising since controlling for firm-specific variables, which dilute common effects, should heighten common effects. However, adding the industry variables leads to a substantial reduction in the correlation, e.g., from in excess of 12% to under 2% in 1996. Following the addition of S&P order flow and returns, the correlation declines further to 0.1%, and is no longer significant. No additional reduction occurs when the non-S&P variables are included. Table 4 also shows that when we add the S&P variables first, the residual correlation immediately drops to 0.1%, and the industry and non-S&P variables offer no further reduction. Thus, the major contributors to the reduction in order flow commonality for index stocks are S&P and, secondarily, industry, order flow, and returns.

2. Non-S&P Stocks

The regression with order flow in individual non-S&P stocks as the dependent variable provides evidence on the broad market effects hypothesis of market-wide, non-index-related commonality. While the order flow correlations for non-S&P stocks in Table 2 are low, it is still important to examine the influences for

[9] For instance, positive cash index returns result in the cash index being overvalued relative to the futures index, holding futures returns constant. This should prompt selling in the stock market.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

these stocks and to see if our hypothesized factors are able to explain common effects in this sample as well.

The coefficient on aggregate non-S&P order flow is reliably positive. Yet, the coefficient is 30% as large as the corresponding S&P order flow coefficient in the regressions for S&P stocks, and its economic significance is also appreciably smaller. For example, a one standard deviation increase in non-S&P order flow increases trading in a non-S&P stock by an average of only 0.09 trades, compared to an average total of 0.6 trades per 15-minute interval in 1986. While broad market effects are weaker than indexing effects, however, non-S&P stocks clearly have their own common influences, perhaps due to aggregate flows to stock funds in general or to broad allocation shifts on the part of individuals and institutions. The effect of lagged broad market returns, while positive and generally significant, is also weaker than that of index returns on index constituents.

Industry effects are present for non-index stocks as well. The effect of industry order flow on stock order flow is generally positive and significant, and that of industry returns is mixed in 1986, but positive in 1996. The industry order flow coefficients are comparable to those in the S&P regressions. Note that order flow for non-S&P stocks is negatively affected by aggregate S&P order flow and largely unaffected by lagged S&P returns. However, the coefficient on index order flow is small relative to that on broad market order flow.[10]

As with the S&P sample, we estimate the model sequentially in order to gauge the role of the common factors in explaining order flow commonality. The results are presented in Panel B of Table 4. Comparing the different permutations, it is apparent that accounting for only broad market effects eliminates what little commonality there is in non-index stocks, driving the residual correlations to levels near zero. Unlike the S&P sample, accounting for index variables does not lead to reductions in common effects. There is some role for industry effects, but only when they are added first. This conclusion accords well with the intuition that common order flow effects in the broader sample of stocks, while weaker than those for index constituents, are associated with market-wide and industry factors. We note that the residual order flow correlations for S&P stocks when indexing effects are removed first are on par with the raw order flow correlations for non-S&P stocks (from Table 2). Once again, this is consistent with most of the order flow commonality in S&P stocks being due to indexing effects, which are largely absent in order flow for non-S&P stocks.

As a final comment, Table 3 also shows that the explanatory power of the regressions is greater for S&P stocks. The specifications show an ability to explain 20% (13%) of the order flow variation for individual index stocks during the end (rest) of the day in each year, but only approximately 5% of the order flow variation for individual non-S&P stocks. The greater explanatory power for index constituents reflects both the strong effects of indexing and also, possibly, the fact that indexing factors are easier to capture than are broad market factors. Overall,

[10] A concern is that our results are biased by the sampling procedure for non-S&P stocks (e.g., industry representation in the quintiles is somewhat uneven). To check for bias, we examine the average coefficients from our main tests within each non-S&P volume quintile. The magnitude and significance of the coefficients are similar across the quintiles and show no pattern, which reassures us that the non-S&P results are not biased by our sampling procedure.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

our results demonstrate that the hypothesized common factors—index, industry, and broad market order flow and returns—do well at explaining commonality in order flow. An important benefit of the zero residual correlations is that the statistical tests on the regression coefficients, which assume independence, are correctly specified.[11]

D. The Importance of Correlated Order Flow

The importance of common effects in order flow ultimately rests on the extent to which they impact prices. The main goal of this section is to examine the extent to which correlated returns are driven by correlated order flow. In this context, we compare the relative importance of stock order flow and aggregate order flow. A secondary goal is to assess the predictions of price formation models.

We carry out the analysis using a regression specification in which returns for a stock are related to aggregate S&P, non-S&P and industry order flow, order flow for the stock, and stock order flow interacted with aggregate S&P, non-S&P, and industry order flow,

$$(2)\quad R_{it} = a_0 + a_1 \mathrm{OF}_{\mathrm{SP500},t} + a_2 \mathrm{OF}_{\text{Non-S\&P},t} + a_3 \mathrm{OF}_{\mathrm{Ind},t} + b_0 \mathrm{OF}_{it} + b_1 \mathrm{OF}_{it}\mathrm{OF}_{\mathrm{SP500},t} + b_2 \mathrm{OF}_{it}\mathrm{OF}_{\text{Non-S\&P},t} + b_3 \mathrm{OF}_{it}\mathrm{OF}_{\mathrm{Ind,t}} + e_{it},$$

where R_{it} is the return for stock i in 15-minute period t, OF_{it} is period t order flow for stock i, and $\mathrm{OF}_{\mathrm{SP500},t}$, $\mathrm{OF}_{\text{Non-S\&P},t}$ and $\mathrm{OF}_{\mathrm{Ind},t}$ are aggregate order flow for S&P 500 stocks, non-S&P stocks, and the stock's industry in period t (all variables are standardized).[12]

The specification in (2) allows us to examine rich models of price formation. The common theme in these models is that order flow is associated with price revisions, i.e., causality runs from order flow to price changes. The models differ, however, in their assumptions regarding the information set used by the specialist in setting prices. Traditional single asset microstructure models (e.g., Glosten and Milgrom (1985)) suggest that a specialist conditions prices solely on order flow in his stock, paying no attention to order flow in other stocks. Multi-asset microstructure models (e.g., Caballe and Krishnan (1994)) suggest that the specialist also learns from order flow in related assets. The nature of the specification allows us to examine the predictions of both sets of models. Specifically, if a specialist conditions prices on order flow in his stock, only b_0 (and possibly a_0) will be non-zero. On the other hand, if the specialist adjusts prices on the basis of order flow in other stocks, one or more of the other coefficients will be non-zero, depending on which order flow is informative to the specialist. For instance, if aggregate S&P order flow directly causes revisions in the value of stock i, a_1 will be

[11] Summarizing the effects of the own stock variables, the coefficient on lagged stock returns is negative and that on lagged order flow is positive. This is consistent with prior evidence (e.g., Hasbrouck (1991)).

[12] Specification (2) follows from a model of the form $R_{it} = a_t + b_t \mathrm{OF}_{it} + e_{it}$ where a_t and b_t are time-varying regression coefficients that depend linearly on the levels of S&P, non-S&P, and industry order flow. We also add the first lag of each order flow variable to allow for lagged direct effects on prices. While the coefficients on the lagged order flow terms are presented in Table 5, these terms are suppressed in (2) for ease of presentation.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

non-zero. Similarly, b_1 will be non-zero if there is an interaction effect between order flow for the entire S&P 500 and for stock *i*.

Table 5 presents the results for specification (2). The results are summarized for S&P and non-S&P stocks, for the last hour and the rest of the day, and for 1986 and 1996, with these partitions designed to shed light on the effects of differences in institutional trading intensity. Column 2 shows that, in every case, positive (negative) order flow in a stock leads to positive (negative) returns. We show that incorporating own order flow leads to the most substantial reduction in return comovement in both samples. In other words, the important driver of correlated returns is correlated stock order flow. Columns four through 12 demonstrate the effects of order flow commonalities.

TABLE 5
Return Regressions

		Own		Industry		S&P		Non-S&P		Own∗			
	Inter-cept	OF	OF_{-1}	OF	OF_{-1}	OF	OF_{-1}	OF	OF_{-1}	S&P OF	N S&P OF	Ind OF	Adj. R^2
Panel A. S&P													
1986													
9:30AM–3:00PM	−0.026*	0.215*	−0.022*	0.008*	0.000	0.056*	0.024*	0.103*	−0.050*	−0.024*	0.004*	−0.003	0.098*
	0.000*	0.000*	0.000*	0.000*	0.430	0.000*	0.000*	0.000*	0.000*	0.000*	0.033*	0.434	0.000*
3:00–4:00PM	−0.024*	0.226*	0.000	0.018*	0.000	0.052*	0.018*	0.149*	−0.088*	−0.016*	−0.005	0.000	0.138*
	0.000*	0.000*	0.714	0.000*	0.964	0.000*	0.000*	0.000*	0.000*	0.003*	0.183	0.490	0.000*
1996													
9:30AM–3:00PM	−0.048*	0.378*	−0.021*	0.018*	0.007*	−0.015*	0.062*	0.138*	−0.045*	−0.002	−0.010*	0.009*	0.247*
	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.183	0.000*	0.000*	0.000*
3:00–4:00PM	−0.052*	0.349*	−0.019*	0.020*	0.003*	−0.019*	0.056*	0.169*	−0.050*	−0.001	0.003	0.000	0.253*
	0.000*	0.000*	0.000*	0.000*	0.052*	0.000*	0.000*	0.000*	0.000*	0.556	0.933	0.540	0.000*
Panel B. Non-S&P													
1986													
9:30AM–3:00PM	−0.010*	0.424*	−0.009*	0.002*	0.000	0.019*	0.029*	0.013*	0.002	0.002	−0.010*	0.003	0.182*
	0.000*	0.000*	0.000*	0.037*	0.372	0.000*	0.000*	0.000*	0.352	0.962	0.000*	0.258	0.000*
3:00–4:00PM	0.000	0.426*	0.000	0.006*	0.001	0.015*	0.030*	0.015*	−0.002	−0.003	−0.010*	0.001	0.215*
	0.921	0.000*	0.549	0.066*	0.896	0.000*	0.000*	0.000*	0.331	0.479	0.084*	0.758	0.000*
1996													
9:30AM–3:00PM	−0.019*	0.449*	−0.005*	0.008*	0.005*	−0.016*	0.055*	0.019*	−0.007*	0.013*	−0.015*	0.000	0.217*
	0.000*	0.000*	0.012*	0.000*	0.004*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.752	0.000*
3:00–4:00PM	−0.020*	0.415*	0.000	0.015*	0.003	−0.023*	0.052*	0.019*	−0.003	0.001	−0.004	0.001	0.195*
	0.000*	0.000*	0.559	0.001*	0.958	0.000*	0.000*	0.000*	0.829	0.749	0.921	0.760	0.000*

Table 5 presents estimates of equation (2). OF stands for order flow (the net number of trades). The regressors are the stock's own order flow, industry order flow, aggregate order flow for all stocks in the S&P 500, aggregate order flow for non-S&P stocks, the first lag of these order flow variables (shown as OF_{-1}), and the interactions between stock order flow and industry order flow, aggregate S&P order flow and non-S&P order flow (shown as Own∗Ind OF, Own∗S&P OF, Own∗N S&P OF). The specification is estimated for each stock and for each year, separately using the time series of 15-minute observations covering the hours 9:30AM–3:00PM (22 observations per day) and the hour (3:00–4:00PM) at the end of the day (four observations per day). All variables are standardized by their time-series standard deviations. For example, the dependent variable in the first row is the return for an individual S&P stock in a given 15-minute interval between 9:30AM and 3:00PM in 1986. The cross-sectional medians of the regression coefficients and the adjusted R^2 are presented, along with their *p*-values, for each sample, for 1986 and 1996, and by time of day. *indicates significance at the 10% level or better.

1. S&P Stocks

Considering the sum of the coefficients on contemporaneous order flow and its first lag, it is seen that the direct effect of aggregate industry, indexing, or broad market order flow on returns is positive in both years. Curiously, the effect of broad market order flow exceeds the effect of index order flow which, in turn,

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

exceeds that of industry order flow. Overall, consistent with an informational role for aggregate order flow, it has significant incremental effects on individual stock prices. In economic terms, the impact of a one standard deviation buy imbalance in the S&P index is to raise prices by three basis points in 1986 and by one basis point in 1996. Thus, the direct effect of aggregate imbalances is to create a less favorable environment for individual stock imbalances of the same sign.

In 1986, the negative coefficient on the interaction of stock order flow and S&P order flow implies that the price impact of own order flow is smaller when order flow in the index is of the same sign, presumably because such trades are more likely to be part of a basket trade and hence uninformed. Conversely, if order flow in a stock is of the opposite sign to that in the index, the price impact is larger due to the increased likelihood that it represents informed trading. In economic terms, the price impact of a one standard deviation (three trade) stock buy imbalance is 10 basis points, given a zero order imbalance in the S&P index, and this drops to nine basis points if the S&P experiences a simultaneous one standard deviation buy imbalance. By 1996, the interaction effect, while still negative, is no longer significant, plausibly because the increased scale of index fund trading creates inventory risk for specialists. The effect of non-S&P or industry order flow on the price impact of stock order flow is negligible in both years.

Table 6 reports the residual return correlation when we estimate the model sequentially, starting with own order flow and then adding industry, S&P, and broad market order flow (and the appropriate interaction effects). This table can be interpreted analogously to Table 4. As Panel A shows, the addition of stock order flow leads to a reduction of between 50% and 60% in the return correlation, for instance, from 14% to 6% at the end of the day in 1996. The addition of industry order flow results in a further substantial drop in the correlation, to approximately 2% in each year, and S&P and non-S&P order flow do little to lower these correlations. When we change the ordering and add index order flow followed by industry order flow, the return correlation displays a matching decline at each step. The similar influences of industry and index order flow show that both are important in explaining return comovement. However, own order flow seems to be the principal driver of the return correlation across index constituents.

2. Non-S&P Stocks

The results are in Panel B of Table 5. As with index constituents, aggregate order flow has a direct effect on prices. Again, it is slightly odd that S&P order flow has the largest effect on non-index stock returns, followed by non-S&P and then industry order flow. The negative coefficient on the interaction of stock order flow and non-S&P order flow implies the presence of a basket discount for non-S&P stocks in both years except at the end of the trading day. The absence of a discount at the end of the day is consistent with institutional trading containing information about prices or imposing inventory risk on specialists. Also, the net effect when stock order flow and S&P order flow are of the same sign is positive in 1986 and during 9:30AM–3:00PM in 1996, but is negative at the end of the day in 1996.

Examining the stage-wise residual correlation in Panel B of Table 6, we see that stock order flow brings the return correlation down to less than 0.5% in 1986,

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 6

Marginal Impact of Hypothesized Factors on Common Effects in Returns

	Corr.	Adj. R^2	Corr.	Adj. R^2	Corr.	Adj. R^2	Corr.	Adj. R^2
Panel A. S&P Stocks								
	1986				1996			
	Just Own OF		Just Own OF		Just Own OF		Just Own OF	
9:30AM–3:00PM	0.026	0.068	0.026	0.068	0.060	0.201	0.060	0.201
3:00–4:00PM	0.035	0.089	0.035	0.089	0.062	0.192	0.062	0.192
	Add Industry		Add S&P		Add Industry		Add S&P	
9:30AM–3:00PM	0.016	0.081	0.013	0.092	0.023	0.235	0.021	0.243
3:00–4:00PM	0.016	0.118	0.012	0.133	0.022	0.234	0.021	0.248
	Add S&P		Add Industry		Add S&P		Add Industry	
9:30AM–3:00PM	0.013	0.093	0.013	0.093	0.020	0.245	0.020	0.245
3:00–4:00PM	0.012	0.134	0.012	0.134	0.020	0.251	0.020	0.251
	Add Rest of Mkt.		Add Rest of Mkt.		Add Rest of Mkt.		Add Rest of Mkt.	
9:30AM–3:00PM	0.012	0.098	0.012	0.098	0.019	0.247	0.019	0.247
3:00–4:00PM	0.012	0.138	0.012	0.138	0.020	0.253	0.020	0.253
Panel B. Non-S&P								
	1986				1996			
	Just Own OF		Just Own OF		Just Own OF		Just Own OF	
9:30AM–3:00PM	0.003	0.176	0.003	0.176	0.011	0.206	0.011	0.206
3:00–4:00PM	0.004	0.195	0.004	0.195	0.010	0.178	0.010	0.178
	Add Industry		Add Rest of Mkt.		Add Industry		Add Rest of Mkt.	
9:30AM–3:00PM	0.001	0.177	0.001	0.180	0.004	0.214	0.004	0.214
3:00–4:00PM	0.001	0.203	0.001	0.208	0.004	0.189	0.004	0.189
	Add S&P		Add Industry		Add S&P		Add Industry	
9:30AM–3:00PM	0.001	0.181	0.001	0.179	0.004	0.214	0.003	0.215
3:00–4:00PM	0.001	0.206	0.001	0.215	0.004	0.192	0.003	0.194
	Add Rest of Mkt.		Add S&P		Add Rest of Mkt.		Add S&P	
9:30AM–3:00PM	0.001	0.182	0.001	0.182	0.003	0.217	0.003	0.217
3:00–4:00PM	0.001	0.215	0.001	0.215	0.003	0.195	0.003	0.195

Table 6 is analogous to Table 4, except that the stepwise estimation is performed for equation (2). At each step, we add the order flow and interactive terms in the order shown. For instance, in column 1, we add stock order flow first, followed next by industry order flow and industry order flow interacted with stock order flow, and then by S&P and non-S&P order flows and the associated interaction terms. The final specification corresponds to Table 5.

though it is less successful in explaining return comovement in 1996. When industry or broad market order flow is added next, the return correlation declines 50% further in each year, and the remaining order flow variables (including S&P order flow) do not lower the correlation any further. Thus, own order flow followed by industry and broad market order flow explains return comovement for non-index stocks.

With the exception of S&P stocks in 1986, the average regression R^2 is in excess of 20% in both samples of stocks. Overall, consistent with the premise behind multi-asset models of price formation, aggregate order flow has incremental effects on prices and contributes to return comovement.[13]

[13]We repeat this analysis for buy and sell effective spreads and find that the effects of both stock and aggregate order flow are appreciably weaker. These results are excluded to save space but are summarized in an appendix available upon request.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

E. Principal Components

In this section, we estimate principal components for order flow and returns and compare the extracted principal components with our economic factors to see if there is any correspondence. With principal components, we estimate a set of variables (or factors), smaller than the number of stocks, that collectively maximize the overall explained variation in the series of stock order flow or returns (this is equivalent to least squares, though pooled across stocks and time intervals). Specifically, principal components are mutually orthogonal linear combinations of the stock order flows (or returns) for each interval, with constant weights, that accomplish this goal. We estimate the principal components for each group of stocks, e.g., S&P constituents, and for each year.

The results (not tabulated to conserve space) show that, for both order flow and returns, the first principal component stands out in each year for S&P stocks. In 1986, the first principal component for order flow explains 9% of total variation during 9:30AM–3:00PM and 16% during the last hour of the day. In 1996, the explained fractions rise to 12% and 18%. Displaying similar patterns, the first principal component for returns explains 5% (9%) of return variation during the first 5½ hours (last hour) of the day in 1986 and 13% and 16% in 1996. The remaining principal components are appreciably smaller in each year, explaining in the region of 1% of order flow or return variation.

By contrast, in the non-S&P sample, the explanatory power of the first principal component is weaker, and the difference between the first and subsequent principal components is less pronounced. The first component explains less than 3% (1.5%) of order flow (return) variation in each year, and there is hardly any increase in explained variation at the end of the day or across years.

Table 7 shows the correlation between the first two principal components and the indexing and broad market factors. We do not consider the industry factors in this exercise since they have more localized effects. For S&P stocks, the correlation between the first principal component for order flow and aggregate S&P order flow is 0.98 in 1986 and 0.99 in 1996. This component is also correlated (though less strongly so) with non-S&P order flow. Given the fact that the first principal component is dominant in terms of explained variance, its near-perfect correlation with index order flow reinforces the importance of aggregate index order flow in driving common order flow for index constituents. For non-S&P stocks, the correlation between the first principal component in order flow and aggregate non-S&P order flow is 0.92 in 1986 and 0.95 in 1996. Thus, the dominant principal component in order flow for non-S&P stocks is strongly associated with aggregate broad market order flow. The remaining correlations are much smaller.

We repeat the analysis for the first two principal components extracted from returns for the two samples of stocks. For S&P stocks, the first principal component is highly correlated with aggregate S&P order flow (correlation of 0.82 in 1986 and 0.91 in 1996) while the second principal component is highly correlated with non-S&P order flow, though only in 1986. The other correlations are not large. For non-S&P stocks, the correlation between the first principal component and non-S&P order flow is large (0.82 in 1986 and 0.85 in 1996). Thus, the sta-

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 7

Correlations of Principal Components and Hypothesized Sources of Common Effects

	1986				1996			
	Order Flow		Returns		Order Flow		Returns	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Panel A. S&P								
S&P OF	0.980*	0.103*	0.821*	0.233*	0.988*	0.102*	0.912*	0.048*
	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.000*
Non-S&P OF	0.675*	−0.010	0.538*	0.328*	0.692*	0.086*	0.668*	0.045*
	<0.0001*	0.423	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.000*
S&P Ret$_{-1}$	0.385*	−0.014	0.246*	0.376*	0.152*	−0.134*	0.046*	0.180*
	<0.0001*	0.269	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.000*	<0.0001*
Non-S&P Ret$_{-1}$	0.329*	−0.037*	0.183*	0.281*	0.121*	−0.100*	0.039*	0.173*
	<0.0001*	0.005*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.001*	<0.0001*
Panel B. Non-S&P								
S&P OF	0.665*	0.097*	0.657*	0.234*	0.702*	0.029*	0.721*	−0.092*
	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	0.017*	<0.0001*	<0.0001*
Non-S&P OF	0.919*	−0.008	0.817*	−0.049*	0.948*	0.058*	0.848*	0.051*
	<0.0001*	0.526	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*	<0.0001*
S&P Ret$_{-1}$	0.470*	0.022*	0.489*	0.018	0.346*	0.046*	0.279*	0.026*
	<0.0001*	0.092*	<0.0001*	0.178	<0.0001*	0.000*	<0.0001*	0.033*
Non-S&P Ret$_{-1}$	0.515*	−0.047*	0.467*	−0.059*	0.374*	0.028*	0.281*	0.100*
	<0.0001*	0.000*	<0.0001*	<0.0001*	<0.0001*	0.024*	<0.0001*	<0.0001*

The first two principal components (PC1 and PC2) for order flow and returns are compared to the S&P and Non-S&P (broad market) factors. Their correlations along with the *p*-values of these correlations are presented. *indicates significance at the 10% level or better.

tistical factors in returns are associated mainly with aggregate order flow for the sample under consideration.

Finally, we extract principal components from the residual order flow and return series after controlling for the economic factors. The residuals are obtained from regressions (1) and (2). If our hypothesized sources truly explain common effects, the principal components should become less important. Our untabulated results reveal this to be the case. For instance, the first principal component explains less than 1.5% of the order flow variation in each year for index constituents and less than 0.75% of the variation for non-index stocks. Similar results hold for residual returns. Thus, the economic factors have succeeded in lowering statistically measured common effects in order flow and returns.

V. Common Effects and Index Addition: An Event Study

Our results suggest that indexing is an important source of order flow and return comovement, and imply that comovement with index variables should increase significantly after a stock is added to a major index. This section contains an index addition event study on comovement.

A. Sample and Data

We use a sample of 205 additions to the S&P 500 index over the period 1984–2001. These additions are for NYSE stocks alone, paralleling our analysis in the rest of the paper. We collect the effective date of each addition from biannual

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

editions of the S&P stock guide. The additions are evenly dispersed through time, with a slight increase toward the end of the sample period (not tabulated). Our tests compare intraday comovement over two "steady state" periods (−260, −20) and (+20, +260), where day 0 is the addition date. We exclude the 39-day window around the addition since this period is characterized by heavy rebalancing and is therefore unusual. Comparing the two steady state windows, we find that the cross-sectional median of the median number of trades per day increases from 93 to 154.[14] There is a slightly larger increase in both total trades and order flow in the last hour of the trading day.

B. Results

1. Correlations

Table 8 provides the correlation of stock order flow, returns, and trading costs with the corresponding aggregate variables over the pre- and post-addition periods, and shows that several correlations are dramatically larger in the post-addition period. The correlation of stock order flow with broad market order flow increases by 40%, and that with industry order flow increases by 70%. However, the largest increase occurs in the correlation between stock and index order flow, which more than doubles post-addition. This fact and the fact that the increase in the correlation is largest at the end of the day are consistent with the indexing hypothesis. While the return and trading cost correlations mimic the pattern in the order flow correlation, the increase in each is less pronounced.

2. Intraday Order Flow Regressions

We estimate a variant of equation (1), regressing stock order flow on own lagged returns and order flow, as well as aggregate S&P, non-S&P, and industry order flow and lagged returns (every variable is standardized). In order to capture addition effects, each variable is interacted with an addition dummy, which is one post-addition and is zero otherwise. Pre-addition, a stock is part of the non-S&P population and the broad market effects hypothesis predicts that its order flow will be sensitive to broad market order flow and returns rather than to indexing factors. Once the stock is added to the index, the indexing hypothesis predicts that its order flow will become sensitive to index order flow and returns, and comparatively insensitive to broad market effects.

As the first two columns of Table 9 show, the predictions of the indexing hypothesis are supported by the data (we only tabulate the coefficients on the aggregate variables). Comovement with index order flow is negative to begin with but increases sharply post-addition, especially at the end of the day. The coefficient on index order flow increases by a factor in excess of 10 at the end of the day and by a factor of six during the rest of the day. The coefficient on broad market order flow is significantly above zero before addition but declines sharply post-addition. Interestingly, the overall post-addition effect of broad market order flow is close to zero.

[14] Analysis of the three-day period centered on the addition date shows that the daily median is much larger at 241 trades. This period is also characterized by a spike in buy imbalances, with median (mean) net buys of eight (33) trades per day.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 8

Event Study of Index Addition: Correlations

	Before	After
Correlation between Stock's Order Flow and:		
Industry Order Flow		
9:30AM–3:00PM	0.184	0.309
3:00–4:00PM	0.213	0.374
S&P Order Flow		
9:30AM–3:00PM	0.136	0.290
3:00–4:00PM	0.170	0.369
Non-S&P Order Flow		
9:30AM–3:00PM	0.171	0.215
3:00–4:00PM	0.203	0.281
Correlation between Stock's Return and:		
Industry Return		
9:30AM–3:00PM	0.226	0.259
3:00–4:00PM	0.234	0.306
S&P Return		
9:30AM–3:00PM	0.181	0.258
3:00–4:00PM	0.208	0.312
Non-S&P Return		
9:30AM–3:00PM	0.191	0.207
3:00–4:00PM	0.214	0.252
Correlation between Stock's Effective Spread (Buy) and:		
Industry Effective Spread (Buy)		
9:30AM–3:00PM	0.040	0.041
3:00–4:00PM	0.056	0.078
S&P Effective Spread (Buy)		
9:30AM–3:00PM	0.013	0.014
3:00–4:00PM	0.027	0.031
Non-S&P Effective Spread (Buy)		
9:30AM–3:00PM	0.023	0.021
3:00–4:00PM	0.059	0.064
Correlation between Stock's Effective Spread (Sell) and:		
Industry Effective Spread (Sell)		
9:30AM–3:00PM	0.077	0.090
3:00–4:00PM	0.108	0.121
S&P Effective Spread (Sell)		
9:30AM–3:00PM	0.046	0.060
3:00–4:00PM	0.077	0.075
Non-S&P Effective Spread (Sell)		
9:30AM–3:00PM	0.071	0.072
3:00–4:00PM	0.107	0.117

The sample consists of 205 stocks added to the S&P 500 between 1984 and 2001. Fifteen-minute order flow, returns, and effective spreads are examined for two intraday periods both before and after index addition. The period labeled Before covers days −260 to −20 relative to the addition date and the period labeled After covers days +20 to +260. For example, in the first panel, before being added to the index, the average correlation between an added stock's 15-minute order flow and its industry's order flow between 9:30AM to 3:00PM is 0.184.

The significant increase in the coefficient on industry order flow might appear surprising. Besides the growing importance of industry or sector funds, a possible explanation is that industry funds buy more of a stock after it has been added to the index due to the certification or enhanced awareness conferred by addition. Note that the effect of industry order flow pales in comparison with that of index order flow. The coefficient on S&P returns is close to zero pre-addition but turns negative post-addition, and the coefficients on broad market and industry returns are small and largely unchanged by index addition. Overall, these results—showing increased comovement with index order flow and reduced comovement with broad market order flow, post-addition—strongly support the indexing hypothesis.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 9

Event Study of Index Addition: Effects on Order Flow

	9:30AM–3:00PM	3:00–4:00PM
Intercept	0.0409* [0.000]*	0.0352* [0.000]*
Post-Addition Dummy	0.0132* [0.000]*	0.0084 [0.160]
Industry OF	0.0194* [0.000]*	0.0316* [0.000]*
Industry OF * Dummy	0.0396* [0.000]*	0.0199* [0.041]*
S&P OF	−0.0366* [0.000]*	−0.0306* [0.003]*
S&P OF * Dummy	0.2291* [0.000]*	0.3247* [0.000]*
Non-S&P OF	0.1115* [0.000]*	0.1319* [0.000]*
Non-S&P OF * Dummy	−0.1206* [0.000]*	−0.1517* [0.000]*
S&P Ret_{-1}	0.0030* [0.025]*	0.0034 [0.179]
S&P Ret_{-1} * Dummy	−0.0336* [0.000]*	−0.0326* [0.000]*
Non-S&P Ret_{-1}	−0.0230* [0.000]*	−0.0237* [0.000]*
Non-S&P Ret_{-1} * Dummy	0.0058 [0.281]	−0.0061 [0.477]
Industry Ret_{-1}	0.0115* [0.000]*	0.0036 [0.554]
Ind Ret_{-1} * Dummy	−0.0085* [0.006]*	0.0074 [0.712]
Adj. R^2	0.1430 [0.000]	0.1840 [0.000]

Table 9 summarizes regressions measuring the effects of index addition on order flow comovement. The dependent variable is individual stock order flow in a 15-minute interval. OF stands for order flow (the net number of trades). The regressors are industry order flow, aggregate order flow for all stocks in the S&P 500, aggregate order flow for non-S&P stocks, lagged aggregate returns, and lagged own returns and order flow (the coefficients on own returns and order flow are not tabulated). The estimation period covers days −260 to −20 and +20 to +260 relative to a stock's addition to the S&P 500. The post-addition dummy is one for the latter period and is zero otherwise. All variables are standardized by their time-series standard deviations. The model is separately estimated for two periods, 9:30AM–3:00PM and 3:00–4:00PM, using all the 15-minute observations for these two periods. The cross-sectional medians of the stock-specific regression coefficients and the adjusted R^2 are presented, with *p*-values in brackets. *indicates significance at the 10% level or better.

3. Intraday Return Regressions

Barberis, Shleifer, and Wurgler (2003), henceforth BSW, document that the comovement of stock returns with index returns (measured by beta or by the market model regression R^2) increases after a stock is added to the S&P 500. Paralleling BSW, we estimate the effects of index addition on intraday return comovement via a regression of stock returns on index, broad market, and industry returns, and these returns interacted with an addition dummy. Under the indexing hypothesis, a stock's return sensitivity to index returns should increase, while its sensitivity to broad market returns should decline, after addition.

Using 15-minute returns, we estimate the regression for the first 5½ hours of the day and the last hour of the day, and the results are presented in Table 10. Note that the slope coefficients will be smaller than expected, since the variables are standardized. The estimated slope coefficient on S&P 500 returns is positive even

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

before addition but increases significantly after addition, the median increasing from 0.018 to 0.180 for the first 5½ hours in the day and from 0.03 to 0.214 during the last hour of the day. At the same time, the sensitivity to broad market returns declines sharply post-addition, with the overall post-addition coefficient being slightly negative. The coefficient on industry returns is significantly above zero pre-addition and essentially unchanged by addition. These results, showing a large increase in the sensitivity to index returns, are consistent with BSW's results, and with the indexing hypotheses.

TABLE 10

Event Study of Index Addition: Effects on Returns

	9:30AM–3:00PM	3:00–4:00PM
Intercept	0.0073* [0.000]*	−0.0032 [0.361]
Post-Addition Dummy	−0.0115* [0.000]*	0.0045 [0.195]
Industry Return	0.1483* [0.000]*	0.1609* [0.000]*
Industry Ret * Dummy	0.0152* [0.006]*	0.0075* [0.060]*
S&P Return	0.0175* [0.000]*	0.0296* [0.002]*
S&P Ret * Dummy	0.1618* [0.000]*	0.1842* [0.000]*
Non-S&P Ret	0.0431* [0.000]*	0.0415* [0.000]*
Non-S&P Ret *Dummy	−0.0908* [0.000]*	−0.1201* [0.000]*
Adj. R^2	0.0690 [0.000]	0.0937 [0.000]

Table 10 summarizes regressions measuring the effects of index addition on return comovement. The dependent variable is an individual stock's 15-minute return. The regressors are aggregate 15-minute returns for all stocks in the same industry, the S&P 500, and the population of Non-S&P 500 stocks. The estimation period covers days −260 to −20 and +20 to +260 relative to a stock's addition to the S&P 500. The post-addition dummy is one for the latter period and is zero otherwise. All variables are standardized by their time-series standard deviations. The model is separately estimated for two periods, 9:30AM–3:00PM and 3:00–4:00PM, using all the 15-minute observations for these two periods. The cross-sectional medians of the stock-specific regression coefficients and the adjusted R^2 are presented, with *p*-values in brackets. *indicates significance at the 10% level or better.

4. The Sensitivity of Returns to Order Flow

In a similar specification to (2), we regress stock returns on stock and aggregate order flow, and include interactive terms to capture post-addition effects. To keep matters manageable, we do not include the interaction of stock and aggregate order flow in this specification. Table 11 shows that while own order flow has a positive effect on prices, this effect declines by 25% post-addition. A plausible explanation for this decline is that, once a stock is added to the index, a trade in the stock is more likely than before to be part of a basket trade and therefore uninformed.

The coefficient on industry order flow is positive pre-addition and is unchanged by addition. The coefficient on broad market order flow is positive pre-addition and declines somewhat post-addition. Accompanying this is a significant increase in the effect of index order flow on returns with the coefficient on index order flow increasing by 75% at the end of the day. The enhanced price sensitivity

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

TABLE 11

Event Study of Index Addition: Relating Returns to Order Flow

	9:30AM–3:00PM	3:00–4:00PM
Intercept	−0.0003* [0.000]*	−0.0003* [0.000]*
Post-Addition Dummy	−0.0001* [0.000]*	−0.0002* [0.002]*
Own OF	0.0021* [0.000]*	0.0023* [0.000]*
Own OF * Dummy	−0.0005* [0.000]*	−0.0006* [0.000]*
Industry OF	0.0001* [0.000]*	0.0002* [0.001]*
Industry OF * Dummy	0.0000 [0.478]	0.0000 [0.984]
S&P OF	0.0003* [0.000]*	0.0004* [0.000]*
S&P OF * Dummy	0.0001 [0.119]	0.0003* [0.006]*
Non-S&P OF	0.0002* [0.000]*	0.0003* [0.000]*
Non-S&P OF * Dummy	0.0000* [0.034]*	−0.0001* [0.067]*
Adj. R^2	0.1566 [0.000]	0.1746 [0.000]

Table 11 summarizes regressions measuring the effects of index addition on the return order flow relation. The dependent variable is an individual stock's 15-minute return. The regressors are the stock's own order flow, industry order flow, aggregate order flow for all stocks in the S&P 500, and aggregate order flow for Non-S&P stocks. The estimation period covers days −260 to −20 and +20 to +260 relative to a stock's addition to the S&P 500. The post-addition dummy is one for the latter period and is zero otherwise. All variables are standardized by their time-series standard deviations. The model is separately estimated for two periods, 9:30AM–3:00PM and 3:00–4:00PM, using all the 15-minute observations for these two periods. The cross-sectional medians of the stock-specific regression coefficients and the adjusted R^2 are presented, with *p*-values in brackets. *indicates significance at the 10% level or better.

to index order flow suggests that specialists pay more attention to trading in other index constituents after a stock is added to the S&P 500.[15]

VI. Summary and Implications

In this study, we explore the causes and implications of common effects in order flow. Specifically, we examine the importance of indexing, industry and market-wide influences on order flow, returns and trading costs, while controlling for own stock effects, and use membership in the S&P 500 to identify stocks that are subject to strong indexing effects. In addition, we analyze time-of-day and intertemporal effects, since institutional trading activity is likely to increase at the end of the day and to have been greater in recent years.

Our results show powerful common effects in order flow and returns for S&P 500 constituent stocks, and an increase in the strength of these effects at the end of the day and in 1996 relative to 1986. Consistent with a leading role for index-driven trades in generating correlated order flow, order flow for S&P constituents

[15]Coefficient estimates from the same experiment for effective spreads show that their sensitivity to own order flow declines significantly (by 25%–30%) post-addition. Also, the effect of aggregate index order flow on buy (sell) effective spreads, while not consistently significant pre-addition, is negative (positive) post-addition. These results are consistent with an increased probability of order flow being regarded as uninformed after a stock is added to the S&P 500, and with basket imbalances leading to improved trading costs for trades whose sign matches that of the imbalance.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

is associated mainly with aggregate index order flow and returns. Statistically significant, though economically small, common effects exist in order flow for non-index stocks, driven by industry and broad market order flow and returns. Even though common effects are large only for constituents of the S&P 500, these stocks represent approximately two-thirds of U.S. market capitalization. Most of the return covariation for both S&P and non-S&P stocks is explained by correlated stock-specific order flow and aggregate order flow, and the price effects of stock order flow appear to depend on the level and direction of aggregate order flow. Trading costs show only weak common effects.

Principal components estimation reveals high correlations between the dominant statistically extracted factors in order flow and returns in the two samples and aggregate index and broad market order flow. An event study of index addition shows that stocks added to the S&P 500 index experience increased order flow and return comovement with index order flow and returns and reduced comovement with broad market order flow and returns.

The link between common order flow and returns has practical implications. First, by increasing return correlations, common order flow could be undermining the effectiveness of portfolio diversification and indexing strategies (see related evidence in Barberis, Shleifer, and Wurgler (2003)). Since common effects are strong only for the constituents of major indexes, it should be possible for an investor to mitigate the pricing effects associated with the growing importance of index funds by holding a broader portfolio consisting of non-index stocks (see also Hasbrouck and Seppi (2001) and Chordia et al. (2000), (2002)). Additionally, given specialists' significant price adjustments in the face of aggregate order flow, investors with discretion could time their transactions to take advantage of common order flow. For instance, an investor who wishes to buy shares in an index constituent might want to time this purchase to coincide with aggregate buys (since this leads to more favorable prices).

From an academic perspective, the incremental explanatory power of aggregate order flow for returns has implications for theoretical models of specialist behavior. Specifically, this result suggests that the specialist learns from order flow in other stocks, and supports multi-asset models of price formation such as Caballe and Krishnan (1994) and others. To our knowledge, the only study of basket trading effects on individual stocks is that of Hasbrouck (1996) who, using TORQ data, finds that program and non-program trades have similar price impacts. Our results also contribute to the academic debate on what drives return covariation. We find that order flow accounts for a huge fraction of the correlation in returns. This leaves only a small fraction of return comovement to be explained by, for instance, pure public information, which moves prices without inducing order flow.

Lastly, our results showing that common effects in trading costs are weak relative to those in order flow or returns are consistent with Hasbrouck and Seppi, who also employ intraday frequencies. Chordia et al. (2000), by contrast, document stronger common effects in changes in daily measures of trading costs. Future research should be directed at understanding the reasons for the divergence in results, with potential explanations including differences in frequency of aggregation and in estimation strategy.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press

References

Admati, A. "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities." *Econometrica*, 53 (1985), 629–658.

Admati, A., and P. Pfleiderer. "A Theory of Intraday Patterns: Volume and Price Variability." *Review of Financial Studies*, 1 (1988), 3–40.

Barberis, N.; A. Shleifer; and J. Wurgler. "Comovement." Working Paper, Harvard, Chicago, NYU (2003).

Caballe, J., and M. Krishnan. "Imperfect Competition in a Multi-Security Market with Risk Neutrality." *Econometrica*, 62 (1994), 695–704.

Chordia, T.; R. Roll; and A. Subrahmanyam. "Commonality in Liquidity." *Journal of Financial Economics*, 56 (2000), 29–64.

_____. "Order Imbalance, Liquidity and Market Returns." *Journal of Financial Economics*, 65 (2002), 111–130.

Domowitz, I., and X. Wang. "Liquidity, Liquidity Commonality and Its Impact on Portfolio Theory." Working Paper, Penn State (2002).

Easley, D.; M. O'Hara; and P. S. Srinivas. "Option Volume and Stock Prices: Evidence on where Informed Traders Trade." *Journal of Finance*, 53 (1998), 431–465.

Edelen, R. "Investor Flows and the Assessed Performance of Open-End Mutual Funds." *Journal of Financial Economics*, 53 (1999), 439–466.

Edelen, R., and J. Warner. "Aggregate Price Effects of Institutional Trading: A Study of Mutual Fund Flow and Market Returns." *Journal of Financial Economics*, 59 (2001), 195–220.

Glosten, L., and P. Milgrom. "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *Journal of Financial Economics*, 14 (1985), 71–100.

Hasbrouck, J. "Measuring the Information Content of Stock Trades." *Journal of Finance*, 46 (1991), 179-207.

_____. "Order Characteristics and Stock Price Evolution: An Application to Program Trading." *Journal of Financial Economics*, 41 (1996), 129–149.

Hasbrouck, J., and D. Seppi. "Common Factors in Prices, Order Flows and Liquidity." *Journal of Financial Economics*, 59 (2001), 383–411.

Huberman, G., and D. Halka. "Systematic Liquidity." *Journal of Financial Research*, 24 (2001), 161–178.

Kumar, P., and D. Seppi. "Information and Index Arbitrage." *Journal of Business*, 67 (1994), 481–509.

Lee, C., and M. Ready. "Inferring Trade Direction from Intraday Data." *Journal of Finance*, 46 (1991), 733–746.

Lewellen, J. "Momentum and Autocorrelation in Stock Returns." *Review of Financial Studies*, 15 (2002), 533–563.

Lo, A., and J. Wang. "Trading Volume: Definitions, Data Analysis and Implications of Portfolio Theory." *Review of Financial Studies*, 13 (2000), 257–300.

Moskowitz, T., and M. Grinblatt. "Do Industries Explain Momentum?" *Journal of Finance*, 54 (1999), 1249–1290.

O'Hara, M. "Presidential Address: Liquidity and Price Discovery." *Journal of Finance*, 58 (2003), 1335–1354.

Subrahmanyam, A. "A Theory of Trading in Stock Index Futures." *Review of Financial Studies*, 4 (1991), 17–51.

Tkac, P. "A Trading Volume Benchmark: Theory and Evidence." *Journal of Financial and Quantitative Analysis*, 34 (1999), 89–114.

https://doi.org/10.1017/S0022109000001733 Published online by Cambridge University Press