University of Alberta

Three Essays on the Interplay between Trading and Business Conditions

by

Nuri Volkan Kayacetin

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Finance

Faculty of Business

Spring 2010
Edmonton, Alberta

Examining Committee

Aditya Kaul, Department of Finance and Management Science

Vikas Mehrotra, Department of Finance and Management Science

Akiko Watanabe, Department of Finance and Management Science

Karim Jamal, Department of Accounting and Management Information Systems

Valentina Galvani, Department of Economics

Tarun Chordia, Department of Finance, Emory University

Dedication

This thesis is dedicated to my wonderful parents, Nuray and Adnan Kayacetin, without whose unconditional love and support it would not have been possible, and to the memory of my grandfathers, Hasan Ciftci and Nuri Kayacetin, who have been with me, watching over me, in every step I take.

Abstract

The first essay provides evidence on the origins of the size and value premiums by examining how order flow in the SMB and HML portfolios relates to economic conditions and investor sentiment. We find that buying pressure for both SMB and HML is lower (increases) when economic conditions are expected to deteriorate (improve), while it is unrelated to proxies for investor sentiment and sales growth. These findings are consistent with big stock and value stocks being regarded as hedges against adverse shifts in economic conditions, and support a rational state variable interpretation of the size and value premiums.

The second essay finds that the marketwide average of individual stock order flows and the difference between the average order flow for big stocks and the average order flow for small stocks (order flow differential) predict growth rates in real GDP, industrial production, and corporate earnings. The predictive significance of these two measures is robust to controls for return factors, suggesting a role for order flow in forecasting stock returns. Consistently, we show that an increase in the order flow differential forecasts higher returns for ten size-sorted portfolios and significantly greater market and size premiums in the subsequent quarter, even after accounting for a large host of variables. These findings are consistent with a world where aggregate order flow brings together dispersed information from heterogeneously informed investors.

The third essay shows that stocks that are harder to value (stocks with less valuable growth options and more dispersed analyst forecasts) and stocks that attract less uninformed trading activity (small stocks, illiquid stocks, stocks not covered by analysts) have higher price impacts, greater probabilities of informed trading, and more private information in returns. In the time-series, reductions in trading activity and consumer sentiment increase the average price impact of trading and reduce the share of firm-specific information in returns. Recessions see high price impacts, low trading activity, and a smaller share of private signals in price movements. This reduction in private information seems to have an impact on the informativeness of prices for corporate managers: the sensitivity of corporate investment to the prices is significantly lower during recessions.

Acknowledgement

I am deeply grateful to my supervisor, Dr. Aditya Kaul, for the unceasing support and continuous encouragement he has provided me with throughout my doctoral studies. This work has benefited greatly from the many inspiring discussions we had and the pursuit of perfection that he has brought to any task that he undertakes. May our best work together always be the next one!

I also wish to express my gratitude to Dr. Vikas Mehrotra, Dr. Akiko Watanabe, Dr. Valentina Galvani, Dr. Karim Jamal, and Dr. Tarun Chordia, the distinguished members of my dissertation committee, for the constructive criticism, insightful comments, and valuable suggestions they have provided during the writing of this thesis. This work would have not been complete without them.

Of course, thanks a million to all my colleagues in the PhD program, some of them doctors already and some are soon to be for the wonderful experiences we had together, and to my friends in Edmonton, for making my stay here full of fond memories, laughter, and love.

I am indebted to the University of Alberta School of Business faculty and staff members for being valuable parts of the friendly and stimulating atmosphere that nurtured my interpersonal and intellectual qualities and helped me cope with the tasks I had assumed during the course of my doctoral studies. Special thanks to Jeanette Gosine, Kathy Harvey, Keltie Tolmie and Louise Herbert for being wonderful.

Lastly, and most importantly, I thank my parents: Nuray and Adnan Kayacetin, my brother: Cihan Kayacetin, my cousins: Hakan and Kaan Gungor and Nihal and Mustafa Koc, my uncles: Bulent Berkoz, Halis Gungor, and Yasar and Recep Ciftci, my aunts: Nigar Gungor and Aysen Berkoz, and my grandmothers: Fahriye Ciftci and Methiye Kayacetin, for being parts of our big and happy family. The many smiles that you have put on my face can never be erased!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis is comprised of three essays, the common theme of which is the interplay between trading activity in the stock market and business conditions. The objective of these three essays is to document, critically evaluate, and explain the effects of business fluctuations on investors' trades in the stock market and the implications of the resulting pattern of behavior on expected stock returns, market liquidity, and the informational efficiency of prices.

The **first essay** characterizes how investors trade the small-minus-big (SMB) and high-minus-low (HML) portfolios in response to business fluctuations, since the origins of the respective size and value premiums, *SMB* and *HML*, captured by these portfolios are still the subject of debate. Fama and French (1996) suggest that these premiums arise as rational investors seek protection against adverse changes in underlying state variables. Lakonishok, Shleifer, and Vishny (1994) argue that *HML* is the outcome of irrational extrapolation of recent earnings performance. While Lakonishok et al. do not focus on the size premium, their irrational extrapolation hypothesis can be extended to explain why small and large firms might have different returns. Hence, *SMB* is also under scrutiny.

We approach this issue from a new angle by analyzing how investors trade the portfolios that capture the size and value premiums. Our hypothesis is that if SMB and HML are hedge portfolios, investors should buy or sell these portfolios based on their expectations about economic conditions. By contrast, if behavioral biases are at work, their trades would be myopic, based more on the current state of the economy and investor sentiment and less on future economic conditions. To test these hypotheses, we estimate monthly order flows for SMB and HML (denoted *OFSMB* and *OFHML*) over the period January 1988 through December 2004 and relate these order flows to proxies for rational and behavioral influences. The proxies for rational effects include the default and term spreads (*DEF* and *TERM*) to capture investor expectations about future economic conditions, implied volatility from CBOE index options (*VXO*) to capture aggregate uncertainty, and the Pastor-Stambaugh (2003) liquidity measure to capture marketwide liquidity (*LIQ*). As proxies for behavioral effects, we include the National Association of Purchasing Managers Index (*NAPM*) to measure aggregate sales performance, and the Baker-Wurgler (2006) investor sentiment measure (*SENT*). As a benchmark, we also analyze the order flow for a portfolio of all New York Stock Exchange (NYSE) stocks (*OFMKT*).

Our results favor the rational asset pricing view rather than the behavioral view. None of the three aggregate order flow series is related to investor sentiment while *NAPM* is significant only for *OFHML*. By contrast, *OFSMB* and *OFHML* are related strongly to the proxies for expected economic conditions, *DEF* and *TERM*, and *OFHML* is also influenced by aggregate uncertainty and marketwide liquidity, *VXO* and *LIQ*. Both *OFSMB* and *OFHML* decline significantly in the face of an anticipated deterioration in economic conditions while the reverse happens when conditions are expected to improve. Examining size and book-to-market (BM) subcategories, we find that these patterns originate from investors' tendency to buy (sell) small, high BM stocks when economic prospects improve (worsen). We also find that *OFMKT* is related positively to *TERM*, a reasonable relation between aggregate buying pressure for stocks and economic conditions.

Further analysis indicates that these relations between order flows and the business cycle variables are not driven by investor over-reaction to business conditions.

When we repeat this analysis for the three return factors (*MKT*, *SMB*, and *HML*), we find that an increase in *TERM* leads to higher market and SMB returns. In addition, aggregate volatility has explanatory power for *HML*, with an increase in *VXO* leading to a decline in *HML*. Last, there is a negative relation between *SMB* and *SENT*. Overall, the return results are not as clear as the order flow results, suggesting that business-cycle effects are more visible in order flows than returns. The weaker results for returns could be explained by two confounding return effects. For instance, an increase in *TERM* in month *t* will lead to a higher return in month *t*+1 if this (good) news is incorporated into stock prices with a lag or to a lower return (due to the countercyclical nature of expected returns) if prices have already incorporated the signal. Month *t*+1 order flow, by contrast, will be positive if investors react with a lag and random otherwise.

Our analysis yields other results of interest. First, we find that (a) individual stock order flows display additional comovement within their size-BM groups, and (b) stocks that switch size/BM categories see their order flow comove more with the order flow of the portfolio they move to and less with the order flow of the portfolio they leave. This evidence of additional comovement in trades related to size and BM extends prior research by Hasbrouck and Seppi (2001) and Harford and Kaul (2005) showing common effects in order flows. It also is in line with the results in Huberman and Kandel (1987) and Chan and Chen (1991), who document covariation in returns related to size and relative distress not captured by the market return. Second, we quantify the amount of variation in size/BM-sorted portfolio returns that can be attributed to comovement in order flows. Here, we compare the explanatory power of the Fama-French (1993) three-factor model for portfolio returns before and after the portfolio returns are adjusted for trading effects, and find an economically material decline in the model R^2s of one-third or more. This is consistent with the intuition that the factor structure of trades will be inherited by returns unless market-makers purge the effects of the trade factors while setting prices (e.g. Kyle, 1985; Hasbrouck and Seppi, 2001).

In a setting where information is distributed heterogeneously across agents, net order flow for broad portfolios may aggregate dispersed information and provide a valuable signal about how investors bet on their expectations about fundamentals with their wallets. Indeed, a recent literature provides evidence that aggregate order flow in the foreign exchange and bond markets reveals information about macroeconomic fundamentals (Brandt and Kavajecz, 2004; Green, 2004; Pasquariello and Vega, 2006; Evans and Lyons, 2009; Beber, Brandt, and Kavajecz, 2008). My **second essay** adds to this literature by investigating the predictive power of equity market order flow for (a) macroeconomic growth and (b) stock returns.

Our analysis in this part focuses on two distinct aggregate order flow measures. The first measure, *market order flow* (*OFM*), is the cross-sectional average of individual stock order flows estimated from intraday trade and quote data using the Lee-Ready (1991) algorithm. *OFM* parallels the aggregate bond and foreign exchange market order flows studied elsewhere, and captures overall buying or selling pressure exerted by trade initiators who place market orders and demand immediacy. Provided that there is a class of investors who trade solely for liquidity reasons, we conjecture that *OFM* should reflect the exchange that take place between these liquidity traders and relatively more sophisticated portfolio optimizers that is brought about by the effect of changing consumption and investment opportunities on the optimal portfolio allocation to stocks.

The second measure, *order flow differential* (*OFD*), is novel and specific to the stock market. We define *OFD* as the difference between the average buying pressures generated by the active big and small stock traders in a given period. We conjecture that this measure should capture the time variation in intertemporal hedging demand induced by the strategic behavior of investors who wish to hedge against adverse changes in future consumption and investment opportunities. That is, as small stock returns are more sensitive to marketwide fluctuations than big stock returns, deterioration of economic expectations (or an accompanying increase in risk aversion) should result in a disproportionate decline in the fraction of wealth allocated to small stocks in relation to the fraction of wealth allocated to big stocks. This would lead to an exchange of securities between sophisticated hedgers and liquidity traders, which is eventually picked up by *OFD*.

Using stock-level order flows constructed from high frequency data, we compile the two order flow aggregates quarterly over the period January 1988 through December 2004. We start our analysis by examining the predictive power of *ODM* and *ODD* for future economic output growth, as measured by the quarterly growth rates of real GDP, industrial production, and corporate earnings (*QPG*, *QYG*, and *QEG*). Our results show that *OFM* is related positively to future growth rates for real GDP and industrial production, but not for corporate earnings: a one standard deviation increase in *OFM* forecasts an increase of about 0.21 to 0.38 standard deviations in *QPG* (0.23 to 0.41 percent) and about 0.18 to 0.35 standard deviations in *QYG* (0.10 to 0.18 percent) over the four subsequent quarters. *OFD*, on the other hand, is related negatively to all the three proxies for future economic growth and its predictive power is even stronger: a one standard deviation increase in *OFD* forecasts a decline of about 0.33 to 0.48 standard deviations in *QPG* (0.36 to 0.51 percent), 0.26 to 0.39 standard deviations in *QYG* (0.13 to 0.20 percent), and 0.21 to 0.31 standard deviations in *QEG* (1.08 to 1.64 percent). These relations are robust to the inclusion of the lagged economic growth rates and contemporaneous return factors from a four-factor model including the excess market return and the size, value, and momentum premiums.

The findings above parallel the evidence from the foreign exchange market reported in Evans and Lyons (2009) that the information in order flows is not captured by returns. A potential explanation, suggested by the work of Chan (1993), is that market makers are unable to immediately extract the marketwide component of a noisy firm-level signal (embedded in order flow, in our case) and, instead, assimilate this information over time as they learn from the signals of other stocks in subsequent periods. Note that the noise thus induced in returns may not wash away in aggregation if it is correlated across market-makers. The hypothesis that follows from this reasoning is that, if the macroeconomic signal in the order flow measures is not impounded in prices in a timely manner, our aggregate order flow measures should predict stock market returns. To address this issue, we regress the future quarterly returns for ten size-sorted portfolios and the future realizations of the market, size, value, and momentum premiums on the quarterly changes in *OFM* and *OFD*. We expect a positive relation between *OFD* and expected returns as the hedging component of demand will be more pronounced when future outlook is dim and risk aversion is high. The relation between *OFM* and expected returns, on the other hand, can go either way: it may be positive if the information in market order flow is only partially incorporated into prices (the noisy macro signal story) or negative because of the negative link between realized and expected returns. In order to ensure that the information in the two order flow variables are unique, we extend the set of control variables with several business cycle indicators (default spread, term spread, forecasted earnings growth, and new equity additions: *DEF*, *TERM*, *FEG* and *NEQ*), proxies for liquidity and investor sentiment (Pastor-Stambaugh (2003) liquidity measure and the Baker-Wurgler (2006) sentiment index: *LIQ* and *ΔSENT*), and lagged portfolio returns.

Our tests reveal that *OFM* and *OFD do* have significant predictive power for stock market returns. An increase in *OFM* in quarter *t* forecasts higher quarter *t*+1 returns for most size-sorted decile portfolios (with the exception of the three largest portfolios), but not for any of the four return premiums. Controlling for *OFD*, a one standard deviation change in *OFM* forecasts an increase of 0.21 to 1.41 percent (0.03 to 0.13 standard deviations) in the decile portfolio returns. This forecast power, however, is mostly subsumed when contemporaneous return factors are added to the model as controls, and disappears totally with the inclusion of the rest of the control variables. Unlike *OFM,* the forecast power of *OFD* is robust to the inclusion of the contemporaneous return factors, business-cycle indicators, marketwide liquidity, and investor sentiment. Keeping all else constant, a one percent increase in *OFD* forecasts an increase of 0.40 percent (0.15 standard deviations) in the excess market return, 0.44 percent (0.24 standard deviations) in *SMB* , and a rise between 0.76 and 2.79 percent (0.11 and 0.21 standard deviations) in the decile portfolio returns.

The positive relation between *OFM* and subsequent small stock returns shows that it takes time for small stock prices to fully reflect the signal embedded in marketwide order flow. This is (a) plausible as the macro signal would be easier to detect for market makers in big stocks since the noise is diversified to a certain extent due to the greater scale of such firms' operations and (b) consistent with the lead-lag relation between big and small stock returns first documented in Lo and MacKinlay (1990). The finding that the explanatory power of *OFM* is subsumed when we account for liquidity and other controls is in line with Albuquerque et al. (2008), who find that a simple statistical factor of equity-market order flows captures mostly liquidity. The strength and robustness of the forecast power of the order flow differential across size deciles, on the other hand, signals a more pervasive effect. The positive relation between *OFD* and subsequent returns is consistent with investors reallocating portfolios from more to less procyclical assets as risk aversion increases prior to economic downturns. Further investigation confirms that the observed effect is distinct from liquidity: a size-controlled order flow differential between liquid and illiquid stocks behaves much like *OFM* does, and fails to achieve the strong explanatory power displayed by *OFD*. Ultimately, the evidence that the information in our aggregate order flows is not incorporated into stock prices for extended periods is striking. In particular, it is intriguing that common return factors, including the excess market return and *SMB* —closely linked to *OFM* and *OFD*—do not subsume the signal in aggregate order flows. To the best of our knowledge, this paper is the first to analyze the predictive content of equity market order flows for fundamentals and expected stock returns.

A recent paper by Beber, Brandt, and Kavajecz (2008) analyzes order flow movements across sectors of the economy and shows that a portfolio based on cross-sector order flows dominates the market portfolio, particularly during economic downturns. We view the two papers as complementary, and our analysis differs from theirs in several respects. First, Beber et al. (2008) use sector order flows and returns to predict the Chicago FED National Activity Index and stock and five-year bond returns, while we use aggregate order flow to predict real GDP, production, and earnings growth. Second, we introduce *OFD* as a novel proxy that captures time-variation in intertemporal hedging demand and forecasts future fundamentals and stock returns. Third, by controlling for a host of economic indicators and return factors, we verify the uniqueness of the signal contained in our measures. The results in Beber et al. (2008) are mostly from univariate relations between sector order flows, macro fundamentals, and returns.

In their seminal paper, Grossman and Stiglitz (1980) argue that securities markets are characterized by an "equilibrium degree of disequilibrium, where security prices reflect the information of informed individuals, but only partially, so that those who expend resources to obtain information receive compensation." This equilibrium degree of disequilibrium is characterized by a trade-off between the informativeness of the price system and the incentives that the individuals in the system have for acquiring private information. Clearly, in a market state where prices are perfectly informative, there is no room for arbitrageurs to function, and in a market state where arbitrageurs do not function, it is paradoxical to have perfectly informative prices. Based on their model, Grossman and Stiglitz (1980) make three propositions relevant to our paper. They posit (a) as the proportion of informed individuals increases, the price system becomes more informative, (b) a greater level of noise would render the price system less informative for uninformed individuals and lead to an increase in the proportion of individuals who are informed, and (c) as the cost of obtaining information increases, the equilibrium proportion of individuals who are informed will be smaller.

Do higher trading costs hinder arbitrage activity and reduce the amount of private information generated by arbitrageurs? Does greater private information generation increase trading costs through alleviating the adverse selection problem faced by market makers? What is the role of liquidity in determining the nature of these interrelationships? How do liquidity, trading costs, and private information generation vary across the business cycle? Are there real effects associated with reductions in private information generation? The **third essay** tries to address these questions through investigating the interrelationship between trading costs, trading activity, and the share of firm-specific information in price movements over an 83-year period from 1926 to 2008, focusing on business-cycle patterns in these variables. Trading activity and trading costs are proxied, respectively, by share turnover (STO) and the average price impact of trading (PIM), defined as the change in price implied by a $1 million trade (in 2008 dollars) following Amihud (2002). Consumer sentiment (SEN) is used as an instrument to capture the general mood of individuals in the economy. Following the insights in Roll (1988) and Morck et al. (2000), the share of firm-specific information in price movements (FSI), computed as one minus the market model R^2, is used as a measure of the informational efficiency of the pricing system. Following Chen, Goldstein, and Jiang (2007), the informativeness of prices for corporate managers is measured as the sensitivity of corporate investment in a given year to the normalized price (Tobin's Q) at the end of the previous year. I also use data on financial analysts' forecasts to construct additional measures of firms' informational environment. These measures are the number of analysts providing earnings-per-share estimates for a given firm (NUM) and the dispersion of these analysts' forecasts (FDISP).

My main results are as follows. I first show that both price impact and FSI display discernible business-cycle patterns, with PIM increasing, and FSI declining significantly during recessions. Two-way causality tests in the spirit of Granger (1969) and Sims (1972) reveal, at the market-level, FSI is caused by consumer sentiment and price impact, and price impact is in turn caused by sentiment and share turnover. Time-series tests of the interrelationship between these variables indicate that FSI is related negatively to contemporaneous and lagged changes in sentiment and PIM, while PIM itself is related negatively to STO and sentiment. Next, I study PIM, the probability of informed trading (PIN), and FSI at the firm level and investigate possible sources of cross-sectional variability in these variables. Here, I find that PIM and PIN are lower for big firms, growth firms, firms whose stock is more liquid, firms with extensive analyst coverage, and firms with lower analyst forecast dispersion. Note that these are stocks that potentially grab the attention of liquidity traders (e.g. Barber and Odean, 2008). In line with this observation, I show that such stocks experience the greatest increase in the probability of informed trading as a recession hits the economy. FSI, on the other hand, is greater for small stocks, value stocks, less liquid stocks, and for stocks with higher trading costs, more disperse analyst forecasts, or little or no analyst coverage.

Collectively, these findings are consistent with a world where a decline in uninformed investor activity aggravates the adverse selection problem faced by market-makers–an effect that is distinct from an increase in the adverse selection problem due to greater informed trading activity. The market-makers rationally respond by adjusting their pricing functions, driving up trading costs. The increase in trading costs reduces the amount of firm-specific information that is incorporated into stock prices through informed trading, since a certain fraction of the signals that used to be profitable in low trading cost regimes will not be worth trading on based on the new, and worsened, terms of trade. In the end, we face a strategic interaction where market-makers know that a trade, if executed, is more likely to come from an informed trader and informed traders know that a trade, if executed, will be less profitable since the market maker knows that the trade is more likely to be information-based. The end-result of this interaction between informed traders and the market-maker is an equilibrium which is optimal for both parties playing this game, but potentially suboptimal for the informational efficiency of the market since part of the relevant firm-specific information is left unincorporated into stock prices. This reduction in information-based trading appears to have a material effect on the informativeness of prices for corporate managers. The sensitivity of corporate investment to the information in prices is 62.5% lower in recessions, underlining the importance of well-functioning, informationally-efficient securities markets.

In a related study, Chordia, Roll, and Subrahmanyam (2001) conduct a higher frequency analysis of aggregate market spreads, depths, and trading activity for U.S. stocks over the period 1988 through 1998. The authors demonstrate that the increase in spreads in down markets is greater in magnitude than the decline in spreads during up markets, while the effect of up and down markets on trading activity is roughly symmetric. This asymmetric relation between the marketwide averages for spreads and returns is consistent with the notion in our paper that the greater price impacts during recessions come about as a result of a decline in uninformed trading activity instead of an increase in informed arbitrage activity.

To the best of my knowledge, this is the first paper to analyze the interrelationship among trading costs, trading activity, and firm-specific information dissemination over a sample period as long as ours. The results of this analysis are important for several reasons. First, delineating the link between informed trading and adverse selection problem in financial markets is beneficial for future research. My results imply that the activity of uninformed traders has an important influence on measures of adverse selection such as PIN and PIM, an influence distinct from that of the intensity of information-based trading. Second, by characterizing the business-cycle patterns in trading costs and private information generation, we provide perspective on the deadweight costs of recessions for the functioning of financial markets. In doing so, we quantify the approximate effect of the high trading cost – low informed arbitrage activity market regimes observed during recessions on the informativeness of prices in guiding corporations' investment decisions. The significant decline in private information generation and the accompanying reduction in price informativeness suggest that promoting trading activity by households and other non-arbitrageurs at all times is critical for the informational efficiency of financial markets. In this sense, the role played by financial analysts in encouraging their clienteles to trade more actively may paradoxically be benefiting the market as a whole. Policy-makers may, hence, find it worthwhile to promote trading activity in financial markets through improving shareholder property rights and providing unsophisticated investors with easier-access investment vehicles in order to attract greater non-arbitrage demand.

Chapter 2

Trade-Based Origins of the Size and Value Premiums

2.1 Introduction

The origins of *SMB* and *HML* are still the subject of debate.[1] Fama and French (1996) suggest that these premiums arise as rational investors seek protection against adverse changes in underlying state variables. Lakonishok, Shleifer, and Vishny (1994) argue that *HML* is the outcome of irrational extrapolation of recent earnings performance. While Lakonishok et al. do not focus on the size premium, their irrational extrapolation hypothesis can be extended to explain why small and large firms might have different returns. Hence, *SMB* is also under scrutiny.

[1] SMB (small-minus-big) is a portfolio long in small stocks and short in big stocks. HML (high-minus-low) is a portfolio long in high book-to-market stocks and short in low book-to-market stocks. We use italics, *SMB* and *HML*, to denote the returns for these portfolios.

In this paper, we approach this issue from a new angle by analyzing how investors trade the portfolios that capture the size and value premiums. Our hypothesis is that if SMB and HML are hedge portfolios, investors should buy or sell these portfolios based on their expectations about economic conditions. By contrast, if behavioral biases are at work, their trades would be myopic, based more on the current state of the economy and investor sentiment and less on future economic conditions. To test these hypotheses, we estimate monthly order flows for SMB and HML (denoted *OFSMB* and *OFHML*) over the period January 1988 through December 2004 and relate these order flows to proxies for rational and behavioral influences.[2] The proxies for rational effects include the default and term spreads (*DEF* and *TERM*) to capture investor expectations about future economic conditions, implied volatility from CBOE index options (*VXO*) to capture aggregate uncertainty, and the Pastor-Stambaugh (2003) liquidity measure to capture marketwide liquidity (*LIQ*). As proxies for behavioral effects, we include the National Association of Purchasing Managers Index (*NAPM*) to measure aggregate sales performance, and the Baker-Wurgler (2006) investor sentiment measure (*SENT*). As a benchmark, we also analyze the order flow for a portfolio of all New York Stock Exchange (NYSE) stocks (*OFMKT*).

Our results favor the rational asset pricing view rather than the behavioral view. None of the three aggregate order flow series is related to investor sentiment while *NAPM* is significant only for *OFHML*. By contrast, *OFSMB* and *OFHML* are related strongly to the proxies for expected economic conditions, *DEF* and *TERM*, and *OFHML* is also influenced by aggregate uncertainty and marketwide liquidity, *VXO* and *LIQ*. Both *OFSMB* and *OFHML* decline significantly in the face of an anticipated deterioration in economic conditions while the reverse happens when conditions are expected to improve. Examining size and book-to-market (BM) subcategories, we find that these patterns originate from investors'

[2] We identify each trade as buyer- or seller-initiated using the Lee-Ready (1991) algorithm. For each stock, order flow is computed as the difference between monthly buy and sell volumes, scaled by the total number of shares traded. These stock-level order flows are then aggregated to obtain portfolio order flows.

tendency to buy (sell) small, high BM stocks when economic prospects improve (worsen). We also find that *OFMKT* is related positively to *TERM*, a reasonable relation between aggregate buying pressure for stocks and economic conditions. Further analysis indicates that these relations between order flows and the business cycle variables are not driven by investor over-reaction to business conditions.

When we repeat this analysis for the three return factors (*MKT*, *SMB*, and *HML*), we find that an increase in *TERM* leads to higher market and SMB returns. In addition, aggregate volatility has explanatory power for *HML*, with an increase in *VXO* leading to a decline in *HML*. Last, there is a negative relation between *SMB* and *SENT*. Overall, the return results are not as clear as the order flow results, suggesting that business-cycle effects are more visible in order flows than returns. The weaker results for returns could be explained by two confounding return effects. For instance, an increase in *TERM* in month t will lead to a higher return in month t+1 if this (good) news is incorporated into stock prices with a lag or to a lower return (due to the countercyclical nature of expected returns) if prices have already incorporated the signal. Month t+1 order flow, by contrast, will be positive if investors react with a lag and random otherwise.

Our analysis yields other results of interest. First, we find that (a) individual stock order flows display additional comovement within their size-BM groups, and (b) stocks that switch size/BM categories see their order flow comove more with the order flow of the portfolio they move to and less with the order flow of the portfolio they leave. This evidence of additional comovement in trades related to size and BM extends prior research by Hasbrouck and Seppi (2001) and Harford and Kaul (2005) showing common effects in order flows. It also is in line with the results in Huberman and Kandel (1987) and Chan and Chen (1991), who document covariation in returns related to size and relative distress not captured by the market return. Second, we quantify the amount of variation in size/BM-sorted portfolio returns that can be attributed to comovement in order flows.

Here, we compare the explanatory power of the Fama-French (1993) three-factor model for portfolio returns before and after the portfolio returns are adjusted for trading effects, and find an economically material decline in the model R^2s of one-third or more.[3] This is consistent with the intuition that the factor structure of trades will be inherited by returns unless market-makers purge the effects of the trade factors while setting prices (e.g. Kyle, 1985; Hasbrouck and Seppi, 2001).

The rest of the paper is organized as follows. Section 2 discusses our research questions and the relevant literature. Section 3 describes the data and variables. Section 4 examines the existence and pricing effects of size and BM-driven commonality in order flows. Section 5 studies the determinants of the common factors in order flows and returns. Section 6 concludes.

2.2 Literature Review and Research Questions

2.2.1 Size and Value Commonality in Order Flow

Order flow measures the net buying pressure faced by passive liquidity suppliers. The link between order flow and daily and intraday price movements is well-established.[4] One strand of research relevant to our paper studies common effects in order flow and returns. Hasbrouck and Seppi (2001) show that intraday order flows and returns for the Dow Jones constituents are characterized by common factors and that roughly two-thirds of the commonality in returns is due

[3] In order to mitigate the simultaneity between order flows and returns, we adjust the order flow factors for each portfolio by excluding the order flow for the portfolio under consideration. The trade-adjusted portfolio returns are defined as the residuals from regressions of portfolio returns on the innovations in the three order flow factors.

[4] Chan and Fong (2000) find that order flow explains a substantial portion of daily price movements. Chordia, Roll, and Subrahmanyam. (2002) document a strong relation between returns and order flow at the market level and show that order flow is high (low) after negative (positive) market returns. Chordia and Subrahmanyam (2004) focus on the relation between daily order flow and returns for individual stocks and document a strong positive contemporaneous association. The authors also show that the returns on size-stratified portfolios of stocks display differential sensitivities to order flow.

to the commonality in order flows. Harford and Kaul (2005) find that common effects in order flows and returns are pervasive, though stronger for index constituents than non-index firms, and that marketwide and correlated stock-specific order flows explain a large fraction of the correlation in returns.

Do trades display additional comovement within size-BM groups? One possible reason why they might is comovement in firm fundamentals. A series of papers that includes Berk, Green, and Naik (1999) and Cooper (2006) suggests that size and BM capture differences in firm fundamentals.[5] If investors recognize that the performance of small (high BM) firms will differ from that of big (low BM) firms as economic conditions change, their trades could display size-based or BM-based commonality. An alternative explanation for trade comovement is based on behavioral arguments. Barberis and Shleifer (2003) present a model where investors place assets in different style categories and allocate funds at the style level. They show that returns for assets of the same style comove too much, while returns for assets of different styles comove too little. Further, assets that switch styles will find their returns comoving more strongly with returns for members of the style category they join and less strongly with returns for members of the category they exit. Supporting this model, Barberis, Shleifer, and Wurgler (2005) find that the returns for stocks added to the S&P 500 index comove more strongly with S&P returns and less strongly with non-S&P returns.[6] As small-cap/large-cap and value/growth are important styles, this explanation predicts additional comovement associated with a stock's size and BM characteristics.

[5] Berk et al. argue that size captures the relative importance of a firm's assets-in-place versus its growth options: growth options are the less important for the value of large firms. Cooper suggests that BM is related to the deviation of a firm's capital stock from its target level. The excess installed capacity of a high BM firm makes its performance more sensitive to economic conditions than that of a low BM firm.

[6] Wahal and Yavuz (2009) provide supporting evidence. They find that future returns for high comovement winner (loser) portfolios are significantly higher (lower) than those of low comovement winner (loser) portfolios and see significantly larger reversals than those for low comovement winner (loser) portfolios.

2.2.2 The Origins of SMB and HML

Fama and French (1996) argue that *SMB* and *HML* represent premiums that compensate investors for bearing risks associated with intertemporal consumption-investment decisions. While these risks are difficult to identify ex-ante, Fama and French (1995), in support of this argument, document the presence of size and value factors in earnings growth. Liew and Vassalou (2000) show that *SMB* and *HML* predict GDP growth in several countries. In contrast, Lakonishok, Shleifer, and Vishny (1994) argue that these premiums arise because investors irrationally extrapolate earnings growth too far into the future. This causes growth stocks to be overpriced and value stocks to be underpriced. Consequently, low BM firms have low future returns and high BM firms have high future returns. This behavioral argument for *HML* can be extended to *SMB*. Given two firms with different market capitalizations, it can be argued that the larger firm has an inflated stock price, due to irrational performance extrapolation, and will provide lower future returns; the opposite holds for small stocks. Consistent with this view, there is evidence suggesting that small firms tend to have performed poorly in the past (e.g. Chan and Chen, 1991).

We evaluate these two explanations by studying the relation between order flow for the SMB and HML portfolios and two sets of variables. Under the rational asset pricing view, *OFSMB* and *OFHML* should vary with factors that measure risks against which investors might want to hedge, including business cycle fluctuations, aggregate uncertainty, and marketwide liquidity. To capture these rational effects, our first set of variables includes two proxies for expected future conditions, the default spread (*DEF*) and the term spread (*TERM*); marketwide uncertainty implied by the prices of S&P 100 index options (*VXO*); and aggregate liquidity (*LIQ*) as measured by Pastor and Stambaugh (2003). The behavioral argument is that low BM stocks and big stocks are over-bought by investors who extrapolate recent growth too far into the future. This argument suggests that *OFSMB* and *OFHML* should relate to current performance and investor sentiment. Accordingly, our second set of variables consists of the National Association of Purchasing Managers Index (*NAPM*, a coincident indicator) and the Baker and Wurgler (2006) investor sentiment measure (*SENT*, capturing the irrational component of investor sentiment).[7]

DEF and *TERM* are well-established as forward-looking economic indicators. Fama and French (1989) argue that *TERM* tracks short-term fluctuations while *DEF* captures longer-term episodes. Chen (1991) shows that a high *DEF* predicts distress and likely poor economic conditions while a high *TERM* signals rapid growth in future periods. *VXO* is the implied volatility on near-term equity index (S&P 100) options, and is expected to be higher when aggregate uncertainty is high. *LIQ* measures the average strength of volume-related return reversals and is generally negative, the more so when liquidity is strained. *NAPM*, based on a survey of purchasing managers, is high when the economy is doing well and sales performance is strong.[8] The sentiment index, *SENT*, is based on six measures of investor sentiment, and is high when investors are excessively bullish and low when they are excessively bearish.

Extant theory, exemplified by Berk, Green, and Naik (1999) and Cooper (2006), suggests that the performance of small firms and high BM (value) firms is more sensitive to economic conditions than that of big firms and low BM (growth) firms. The less procyclical performance of big firms and growth firms should make them good hedges against business cycle fluctuations. Greater uncertainty

[7] The construction of these variables is described in Section 3.3. Our approach is similar to Hvidkjaer (2006), who uses intraday transaction data to analyze how investors trade momentum portfolios.

[8] NAPM announcements are closely followed. Balduzzi, Elton, and Green (2001) list NAPM among the eight major U.S. announcements that affect the prices of all four government bonds they study. Moreover, *NAPM* is highly correlated with contemporaneous growth in the economy. Harris (1991) shows that *NAPM* explains 71% of the variation in real GNP, 58% of the variation in industrial production, and 44% of the variation in capacity utilization. He finds, by contrast, that *NAPM* does not reliably predict business cycle turning points, which makes it an ideal proxy for current conditions. Another reason why *NAPM* is relevant for our purposes is that most of its components are directly related to aggregate sales. If sales growth is the basis for irrational extrapolation, including *NAPM* makes for a meaningful test of the behavioral hypothesis.

increases the value of growth options and thus makes low BM stocks more attractive. Moreover, since aggregate volatility tends to vary opposite to business conditions (e.g. Schwert, 1989), it is plausible that changes in aggregate uncertainty would affect big stocks less than small stocks and growth stocks less than value stocks. Last, there is comovement in liquidity (Chordia, Roll and Subrahmanyam, 2000), and changes in aggregate liquidity are unlikely to affect stocks with different liquidity levels or liquidity risks in the same way. In particular, big and growth stocks tend to be more liquid and to have less liquidity risk than small and value stocks (e.g. Acharya and Pedersen, 2005). Thus, a decline in liquidity may be expected to make large stocks and growth stocks attractive relative to small stocks and value stocks. If investors regard *DEF*, *TERM*, *VXO*, and *LIQ* as important state variables, their trades should be driven by these factors. Based on the relative sensitivities of SMB and HML to economic conditions (discussed above), we would then expect *OFSMB* and *OFHML* to be related positively to *TERM* and negatively to *DEF*, *VXO*, and *LIQ*.

By contrast, the behavioral story predicts no relation between *OFSMB* or *OFHML* and the above state variables. Instead, the story holds that in periods of strong growth and positive sentiment investors tend to buy growth stocks (and, arguably, big stocks) and sell value stocks (and small stocks) in excessive quantities. We hypothesize that the effects of such biases on *OFSMB* and *OFHML* should be magnified when recent economic performance is stronger and when investors are excessively bullish about the economy. The behavioral explanation, therefore, suggests that *OFSMB* and, especially, *OFHML* will be negatively related to *NAPM* and *SENT*.

In response to the question of what we learn from order flow that is not apparent from previous studies of returns, we have a few responses. Fama and French (1989) do not find strong business cycle effects on monthly stock returns; neither do we over our sample period. However, such effects are clearly visible in order flows. We believe that order flow is a cleaner variable to study because a change in fundamentals may have two confounding effects, one related to price adjustment and the other related to expected returns. For instance, due to the countercyclical nature of expected returns, a favorable signal in month *t* about economic conditions (e.g. an increase in *TERM*) is likely to raise prices and reduce expected returns. Thus, if prices rapidly impound the signal, returns in month *t*+1 should be lower. However, if this signal is only partially incorporated into prices at the time it is observed, subsequent adjustment to the signal will lead to high returns in month *t*+1. This will offset (or dominate) the expected return effect.[9] Month *t*+1 order flow, by contrast, will be positive if investors react with a lag and random otherwise. Second, quantifying the contribution of order flow comovement to return comovement is important, as the state variable argument does not necessarily imbue trading with a role.

2.3 Data and Variables

2.3.1 Monthly Order Flows

We obtain intraday trade and quote data from the Institute for the Study of Security Markets (ISSM) dataset for the period 1988 through 1992 and from the Trade and Quote (TAQ) dataset for the period 1993 through 2004.[10] Following Chordia et al. (2002), we restrict the sample to ordinary shares trading on the NYSE in order to ensure that our results are not influenced by differences in trading protocols across venues. After cleaning the intraday data with the filters specified in Chordia and Subrahmanyam (2004), trades are signed using the Lee-

---

[9] In a study of momentum in firm-level stock and bond returns, Gebhardt et al. (2005) find that both stock and bond values under-react to firm fundamentals, bonds the more so. This documented under-reaction may also have implications for the relation between marketwide returns and macroeconomic fundamentals.

[10] Our sample period is defined by the availability of reliable transaction data. The ISSM data go back as far as 1983; before 1988, however, data are missing for several days for the entire population of stocks and are incomplete for several stocks on other days.

Ready (1991) algorithm.[11] For each stock, order flow is computed monthly as the difference between buyer-initiated and seller-initiated trades in terms of (i) number of transactions (*OFN*), (ii) number of shares traded (*OFV),* and (iii) dollars traded (*OFD*). Following Chordia and Subrahmanyam (2004), we standardize order flow by the corresponding measure of total trading activity (i.e. total number of transactions, *TOTN*; share volume, *TOTV*; and dollar volume, *TOTD*) to obtain scaled order flow (*OFNX*, *OFVX*, and *OFDX*).

Panel A of Table 1 presents the time-series averages of the cross-sectional means and standard deviations of the order flow and trading activity measures. The positive mean of approximately 4% for each measure indicates buying pressure in the market between 1988 and 2004. The unscaled order flow means are also positive: 417 trades, 655,800 shares and 26.8 million dollars. These positive values are unsurprising given that: (i) the order flow measures reflect only market orders; and (ii) the U.S. market has been in expansion for a major part of our sample period.[12] The means of total (i.e. unsigned) trading activity are 5566 transactions, 7.8 million shares and $294.5 million.

Panel B reports the cross-sectional averages of the stock-level time-series correlations among the order flow measures. While the average correlations between the monthly share and dollar volume-based order flows (scaled or

[11] The first trade of each day and trades (a) out of sequence, (b) recorded outside NYSE trading hours, (c) with special settlement conditions are removed. Quotes with negative spreads and quotes for which (a) the ask or the bid price moves by more than 50% and (b) the spread is greater than 20% of the quote midpoint ($2) when the quote midpoint is greater (less) than $10 are eliminated. The Lee-Ready algorithm compares the transaction price with the first quote occurring at least five seconds earlier: the trade is classified as a buy (sell) if the price is closer to the ask (bid). If the trade occurs at the quote midpoint, it is classified as a buy (sell) if the price is higher (lower) than that of the preceding trade. Buys (sells) are assigned a positive (negative) sign.

[12] This contrasts with Chordia and Subrahmanyam (2004) who report negative mean values for scaled daily order flows over 1988-1998. Untabulated analysis for 1988-1998 shows that the mean scaled monthly order flows are smaller, but still positive. Our monthly figures can be reconciled with their daily figures if there are more negative order flow days in any month than positive order flow days and the volume generated on positive order flow days is higher than that generated on negative order flow days. As a check, we compute the mean daily order flow in number of trades and dollars for 1988-1998. Our means of 4.71 trades and $560,000 are similar to the daily means of 4.67 trades and $430,000 reported by Chordia and Subrahmanyam.

unscaled) are above 0.90, those between the transaction-based measures and the share or dollar measures are lower, at around 0.50. The mean correlation between the scaled and unscaled order flows is approximately 0.70; thus, it is unsurprising that results based on scaled versus unscaled order flow are similar. We also report the average correlations between order flow and the contemporaneous monthly stock return (shown in italics along the diagonal). Scaled order flow in number of shares or dollars is more highly correlated with returns (0.32) than is the scaled transaction measure, *OFNX* (0.22). This can be explained by the fact that higher volumes are likely to move prices to a greater extent.

2.3.2 Size and Book-to-Market Portfolios

Monthly order flows are merged with data on prices, returns, and number of shares outstanding obtained from the Center for Research in Security Prices (CRSP) monthly files and with annual balance sheet and income statement data from COMPUSTAT. As in Chordia and Subrahmanyam (2004), stocks whose prices exceed $999 in any given year are eliminated from the sample for that year. We use the size and BM breakpoints available at Kenneth French's website to assign stocks to two size (small and big: S and B) and three BM (high, medium, and low: H, M, and L) categories. This categorization results in six portfolios: S/L, S/M, S/H, B/L, B/M, and B/H. In June of each year *t*, stocks are assigned to one of these portfolios based on their market capitalizations at the end of June and BM ratios as of the preceding December-end. For each portfolio, equally-weighted monthly order flow is computed from July of year *t* to June of year *t+1*, at which point the portfolios are reformed.[13]

[13] On average, there are 974 stocks per year, distributed unevenly across the six portfolios: the portfolio B/H is the smallest, with an average of 74 stocks, while B/L is the largest, with an average of 245 stocks. The number of stocks increases steadily from 680 in January 1988 to 1343 in December 2004. Our results are robust to the use of value-weighted instead of equally-weighted order flows. This is not surprising given that the correlation between the two measures is in excess of 0.85 for each portfolio.

Fama and French (1993) show that the time-series of stock returns is explained by a three-factor model that includes the excess return on a broad market portfolio over the risk-free rate (*MKT*) and the returns on a portfolio long in small stocks and short in big stocks (*SMB*) and a portfolio long in high BM stocks and short in low BM stocks (*HML*). We compute order flow for these three portfolios as follows. Market order flow (*OFMKT*) is the cross-sectional average of the monthly order flows for the sample stocks. SMB order flow *(OFSMB)* is the difference between the simple averages of the monthly order flows for the three small stock portfolios (S/L, S/M, and S/H) and the three big stock portfolios (B/L, B/M, and B/H). Similarly, HML order flow *(OFHML*) is the difference between the simple averages of the monthly order flows for the two value portfolios (S/H and B/H) and the two growth portfolios (S/L and B/L). *MKT*, *SMB* and *HML* are the equally-weighted return factors from Kenneth French's website.

The top six rows of Table 2 provide the time-series means and standard deviations for the scaled share-volume based order flow factors and the return factors, as well as the correlations among these variables. The mean value for *OFMKT* is 3.64%, consistent with the bull market run over much of our sample period. However, the mean values for *OFSMB* and *OFHML* are -5.51% and -0.84%, so order initiators are, on average, selling these portfolios. Looking at the return factors, the average monthly excess market return, *MKT*, is 0.84% over the sample period (10.55% annually), while the average size and value premiums, *SMB* and *HML*, are 0.26% and 0.62% (3.17% and 7.70% annually). The monthly standard deviations for the order flow factors—3.99% for OFMKT, 2.79% for *OFSMB* and 2.54% for *OFHML*—indicate considerable time-series variation in each series. The standard deviations for the return factors are also large.

Turning to the contemporaneous correlations among the order flow and return factors, we first note that each order flow factor is, as would be expected, correlated positively and significantly with its return counterpart. *OFMKT* is correlated negatively with *OFHML* and positively with *OFSMB* and *SMB*, with correlation coefficients of -0.28, 0.57, and 0.18. These correlations indicate that small stock returns tend to be high compared to big stock returns and the buying pressure for low BM (small) stocks tends to be high relative to that for high BM (big) stocks when marketwide buying pressure is high. In line with this, (i) *MKT* is correlated positively with *SMB* and negatively with *HML* with correlation coefficients of 0.17 and -0.49, and (ii) *SMB* is correlated negatively with *HML*, with a correlation coefficient of -0.44.

2.3.3 Measuring Business Conditions

As described in Section 2.2, we employ the default and term spreads (*DEF* and *TERM*) to capture expectations about future economic conditions, implied volatility (*VXO*) to capture aggregate uncertainty, and the Pastor-Stambaugh (2003) liquidity measure to capture marketwide liquidity (*LIQ*). The data used to construct *DEF* and *TERM* come from the St. Louis FED (research.stlouisfed.org). *DEF* is computed as the difference between the annualized yields on Moody's seasoned Baa-grade and Aaa-grade bond portfolios, while *TERM* is the difference between the 10-year Treasury constant maturity rate and the annualized one-month Treasury bill rate. *VXO*, which measures the 30-day volatility implied by S&P 100 index option prices, comes from the Chicago Board of Options Exchange website (www.cboe.com). *LIQ* is available from the website of Lubos Pastor, and is based on the cross-sectional average slope coefficient from a stock-level regression of next-day returns on volume interacted with the sign of the contemporaneous excess return.[14]

To capture aggregate sales performance, we add the National Association of Purchasing Managers Index (*NAPM*), available from the St. Louis FED. *NAPM* is based on the survey responses of 300 purchasing managers, and is above (below)

---

[14] Pastor and Stambaugh (2003) adjust this measure to ensure stationarity as well as remove predictability.

50 for expansion (contraction) relative to the previous month.[15] To measure investor sentiment (*SENT*), we use the Baker-Wurgler (2006) index, available from Jeff Wurgler's website. *SENT* is the first principal component of six variables: the closed-end fund discount, NYSE share turnover, the number of IPOs and the average of their first-day returns, the equity share in new issues, and the dividend premium. Each series is orthogonalized with respect to the growth rates of industrial production, consumer durables, non-durables, services, and an NBER recession dummy in order to isolate the component not related to macroeconomic conditions. Denoting the month of the order flow and return observations as month *t*, *DEF*, *TERM*, and *VXO* are measured at the end of the month *t*-1, *NAPM* is from the survey in month *t*-1, while *LIQ* and *SENT* are calculated over month *t*-1.

Descriptive statistics and the correlation matrix for these variables are presented in the bottom six rows of Table 2. The means for *DEF* and *TERM* are 0.85% and 3.00% over the sample period, while the average values of *VXO* and *LIQ* are 21.07% and -2.11 (*LIQ* is multiplied by 100 in order to make its scale comparable to those of the other variables). *NAPM* has a mean of 51.93, indicating that the economy has performed well over much of our sample period. By construction, the mean value of *SENT* is close to 0. The large standard deviations for each variable point to high power in our tests, i.e. for a given slope, higher variability translates into greater precision of the estimate.

Focusing on the significant correlations among these six variables, we see that (i) *DEF* is correlated positively with *TERM* and *VXO* and negatively with *LIQ* and *NAPM*, (ii) *TERM* is correlated positively with *NAPM*, and (iii) *VXO* is correlated negatively with *LIQ* and *NAPM*. *NAPM* is an indicator of current economic conditions and should be high when sales performance is strong. High values of

[15] The survey covers 20 manufacturing industries and the following components: new orders, production, supplier delivery times, backlogs, inventories, prices, employment, export orders, and import orders. These components, with the possible exception of delivery times, are all closely tied to economy-wide sales performance. While services are excluded, *NAPM* is strongly related to the performance of the economy.

DEF indicate below-average economic conditions and high values of *TERM* indicate above-average conditions in future periods. Thus, while the positive correlations between *DEF* and *TERM* and *NAPM* and *TERM* are slight puzzles, the negative correlation between *NAPM* and *DEF* is reasonable. Likewise, volatility tends to be high during slowdowns (Schwert, 1989), so it makes sense that *VXO* is related negatively to *NAPM* and positively to *DEF*. Last, *SENT* is not related significantly to *DEF*, *TERM* or *VXO* because it is orthogonalized using other variables that capture business conditions.

Turning to the correlations between the explanatory variables and the order flow and return factors, we see that *OFMKT* tends to be high when *TERM* or *NAPM* is high. *OFSMB* is correlated positively with *TERM* and *NAPM* and negatively with *DEF* and *VXO*. The correlation between *OFHML* and *VXO* is also negative. These correlations point to a strong procyclical pattern in *OFSMB* and a milder procyclical pattern in *OFHML*. For instance, both order flow factors tend to be low when expected volatility is high. The correlations with the return factors tend to be weaker. Here, the only significant coefficients are the positive correlations between *TERM* and both *MKT* and *SMB.* The fact that *SENT* is not significantly correlated with any of the order flow or return factors provides preliminary evidence against a behavioral interpretation of *SMB* or *HML*.

2.4 Size/BM-Driven Order Flow Comovement and Its Effects on Prices

2.4.1 The Strength of Comovement

Huberman and Kandel (1987) and Chan and Chen (1991) show that there is covariation in returns related to size and relative distress and not captured by the market return. While the presence of size- and BM-based comovement in returns suggests that similar effects should exist in order flows, rapid price adjustment will obviate or attenuate such comovement. Hence, we start by investigating the presence of size and BM influences on order flow. Note that the results in this section provide background for our subsequent analysis both of the importance of order flow comovement in generating return comovement (see Section 4.2) and of the determinants of the size and BM factors in order flows (see Section 5).

For consistency with this subsequent analysis, we measure order flows monthly. We carry out two tests. In our first test, we investigate whether the order flow of the size-BM portfolio to which a stock belongs has incremental explanatory power for that stock's order flow, beyond marketwide and own-stock effects. We run stage-wise regressions of the order flow for each stock on its first lag and the contemporaneous market and portfolio order flows. The latter is the equally-weighted order flow for the size-BM portfolio to which the stock belongs in that year, adjusted to exclude the stock's order flow. We use all the monthly observations available for each stock in this analysis.

The cross-sectional means and medians of the adjusted R^2 statistics from these regressions are presented in Panel A of Table 3. First, we observe that lagged order flow explains about one-fifth of the variation in the current period's order flow for individual stocks. This strong autocorrelation at a monthly frequency suggests that market-makers let their inventories deviate from an unconditional mean for extended periods, consistent with the slow pace of inventory adjustment reported in Hasbrouck and Sofianos (1993) and Madhavan and Smidt (1993).[16] For every order flow measure, there is a statistically significant jump in the mean and median R^2 when portfolio order flow is included in the model, irrespective of the order in which the regressors are added. For the number of trades based measure (*OFNX*), the mean explained variation rises from 19% to 26% when portfolio order flow is added to a model with lagged own order flow, and from 26% to 29% when it is added to a model with lagged own order flow and contemporaneous market order flow. For volume-based order flow, *OFVX*, the

[16] By contrast, Comerton-Forde, Hendershott, Seasholes, and Moulton (2009) show a rapid response of specialist liquidity provision to inventory imbalances.

average R^2 roughly doubles when portfolio order flow is added to a model with lagged own order flow.[17]

Our second test exploits the fact that, over the 1988-2004 sample period, 1607 stocks move between size/BM categories. This allows for an event-study test of the importance of these characteristics in driving comovement. If a stock's size/BM characteristics are critical to comovement, its order flow should comove strongly with the order flow for the portfolio to which it belongs both before and after the switch. Thus, we should see a drop in the comovement of the switching stock's order flow with the order flow for the portfolio it moves from and an increase in comovement with the order flow for the portfolio it switches to.

To assure ourselves of a reasonable number of time-series observations for this test, we require the switching stocks to stay in the pre-switch and post-switch portfolios for at least two years on either side of the move. We identify 318 such stocks and estimate two regressions, one before and the other after the switch. The dependent variable is monthly order flow (scaled number of trades, *OFNX*) for the switching stock, and the two independent variables are the original and new portfolio order flows, which are adjusted to exclude the switching stock's order flow. Panel B of Table 3 presents the results. The average slope coefficient on the original portfolio's order flow is 0.85 in the pre-switch period and drops to 0.20 in the post-switch period, while that on the new portfolio's order flow is 0.24 in the pre-switch period and increases to 0.84 in the post-switch period. Both changes are highly significant and large in economic terms.

The results in Table 3 point to correlated investor trading in stocks with similar size and BM characteristics. This evidence of size/BM comovement in stock order flows is consistent with Barberis and Shleifer (2003), who suggest that allocations occur at the level of broad categories, as well as with rational

[17] The results for *OFDX*, which are almost identical to those for *OFVX*, are suppressed to save space.

comovement due to common size and BM effects in cash flows, reported in Fama and French (1995). The analysis in Section 5 leads us to favor a rational rather than behavioral explanation for this comovement. Our results are also consistent with the finding in Hasbrouck and Seppi (2001) that the variation in fifteen-minute order flows is well-explained by three principal components. Hasbrouck and Seppi go on to show that the principal components for order flows are related to those for returns. In the same spirit, we assess the importance of order flow comovement in generating return comovement.

2.4.2 Contribution of Order Flow to Size and BM Effects in Returns

This subsection examines the contribution of correlated size- and BM-driven trading to the empirical importance of the Fama-French (1993) three-factor model. We proceed by comparing the explanatory power of the three-factor model for the returns on six size-BM sorted portfolios before and after we adjust the portfolio returns for trading-induced effects. The greater the importance of correlated trading, the larger will be the decline in the explanatory power of the three factors after we adjust the portfolio returns for trading. Our approach follows that in Hasbrouck and Seppi (2001) and in Harford and Kaul (2005), who provide a general analysis of the relation between correlated trading and correlated returns using intraday data.

Returns are adjusted for trading in the following three steps. In an efficient market, any relevant information should be inferred from the unanticipated component of order flow. Therefore, we first obtain the unanticipated order flow for each of the six size-BM portfolios as the residual from an AR(3) model for portfolio order flow.[18] In the second step, we address the potential endogeneity of

[18] Box-Ljung tests reveal this to be the appropriate time-series model. As a robustness check, we repeat the analysis with the trade innovation defined as the residual from a regression of order flow on lagged order flow as well as lagged values of the market/business condition variables. Our conclusions are unchanged.

unexpected order flow relative to returns, i.e. the possibility that instead of returns being driven by monthly order flow, it is the other way around. We deal with this problem by reconstructing the order flow factors —$OFMKT^*$, $OFSMB^*$, and $OFHML^*$—separately for each of the six portfolios. These factors are based on unexpected order flow for the six portfolios but crucially exclude the order flow for the portfolio whose return is being adjusted for trading.[19] Thus, for instance, $OFSMB^*$ corresponding to the portfolio S/H is defined as the difference between the average of the unexpected order flows for the remaining small stock portfolios (S/L and S/M) and the average of the unexpected order flows for the big stock portfolios (B/L, B/M, and B/H). We similarly exclude the order flow for S/H while averaging portfolio order flows to get $OFHML^*$ and $OFMKT^*$ for this portfolio. Finally, we regress the excess monthly return for each portfolio on $OFMKT^*$, $OFSMB^*$, and $OFHML^*$, and call the residual from this regression the *trade-adjusted return*.[20]

We regress the raw and trade-adjusted excess returns for the six portfolios on *MKT*, *SMB* and *HML*, and compare the slope coefficients and goodness-of-fit statistics. Panels A and B of Table 4 present the results from these regressions. Before the adjustment for trades, the three return factors explain between 85% (for B/H) and 94% (for B/M) of the variation in the excess returns for the six portfolios. After the portfolio returns are adjusted for trades, the explanatory power of the three factors drops by one-third or more, to between 38% (for B/L) and 54% (for S/H). Additionally, we observe significant declines in many of the slope coefficients. The slopes on *MKT* are halved for all portfolios regardless of size and BM characteristics. The slopes on *SMB* are reduced by about 60% for

[19] The results obtained using unadjusted order flows are qualitatively similar, but more significant.

[20] The Fama-French portfolios used in this analysis also include NASDAQ and AMEX stocks. We use NYSE-based order flows to explain broader portfolio returns in order to reduce mechanical effects—to the extent that our order flows explain the returns to the Fama-French portfolios, the results should be even stronger for portfolio returns based on NYSE stocks alone. Indeed, when we repeat the analysis using NYSE size/BM portfolio returns, the explanatory power of order flow increases.

small stocks, but are not affected (or even increase slightly in magnitude) for big stocks. Finally, the slopes on *HML* drop significantly for high BM stocks, with the decline being more pronounced for S/H (from 0.56 to 0.17) than B/H (from 0.63 to 0.43); the slopes for medium and low BM stocks are unaffected.

Collectively, the large reductions in the model R^2 and many slope coefficients point to the importance of correlated trading in driving common effects in returns. Several questions emerge. First and foremost, are investor trades in the SMB and HML portfolios consistent with a rational hedge portfolio interpretation of these factors, or is a behavioral interpretation more appropriate? More generally, what are the influences on these common factors in order flow? Do the common factors display patterns related to economic conditions? The next section addresses these questions.

2.5 On the Origins of the Size and Value Factors

2.5.1 Analysis of Portfolio Order Flows

One way to distinguish between the rational and behavioral explanations for *SMB* and *HML* is to investigate how investors trade these portfolios. The rational asset pricing explanation suggests that the average returns for big firms and low BM firms are lower than implied by their CAPM betas because the performance of these firms provides a hedge against adverse shifts in consumption/investment opportunities. Small firms and high BM firms have higher-than-expected average returns because their performance provides less of a hedge. To evaluate this explanation, we relate *OFSMB* and *OFHML* to proxies for the state of the economy: aggregate uncertainty (*VXO*), marketwide liquidity (*LIQ*), and two measures of expected economic conditions (*DEF* and *TERM*). Based on the arguments in Section 2.2, we expect *OFSMB* and *OFHML* to be related positively to *TERM* and negatively to *VXO*, *DEF*, and *LIQ*.

The behavioral explanation predicts no relation between investor preferences for big versus small firms and value versus growth firms and these four variables. Rather, the argument is that the demand for growth stocks (and, arguably, big stocks) is excessively high because these stocks have performed relatively well recently and the demand for value (small) stocks is excessively low because these stocks have performed relatively poorly (see Chan and Chen, 1991, for evidence on the relative performance of small and big firms). We hypothesize that such biases should be more pronounced when current economic conditions and sales are strong and when investor sentiment is positive. Thus, the buying pressure for growth stocks and big stocks will increase relative to that for value stocks and small stocks (causing *OFSMB* and *OFHML* to decline) following periods when *NAPM* and *SENT* are high.

To test these predictions, we regress the current month's scaled share volume based order flows for the SMB and HML portfolios (i.e. *OFSMB* and *OFHML*) on *DEF*, *TERM*, *VXO*, *LIQ*, *NAPM*, and *SENT* as of the end of the preceding month, controlling for a deterministic time trend (*TIME*) and the one-month-lagged values of the excess market return, own portfolio return, and own portfolio order flow (*MKT.L*, *RETP.L*, *OFP.L*).[21] As a benchmark case, we also examine the relation between *OFMKT* and these variables. Table 5 presents the coefficient estimates and Newey-West (1987) adjusted t-statistics from these regressions, as well as F-statistics for the null hypotheses that the coefficients on (a) *DEF* and *TERM*, (b) *VXO* and *LIQ*, or (c) *NAPM* and *SENT* are jointly zero in a model comprising the remaining variables.

First, note that *OFMKT*, *OFSMB* and *OFHML* are all related significantly to the variables that capture expectations about future economic conditions: the hypothesis that the coefficients on *DEF* and *TERM* are jointly zero is rejected (the

[21] The lagged market and portfolio returns are included to pick up any contrarian or momentum-based relation between returns and subsequent order flows. The lagged portfolio order flow accounts for persistence in order flow. Since order flow measured in terms of number of trades and dollar volume yields similar conclusions, for brevity, we only report the results for share volume order flow.

p-values from the F-tests are <0.01 for *OFSMB* and <0.05 for *OFMKT* and *OFHML*). *VXO* and *LIQ* and *NAPM* and *SENT*, however, are significant only for *OFHML* and at the 10% level (with negative coefficients). The contrasting significance of the sets of F-tests provides support for the rational asset pricing view of *SMB* and *HML*, but not for the behavioral view. The coefficients on the proxies for expected economic conditions, *DEF* and *TERM*, are also consistent with the rational, state variable view of *SMB* and *HML*. Both *OFSMB* and *OFHML* are associated positively with *TERM* and negatively with *DEF*. Thus, expected deterioration in economic conditions leads to significant declines in the buying pressure for small stocks relative to big stocks and for value stocks relative to growth stocks. In economic terms, a one standard deviation increase in *DEF* leads to a decline of 0.5% (0.17 standard deviations) in *OFSMB* and of 0.3% (0.14 standard deviations) in *OFHML*, while a one standard deviation increase in *TERM* predicts an increase of 0.8% (0.30 standard deviations) in *OFSMB* and 0.4% (0.18 standard deviations) in *OFHML*.

We also observe that *OFMKT* is related positively to *TERM*, so that buying pressure for all stocks increases (drops) if the economy is expected to grow more (less) rapidly. Such behavior is consistent with investors shifting funds away from the stock market, possibly towards safer assets such as bonds, when economic conditions are expected to deteriorate (inter-market flight-to-quality). The results for *OFSMB* and *OFHML* are consistent with investors moving funds from small stocks to big stocks and from value stocks to growth stocks when they expect conditions to deteriorate (intra-market flight-to-quality). In unreported analysis, we study tail effects. We do this by introducing separate interaction effects for the cases where *DEF*, *VXO*, and *SENT* are in their top quartiles or deciles, and *NAPM*, *TERM* and *LIQ* their lowest quartiles or deciles. However, none of these tail effects is significant. Thus, instead of a discrete shift driven by extreme conditions—in particular, given their prior significance, high values of *DEF* and low values of *TERM*—it seems that investors continuously adjust their portfolios in response to signals about the state of the economy.

For further insights, we subdivide *OFSMB* by BM, into $OFSMB_L$, $OFSMB_M$ and $OFSMB_H$, and *OFHML* by size, into $OFHML_B$ and $OFHML_S$.[22] This decomposition sharpens our results in three related ways. First, it allows us to see where the effects just discussed arise. For instance, are the effects of *DEF* and *TERM* on *OFSMB* more pronounced among value or growth stocks? Second, this analysis sheds additional light on the importance of flight-to-quality. Specifically, the performance of small value stocks should be more sensitive to variations in economic conditions than that of large value stocks, since size plausibly provides a buffer against downturns. By contrast, the performance differential between small growth stocks and big growth stocks should not depend as much on economic conditions, since growth stocks are likely to be more financially secure. If such considerations are important to investors, $OFHML_S$ will be more sensitive to changes in economic conditions than is $OFHML_B$. The same logic applies to the *OFSMB* decomposition: if flight-to-quality effects are important, $OFSMB_L$ will be less strongly related to proxies for economic conditions than is $OFSMB_H$. A third advantage is that we can check if the effects of sentiment are masked overall but present in stocks especially prone to sentiment, e.g. small stocks.

The bottom panel of Table 5 reports the results for these finer portfolios. Starting with *OFSMB*, first note that the coefficients on *NAPM* and *SENT* are never significant, jointly or individually. Similarly, *VXO* and *LIQ* are never jointly significant. Second, the null hypothesis that the coefficients on *DEF* and *TERM* are zero is rejected for all three BM categories. Third, consistent with a financial distress argument for high BM firms (Fama and French, 1995), the negative relation between *OFSMB* and *DEF* is stronger for high BM stocks than for low BM stocks. The effect of a one standard deviation increase in *DEF* is a decline of 0.1% in $OFSMB_L$, 0.5% in $OFSMB_M$, and 0.9% in $OFSMB_H$ (0.03, 0.18, and 0.25

---

[22] Thus, $OFHML_S$ is the difference between the order flow for the small-high BM portfolio and the order flow for the small-low BM portfolio and $OFHML_B$ is the difference between the order flows for the big-high BM portfolio and the big-low BM portfolio. Similarly, $OFSMB_L$ is the difference between the order flow for the small-low BM portfolio and the order flow for the big-low BM portfolio, while $OFSMB_H$ is the difference between the order flows for small and big stocks in the high BM category.

standard deviations, respectively). Fourth, the positive relation between *TERM* and *OFSMB* is strongest for high BM stocks: a one standard deviation increase in *TERM* leads to an increase of 1.1% (0.33 standard deviations) in $OFSMB_H$ versus 0.8% (0.23 standard deviations) in $OFSMB_L$ and $OFSMB_M$.

For *OFHML*, the null hypothesis that the coefficients on *DEF* and *TERM* are jointly insignificant is rejected for $OFHML_S$, but not for $OFHML_B$. The slope estimates for $OFHML_S$ also line up well with the flight-to-quality story. $OFHML_S$ responds negatively to an increase in *DEF*, with a one standard deviation increase in *DEF* corresponding to a decline of about 0.8% (0.23 standard deviations) in $OFHML_S$. This negative relation is consistent with a flight of investor funds out of (likely distressed) small high BM stocks towards safer alternatives in response to increased default risk. The relation between $OFHML_S$ and *TERM* is positive, with a one standard deviation increase in *TERM* leading to a 0.7% (0.19 standard deviations) increase in $OFHML_S$. Thus, buying pressure for small value stocks increases relative to small growth stocks in anticipation of stronger economic conditions. Although the hypotheses that the coefficients on *VXO* and *LIQ* and *NAPM* and *SENT* are jointly zero are not rejected for either subcategory, the coefficient on *NAPM* is significantly below zero (at the 10% level) in the model for $OFHML_S$. This negative coefficient is consistent with the prediction of the behavioral story that investors overinvest in small high growth stocks relative to small lower growth stocks when the economy is strong.

The coefficients on the control variables are also of interest. We observe a notable amount of persistence in portfolio order flows. The coefficient on lagged order flow is 0.46 for *OFMKT*, 0.34 for *OFSMB*, and 0.17 for *OFHML*. Persistence is highest for medium BM stocks among the *OFSMB* subcategories, and for big stocks among the *OFHML* subcategories. *OFSMB* and *OFHML* are positively related to the lagged market return. The effect of the lagged market return on *OFSMB* is increasing in BM and its effect on *OFHML* is larger for small stocks than for big stocks. In the presence of the lagged market return, the lagged portfolio return is never significant. Last, there is a negative time trend in *OFHML* and a positive trend in *OFMKT* and *OFSMB*. The trend in *OFSMB* appears to be driven by stocks in the low BM category.

2.5.2 Analysis of Portfolio Returns

Given our result that the common effects in order flows are related to the common effects in returns (see Table 4), we conduct a similar analysis for *MKT*, *SMB* and *HML*, and the respective size and BM subcategories. Table 6 reports the results. First, as with order flows, the hypothesis that the coefficients on *NAPM* and *SENT* are jointly zero is never rejected. The hypothesis that the coefficients on *DEF* and *TERM* are both zero is rejected only for *SMB*, while the hypothesis that *VXO* and *LIQ* are jointly zero is rejected only for *HML*.

Looking at the individual coefficients, there is a positive relation between *SMB* and *TERM* and a negative relation between *HML* and *VXO*. The subcategory results for *SMB* reveal that the positive effect of *TERM* is more pronounced for low and high BM stocks than for medium BM stocks. The negative effect of *VXO* on *HML* is stronger for returns than order flows, suggesting that the adverse effect of increased uncertainty is incorporated into the prices of distressed high BM firms without much need for trading. Last, the excess market return is positively associated with *TERM*. *SENT* and *NAPM* are never individually significant. The control variables are largely insignificant with the exception of the lagged excess market return (*MKT.L*), which is positively associated with *MKT*, *SMB*, and *HML* in the following month. In general, however, the results in Table 6 are weaker than those in Table 5, suggesting that the effects of business cycle fluctuations are less visible in returns than in order flows. As mentioned in Section 2.2, confounding return effects associated with a given macro-economic signal would weaken its predictive power for returns.

Our analysis suggests that a negative assessment of future economic conditions reduces buying pressure for stocks as a whole, possibly boosting safer investments such as bonds. Further, the decline in buying pressure is larger for stocks that are perceived to be riskier: small stocks relative to big stocks and value stocks relative to growth stocks. These effects are weaker in returns, suggesting that order flow is the better object of analysis. Finally, neither recent economic performance nor investor sentiment has much explanatory power for investor trading in all stocks or in the SMB and HML portfolios. Taken together, our results contribute to the debate over the source of the size and value premiums in returns. The fact that investors are attracted to big stocks and growth stocks when economic conditions are expected to be unfavorable is consistent with these stocks being viewed as hedges and therefore commanding unconditionally low expected returns. The opposite holds for small stocks and value stocks. Thus, our analysis of investor trading supports a rational explanation for the existence of the size and value premiums.

2.5.3 Does Order Flow Over-React?

Section 5.1 provides evidence that investor trades in the SMB and HML portfolios are related to forward-looking economic indicators in a manner consistent with a rational interpretation of *SMB* and *HML*. These results do not rule out the possibility of investor over-reaction to changes in these indicators. For instance, investors may buy (sell) SMB and HML in excessive quantities in response to an increase (decrease) in *TERM*, resulting in unduly large and significant slope coefficients on this variable in our regressions and strengthening our results. We address this possibility in this section.

Over-reaction to any variable imparts an upward bias to the slope coefficient on that variable (under-reaction imparts a downward bias). Consider the basic specification we estimate in the Section 5.1:

$$OF_{p,t} = \alpha_p + X_{t-1}\beta_p + Z_{p,t-1}\kappa_P + \upsilon_{p,t} \qquad (1),$$

where $OF_{p,t}$ is the month t order flow for portfolio p (p = *SMB*, *HML* or *MKT*), X_{t-1} is a row vector of the month t-1 realizations of the six key explanatory variables, $Z_{p,t-1}$ is a vector of control variables (some of which are portfolio-specific), and $\upsilon_{p,t}$ is white noise. Here, the coefficients of interest are represented by the column vector, β_P, which can be written as the sum of the true coefficient vector, β_P^*, and an error that captures over-reaction or under-reaction: $\beta_P = \beta_P^* + (\beta_P - \beta_P^*)$.[23] Using this identity, (1) can be rewritten as:

$$OF_{p,t} = \alpha_p + X_{t-1}\beta_p^* + X_{t-1}(\beta_p - \beta_p^*) + Z_{p,t-1}\kappa_p + \upsilon_{p,t} \qquad (2).$$

Here, $X_{t-1}(\beta_p - \beta_p^*)$ represents the error in the response of order flow to the six variables. Now suppose that order flow in month t over-reacts to one or more variables in X_{t-1} and this over-reaction is corrected in month t+1. We make this assumption for tractability. While we could assume that over-reaction is corrected slowly over several months, estimation becomes more complicated (requiring non-linear estimation strategies), and the number of coefficients increases significantly. Assuming that over-reaction is corrected in one month, order flow in month t+1 can be written as:

$$OF_{p,t+1} = \alpha_p + X_t\beta_p - X_{t-1}(\beta_p - \beta_p^*) + Z_{p,t}\kappa_p + \upsilon_{p,t+1} \qquad (3).$$

The first term on the RHS of (3) reflects the immediate order flow response to the most recent values of X_t (this includes over-reaction), while the second term represents the correction of the previous period's over-reaction. A rearrangement of (3) yields the specification that we estimate:

$$OF_{p,t+1} = \alpha_p + \Delta X_t\beta_p + X_{t-1}\beta_p^* + Z_{p,t-1}\kappa_p + \upsilon_{p,t} \qquad (4),$$

[23] To keep things manageable, we assume that the coefficients on $Z_{p,t-1}$ are unbiased.

where ΔX_t is the change in the explanatory variables from month t-1 to month t. In this model, therefore, the coefficients on X_{t-1} should measure the true order flow responses (β_P^*) while those on ΔX_t should capture the immediate order flow response (β_P) Over-reaction implies that the estimate of β_P will be more extreme (i.e. more positive or negative) than β_P^*.

Table 7 presents the estimates of the elements of β_p and β_p^* from (4) in the first row, Newey-West t-statistics in the second row, and the difference between these estimates and Wald tests of the hypothesis that the elements of β_p and β_p^* are equal in the third row. The test statistic is distributed $\chi^2(1)$ and based on the Newey-West variance-covariance matrix. Recall that the coefficient on the second lag of each variable (shown by the suffix *.L2*) provides an estimate of the true coefficient, β_p^*, while the coefficient on the first lag of the change in the variable is an estimate of the overall response, β_p.

There are several results of interest. For most variables, the true coefficients, β_p^*, are of the same sign as, and of similar magnitude to, the original coefficients in Table 5. The coefficient on *TERM* is highly significant for all three portfolios and for four of the five sub-portfolios. The coefficient on *DEF* is significant at the 1% level for *OFSMB* and for three sub-portfolios, and is marginally significant for two others. For *OFSMB*, the true coefficients on *DEF* and *TERM* are -1.8 and 0.6 (as compared to -2.2 and 0.6 in Table 5). For *OFHML*, the true coefficient for *DEF* is a marginally insignificant -1.1 (as compared to -1.5 in Table 5), while that for *TERM* is virtually unaffected.

As in Section 5.1, the true coefficients for *VXO* and *LIQ* are significantly below zero for *OFHML*. In a departure from our earlier results, the coefficient on *NAPM* is negative and significant for both *OFMKT* and *OFHML*. The finding of a negative coefficient for *OFHML* is consistent with the behavioral story, which would suggest that investors buy shares in growth relative to value stocks when

NAPM is high and the economy is doing well. However, the negative coefficient on *OFMKT* is at odds with this story, since investors might be expected to buy stocks in general when the economy is doing well. Paralleling earlier results, the coefficients on *SENT* are never significant.

The signs of the actual coefficients, β_p, are usually the same as those of β_p^*. Several coefficients are either significant (e.g. those on *TERM* are significant for *OFSMB*, *OFHML* and two of the three SMB sub-categories) or close to being significant. The coefficient β_p is more extreme than β_p^* for most variables. For instance, β_p and β_p^* for *DEF* are -2.7 and -1.5 for *OFMKT*, -3.8 and -1.8 for *OFSMB*, and -2.6 and -1.1 for *OFHML*. The point estimates on *DEF* are, therefore, consistent with over-reaction, although the differences are not significant. A notable exception is the pair of coefficients on *VXO* for *OFHML*. Here, β_p is significantly more positive than β_p^*, which suggests that order flow actually under-reacts to volatility. For the remaining variables, the Wald test shows that the difference between β_p and β_p^* is not statistically significant.

In sum, our results indicate that order flow for SMB does not over-react to the variables capturing economic conditions. Order flow for HML appears to over-react to *VXO* but not to *DEF* or *TERM*. Thus, the analysis in this section suggests that the significant relations we have uncovered between order flows and the business cycle variables are not driven by investor over-reaction to these variables.

2.6 Concluding Remarks

The size and value premiums have alternately been ascribed to rational or behavioral forces. This paper studies monthly order flows for NYSE stocks in order to provide a new perspective on the sources of these premiums, one based on investor trades. To this end, we construct order flows for the SMB and HML portfolios and examine how they relate to aggregate performance (measured by the NAPM index), expected future conditions (measured by the default and term spreads), aggregate uncertainty (implied volatility on index options), liquidity and investor sentiment.

Our results favor a risk-based explanation. Order flows for SMB and HML decline in response to an anticipated deterioration in economic conditions, i.e. as the default spread increases or the term spread declines. There is some evidence that implied volatility and liquidity matter, but their effects are weaker. These patterns are consistent with big and growth stocks being regarded as hedges against adverse shifts in economic conditions. By contrast, neither the recent strength of the economy nor investor sentiment has significant explanatory power for SMB and HML order flows. This appears to be at odds with the argument that irrational extrapolation of past performance is the source of the value and size premiums. Further analysis indicates that the relations we have uncovered between order flows and the business cycle indicators are not driven by investor over-reaction.

Beyond these key results, we find that (a) there is additional comovement in order flow associated with size and BM and (b) commonality in trading accounts for one-third or more of the explanatory power of the three factor model proposed by Fama and French (1993). These results extend extant evidence from the microstructure literature that the common factors in order flows and returns are closely related (e.g. Hasbrouck and Seppi, 2001). They also point to the empirical importance of size and BM-based trading.

2.7 References

Acharya, V., and L. Pedersen, 2005, Asset Pricing with Liquidity Risk, *Journal of Financial Economics 77*, 375 - 410

Baker, M. and J. Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, *Journal of Finance 61*, 1645 - 1680

Balduzzi, P., E. Elton, and C. Green, 2001, Economic News and Bond Prices: Evidence from the U.S. Treasury Market, *Journal of Financial and Quantitative Analysis 36*, 523 - 543

Barberis, N. and A. Shleifer, 2003, Style Investing, *Journal of Financial Economics 68*, 161- 199

Berk, J., R. Green, and V. Naik, 1999, Optimal Investment, Growth Options, and Security Returns, *Journal of Finance 54*, 1553-1607

Chan, K. C. and N. Chen, 1991, Structural and Return Characteristics of Small and Large Firms, *Journal of Finance 46*, 1467 - 1484

Chan, K. C. and W. Fong, 2000, Trade Size, Order flow, and the Volatility–Volume Relation, *Journal of Financial Economics 57*, 247 - 273

Chen, N., 1991, Financial Investment Opportunities and the Macroeconomy, *Journal of Finance 46*, 529-554

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

Chordia, T., R. Roll, and A. Subrahmanyam, 2002, Order flow, Liquidity, and Market Returns, *Journal of Financial Economics 65*, 111 - 130

Chordia, T. and A. Subrahmanyam, 2004, Order flow and Individual Stock Returns: Theory and Evidence, *Journal of Financial Economics 72*, 485 - 518

Comerton-Forde, C., T. Hendershott, C. Jones, M. Seasholes, and P. Moulton (2009), Time Variation in Liquidity: The Role of Market Maker Inventories and Revenues, *Journal of Finance*, forthcoming.

Cooper, I., 2006, Asset Pricing Implications of Nonconex Adjustment Costs and Irreversibility of Investment, *Journal of Finance 61*, 139 - 170

Fama, E. and K. French, 1989, Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics 25,* 23 – 49

Fama, E. and K. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance 47,* 427-465

Fama, E. and K. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics 33,* 3-56

Fama, E. and K. French, 1995, Size and Book-to-Market Factors in Earnings and Returns, *Journal of Finance 50,* 131–155

Fama, E. and K. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance 51*, 55-84

Gebhardt, W. , S. Hvidkjaer, and B. Swaminathan., 2005, Stock and Bond Market Interaction: Does Momentum Spill Over?, *Journal of Financial Economics 75*, 651 - 690

Harford, J. and A. Kaul, 2005, Correlated Order Flow: Pervasiveness, Sources, and Pricing Effects, *Journal of Financial and Quantitative Analysis 40,* 29-55

Harris, Ethan, 1991, Tracking the Economy with the Purchasing Managers' Index, *Quarterly Review*, Federal Reserve Bank of New York, 61 - 69

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

Hasbrouck, J. and G. Sofianos, 1993, The Trades of Market Makers: An Empirical Analysis of NYSE Specialists, *Journal of Finance 48*, 1175–1199

Huberman,G. and S. Kandel, 1987, Mean-Variance Spanning, *Journal of Finance 42*, 873 - 888

Hvidkjaer, S., 2006, A Trade-Based Analysis of Momentum, *Review of Financial Studies 19*, 457 - 491

Kyle, A., 1985, Continuous Auctions and Insider Trading, *Econometrica 53*, 1315–1336

Lakonishok, J., A. Shleifer, and R. Vishny, 1994, Contrarian Investment, Extrapolation, and Risk, *Journal of Finance 49*, 1541 - 1578

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

Liew, J. and M. Vassalou, 2000, Can Book-to-Market, Size and Momentum be Risk Risk Factors that Predict Economic Growth?, *Journal of Financial Economics 57*, 221-245

Madhavan, A. and S. Smidt, 1993, An Analysis of Changes in Specialist Inventories and Quotations, *Journal of Finance 48*, 1595 - 1628

Newey, W. and K. West, 1987, A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica 55, 703 – 708

Pastor, L. and R. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy111*, 642 – 685

Schwert, G. W., 1989, Why Does Stock Market Volatility Change over Time, *Journal of Finance 44*, 1115- 1153

2.8 Tables

Table 2-1 Descriptive Statistics for Trading Variables

The trades for all ordinary common shares trading on the New York Stock Exchange (NYSE) are signed using the Lee-Ready (1991) algorithm and order flows are estimated at a monthly frequency as the difference between buyer-initiated trades and seller-initiated trades in terms of (i) the number of transactions (*OFN*), (ii) the number of shares traded (*OFV*), and (iii) the dollar volume traded (*OFD*). The total number of transactions (*TOTN*), the total number of shares traded in thousands of shares (*TOTV*), and the total dollar volume traded in millions of dollars (*TOTD*) are computed monthly for each stock and are used to compute the scaled order flow measures (order flow divided by total trading activity): *OFNX*, *OFVX*, and *OFDX*. Panel A presents the time-series averages of the cross-sectional means and standard deviations of the monthly stock-level order flows (both scaled and unscaled) and total trading activity measures, while Panel B displays the cross-sectional averages of the stock-by-stock time-series correlations among these variables. Coefficients significant at the 5% level are shown in boldface. The diagonal entries (in italics) show the contemporaneous time-series correlations between the trading variables and returns.

Panel A: Descriptive Statistics				Panel B: Correlations								
Variable	Unit	Mean	Std Dev	*OFN*	*OFNX*	*OFV*	*OFVX*	*OFD*	*OFDX*	*TOTN*	*TOTV*	*TOTD*
OFN	Trades	417.43	555.27	*0.16*								
OFNX	Percent	3.94	3.99	**0.69**	*0.22*							
OFV	Thousand Shares	655.80	682.80	**0.52**	**0.36**	*0.24*						
OFVX	Percent	3.64	4.60	**0.37**	**0.47**	**0.75**	*0.32*					
OFD	Million Dollars	26.76	22.63	**0.52**	**0.35**	**0.94**	**0.71**	*0.23*				
OFDX	Percent	3.94	4.53	**0.37**	**0.47**	**0.75**	**1.00**	**0.71**	*0.32*			
TOTN	Trades	5566.25	6277.88	**0.51**	**0.15**	**0.32**	0.13	**0.32**	0.13	*0.06*		
TOTV	Thousand Shares	7810.20	6045.97	**0.38**	0.12	**0.29**	0.07	**0.28**	0.07	**0.76**	*0.03*	
TOTD	Million Dollars	294.47	205.40	**0.37**	0.13	**0.28**	0.09	**0.34**	0.09	**0.74**	**0.89**	*0.04*

Table 2-2 Descriptive Statistics for Order Flow and Return Factors

The left panel of this table presents the means and standard deviations of the key variables, while the right panel displays the correlations. The top six rows report the descriptive statistics and correlations of the order flows and returns for the market, SMB, and HML portfolios, computed using the scaled share volume-based measure, *OFVX*. The order flow and excess return on the market portfolio (*OFMKT* and *MKT)* are computed as the cross-sectional average of monthly order flows (returns) for all NYSE stocks. To construct the SMB and HML portfolios, common stocks trading on the NYSE are allocated to two groups based on size (small or big: S or B) and three groups based on book-to-market equity (BM) (low, medium, or high: L, M, or H) at the end of June in each year *t*, using the Fama and French breakpoints. Thus, six size and BM portfolios (S/L, S/M, S/H, B/L, B/M, B/H) are formed for the next twelve months (July of year *t* through June of year *t+1*) for return and order flow measurement. The order flow and return for each portfolio are computed as the equally-weighted averages across the constituent stocks. SMB is a portfolio that is long in the three small stock portfolios (S/L, S/M, S/H) and short in the three big stock portfolios (B/L, B/M, B/H). HML is a portfolio that is long in the two high BM portfolios (S/H and B/H) and short in the two low BM stock portfolios (S/L and B/L). The order flow (return) for SMB (*OFSMB* and *SMB*) are defined as the difference between the simple averages of the monthly order flows (returns) for the three small stock portfolios and the three big stock portfolios. The order flow (return) for HML (*OFHML* and *HML*) are computed as the difference between the simple averages of the order flows (returns) for the two high BM portfolios and the two low BM portfolios. The bottom six rows contain information about the default and term spreads (*DEF* and *TERM*), OEX implied volatility (*VXO*), the Pastor and Stambaugh (2003) liquidity measure (*LIQ*), the National Association of Purchasing Managers Index (*NAPM*), and the Baker-Wurgler (2006) investor sentiment index (*SENT*), all measured with a one-month lag. *DEF* is the difference between the yields for Moody's seasoned Baa-and Aaa-grade bond portfolios. *TERM* is the difference between the 10-year Treasury constant maturity rate and the 3-month Treasury bill rate. Coefficients significant at the 5% level are shown in boldface.

	Descriptive Statistics		Correlations										
	Mean	Std Dev	*OFMKT*	*OFSMB*	*OFHML*	*MKT*	*SMB*	*HML*	*DEF*	*TERM*	*VXO*	*LIQ*	*NAPM*
OFMKT (%)	3.64	4.60	1.00										
OFSMB (%)	-5.51	2.79	**0.57**	1.00									
OFHML (%)	-0.84	2.54	**-0.28**	0.02	1.00								
MKT (%)	0.84	5.21	**0.33**	**0.33**	0.12	1.00							
SMB (%)	0.26	4.08	0.12	**0.35**	0.01	**0.58**	1.00						
HML (%)	0.62	3.74	-0.01	0.01	0.11	**-0.62**	**-0.46**	1.00					
DEF (%)	0.85	0.22	-0.04	**-0.17**	-0.10	0.05	0.05	-0.01	1.00				
TERM (%)	3.00	1.43	**0.38**	**0.44**	-0.05	**0.15**	**0.17**	0.08	**0.30**	1.00			
VXO (%)	21.07	7.01	0.10	**-0.22**	**-0.41**	0.05	-0.02	-0.14	**0.42**	-0.07	1.00		
LIQ	-2.11	5.35	-0.12	0.08	0.14	0.05	0.14	-0.14	**-0.15**	-0.02	**-0.42**	1.00	
NAPM	51.93	5.09	**0.18**	**0.37**	-0.10	-0.02	0.00	0.08	**-0.21**	**0.42**	**-0.25**	0.01	1.00
SENT	0.00	1.01	-0.02	0.07	0.00	0.11	0.09	-0.01	-0.05	0.06	-0.05	0.03	0.06

Table 2-3 Size/BM Induced Commonality in Stock Order Flows

Panel A reports the results from a test where the order flow for each stock is regressed, in stages, on its own lag, and the contemporaneous order flows for the market portfolio and the size-BM portfolio to which it belongs in month *t*. The first stage regression uses the own lagged order flow as the only explanatory variable, while the third stage regression uses all three order flows. The second stage regression is conducted by adding either the market portfolio or the size-BM portfolio order flow to a model that includes the lagged order flow for the stock. The market and portfolio order flows are adjusted to exclude the own stock order flow. Panel A presents the cross-sectional means and medians of the adjusted R-square values from these regressions. The results for two scaled order flow measures (number of transactions, *OFNX*, and share volume, *OFVX*) are reported. Panel B presents regressions of category-switching-stock order flow on the contemporaneous order flows for the original and new size-BM portfolio, both adjusted to exclude the own stock order flow. Stocks that switch between size-BM categories and remain in the pre-switch and post-switch portfolios for at least two years are identified, and two regressions are estimated for the order flow for each switching stock: one before and the other after the switch. The cross-sectional means and standard deviations (in parentheses) of the coefficient estimates are presented in Panel B along with t-statistics testing the null hypothesis that the pre-switch slope is equal to the post-switch slope. The intercepts are suppressed.

Panel A: Stagewise Regressions of Individual Stock Order Flows		
OFNX	Mean R-sq	Median R-sq
Just Own OF	0.19	0.16
Add Portfolio OF	0.26	0.24
Add Market OF	0.29	0.27
Just Own OF	0.19	0.16
Add Market OF	0.26	0.23
Add Portfolio OF	0.29	0.27
OFVX	Mean R-sq	Median R-sq
Just Own OF	0.06	0.04
Add Portfolio OF	0.11	0.09
Add Market OF	0.14	0.12
Just Own OF	0.06	0.04
Add Market OF	0.12	0.10
Add Portfolio OF	0.14	0.12

Panel B: Comovement Test for Stocks Switching Categories		
	Mean Coefficient Estimate	
	Pre-switch	Post-Switch
Old Portfolio Order Flow	0.85	0.20
	1.80	1.71
New Portfolio Order Flow	0.24	0.84
	1.80	1.96
t(PRE - POST)	6.07	-6.66

Table 2-4 Linking Common Factors in Returns and Order Flows

Panel A presents the coefficient estimates, Newey-West (1987) t-statistics, and adjusted R^2 values from time-series regressions of the excess monthly returns for six size-BM portfolios (S/L, S/M, S/H, B/L, B/M, and B/H) on the common return factors: *MKT*, *SMB*, and *HML*. Each of these factors is adjusted to exclude the own portfolio return. Panel B reports the results of these regressions after adjusting the size-BM portfolio returns for trade comovement. The trade-adjusted return is computed, for each portfolio, as the residual from a regression of the portfolio return on the common order flow factors: *OFMKT*, *OFSMB*, and *OFHML*. Each order flow factor is adjusted to exclude the own portfolio order flow in order to mitigate simultaneity between order flows and returns. The size-BM portfolios are as defined in the notes for Table 2.

Panel A: Excess Returns before Adjustment for Trades									
	INT	*MKT*	*SMB*	*HML*	t(*INT*)	t(*MKT*)	t(*SMB*)	t(*HML*)	R^2
S/L	-0.001	1.062	0.565	-0.614	-0.710	31.126	8.078	-6.593	0.88
S/M	0.001	0.944	0.315	0.007	0.626	37.240	7.550	0.143	0.93
S/H	0.001	1.008	0.626	0.559	1.045	30.822	8.560	6.109	0.87
B/L	0.001	0.906	-0.442	-0.638	1.100	49.557	-6.359	-10.649	0.89
B/M	-0.002	1.018	-0.467	-0.026	-2.067	40.479	-11.515	-0.615	0.94
B/H	0.002	0.953	-0.520	0.627	1.830	18.980	-11.228	7.421	0.85
Panel B: Excess Returns adjusted for Trading Effects									
S/L	-0.007	0.587	0.254	-0.674	-3.306	9.997	2.935	-4.652	0.46
S/M	-0.008	0.576	0.112	-0.052	-4.771	10.564	2.114	-0.629	0.54
S/H	-0.007	0.504	0.227	0.165	-4.746	9.546	2.346	1.465	0.40
B/L	-0.005	0.522	-0.445	-0.717	-3.074	11.260	-6.038	-6.060	0.55
B/M	-0.007	0.606	-0.487	-0.144	-5.730	12.800	-7.439	-1.833	0.56
B/H	-0.007	0.562	-0.590	0.432	-3.500	8.988	-7.809	4.235	0.48

Table 2-5 Market Conditions and the Order Flow Factors

This table reports the slope estimates and Newey-West (1987) t-statistics from regressions of month *t* *OFMKT*, *OFSMB*, and *OFHML* on the month *t*-1 values of the default and term spreads (*DEF* and *TERM*), implied market volatility (*VXO*), market liquidity (*LIQ*), the National Association of Purchasing Managers Index (*NAPM*), and investor sentiment (*SENT*), controlling for month *t*-1 market and own portfolio returns (*MKT.L* and *RETP.L*) and own order flow (*OFP.L*), and a deterministic time trend (*TIME*). The intercept terms are suppressed in order to save space. The F-statistic and p-value reported in the third line for each portfolio are from a test of the joint significance of the business cycle variables (*DEF* and *TERM*); market volatility and liquidity (*VXO* and *LIQ*); and recent economic strength and investor sentiment (*NAPM* and *SENT*). *MKT.L* and *RETP.L* are the same in the model where *OFMKT* is the dependent variable. The variables are defined in the notes for Table 2. We also divide *OFSMB* and *OFHML* into subcategories. $OFSMB_L$, $OFSMB_M$ and $OFSMB_H$ are the differences between small and big stock order flows in the low, medium and high BM categories, while $OFHML_S$ and $OFHML_B$ are the differences between high and low BM stock order flows in the small and big size categories. Coefficients significant at the 5% level are shown in boldface.

Panel A: Portfolio Order Flows											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
OFMKT	Coefficient	-1.004	**0.450**	-0.013	-0.020	-0.051	-0.226	0.000		**0.458**	**0.031**
	t-stat	-1.097	2.923	-0.392	-0.689	-1.492	-1.306	0.003		3.385	3.420
	Model F-stat	***3.655***	*0.028*	*0.185*	*0.831*	*1.655*	*0.194*				
OFSMB	Coefficient	**-2.180**	**0.578**	-0.005	0.006	0.020	-0.084	**0.134**	-0.002	**0.340**	**0.007**
	t-stat	-2.641	3.467	-0.157	0.166	0.583	-0.449	3.580	-0.026	4.824	2.172
	Model F-stat	***8.097***	*0.000*	*0.043*	*0.958*	*0.292*	*0.747*				
OFHML	Coefficient	**-1.514**	**0.307**	**-0.048**	**-0.049**	**-0.068**	-0.068	**0.117**	0.056	**0.169**	**-0.020**
	t-stat	-2.164	3.118	-1.861	-1.905	-2.345	-0.562	4.103	0.891	2.225	-6.691
	Model F-stat	***3.234***	*0.042*	***2.389***	*0.094*	*2.558*	*0.080*				
Panel B: Book-to-Market Subcategories for OFSMB											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
OFSMBL	Coefficient	-0.415	**0.543**	-0.014	0.025	0.062	-0.084	**0.127**	-0.039	**0.226**	**0.017**
	t-stat	-0.380	2.745	-0.377	0.601	1.348	-0.475	2.771	-0.571	3.006	4.413
	Model F-stat	***5.308***	*0.006*	*0.342*	*0.711*	*0.996*	*0.371*				
OFSMBM	Coefficient	**-2.301**	**0.548**	-0.005	-0.013	-0.017	0.012	**0.145**	-0.007	**0.322**	0.004
	t-stat	-2.800	3.654	-0.158	-0.457	-0.465	0.050	4.052	-0.081	3.837	1.235
	Model F-stat	***6.277***	*0.002*	*0.073*	*0.930*	*0.100*	*0.905*				
OFSMBH	Coefficient	**-4.311**	**0.781**	-0.013	0.008	0.040	-0.127	**0.152**	0.005	**0.241**	0.003
	t-stat	-3.272	2.915	-0.250	0.158	0.739	-0.574	2.775	0.053	3.798	0.606
	Model F-stat	***7.157***	*0.001*	*0.068*	*0.935*	*0.395*	*0.674*				
Panel C: Firm Size Subcatefories for OFHML											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
OFHMLS	Coefficient	**-3.707**	**0.465**	-0.048	**-0.053**	-0.079	-0.141	**0.139**	-0.026	**0.162**	**-0.027**
	t-stat	-3.057	2.346	-1.191	-2.083	-1.711	-1.145	3.479	-0.314	2.602	**-6.597**
	Model F-stat	***4.969***	*0.008*	*1.175*	*0.311*	*1.660*	*0.193*				
OFHMLB	Coefficient	0.523	0.174	-0.043	-0.039	-0.053	-0.088	**0.097**	0.005	**0.239**	**-0.011**
	t-stat	0.719	1.259	-1.675	-1.178	-1.565	-0.558	3.115	0.100	4.043	-4.015
	Model F-stat	*1.650*	*0.195*	*1.181*	*0.309*	*1.244*	*0.291*				

Table 2-6 Market Conditions and the Return Factors

This table reports the slope estimates and Newey-West (1987) t-statistics from regressions of month *t MKT*, *SMB*, and *HML* on the month *t*-1 values of the default and term spreads (*DEF* and *TERM*), implied market volatility (*VXO*), market liquidity (*LIQ*), the National Association of Purchasing Managers Index (*NAPM*), and investor sentiment (*SENT*), controlling for month *t*-1 market and own portfolio returns (*MKT.L* and *RETP.L*) and the portfolio's order flow (*OFP.L*), and a deterministic time trend (*TIME*). The intercept terms are suppressed in order to save space. The F-statistic and p-value reported in the third line for each portfolio are from a test of the joint significance of the business cycle variables (*DEF* and *TERM*); market volatility and liquidity (*VXO* and *LIQ*); and recent economic strength and investor sentiment (*NAPM* and *SENT*). *MKT.L* and *RETP.L* are the same in the model where *MKT* is the dependent variable. The variables are defined in the notes for Table 2. We also divide *SMB* and *HML* into subcategories. SMB_L, SMB_M and SMB_H are the differences between small and big stock returns in the low, medium and high BM categories, while HML_S and HML_B are the differences between high and low BM stock returns in the small and big size categories. Coefficients significant at the 5% level are shown in boldface.

Panel A: Portfolio Returns											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
MKT	Coefficient	-2.309	**0.723**	0.114	0.072	-0.110	-0.096	**0.265**		-0.076	0.003
	t-stat	-1.033	2.482	1.633	0.584	-1.293	-0.262	2.336		-0.489	0.263
	Model F-stat	*2.181*	*0.116*	*1.293*	*0.277*	*0.871*	*0.420*				
SMB	Coefficient	-1.869	**0.730**	0.072	0.099	-0.058	-0.194	**0.372**	-0.205	-0.227	0.004
	t-stat	-1.102	3.226	1.357	1.143	-1.015	-0.480	4.699	-1.241	-1.052	0.936
	Model F-stat	***4.007***	*0.020*	*1.937*	*0.147*	*0.643*	*0.527*				
HML	Coefficient	1.172	-0.077	**-0.137**	-0.174	0.004	-0.053	**0.132**	0.131	0.052	0.006
	t-stat	0.823	-0.407	-2.171	-1.306	0.074	-0.130	2.340	0.935	0.498	1.096
	Model F-stat	*0.283*	*0.754*	***6.671***	*0.002*	*0.015*	*0.985*				
Panel B: Book-to-Market Subcategories for SMB											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
SMBL	Coefficient	-2.376	**0.713**	0.104	0.044	-0.122	-0.129	**0.459**	-0.240	-0.137	0.004
	t-stat	-1.240	3.078	1.637	0.447	-1.620	-0.375	4.156	-1.381	-1.009	0.635
	Model F-stat	***2.752***	*0.066*	*1.229*	*0.295*	*1.374*	*0.256*				
SMBM	Coefficient	-1.149	**0.509**	0.080	0.139	-0.045	-0.123	**0.268**	-0.169	-0.108	0.001
	t-stat	-0.810	2.678	1.622	1.688	-0.975	-0.327	4.515	-1.269	-0.788	0.290
	Model F-stat	***2.387***	*0.095*	***4.100***	*0.018*	*0.420*	*0.657*				
SMBH	Coefficient	-1.391	**0.787**	0.046	0.111	-0.030	-0.343	**0.345**	-0.144	-0.160	0.003
	t-stat	-0.644	3.066	0.800	1.213	-0.597	-0.747	4.185	-1.156	-1.063	0.739
	Model F-stat	***4.579***	*0.011*	*1.853*	*0.160*	*0.695*	*0.500*				
Panel C: Firm Size Subcatefories for HML											
		DEF	TERM	VXO	LIQ	NAPM	SENT	MKT.L	RETP.L	OFP.L	TIME
HMLS	Coefficient	1.815	-0.085	**-0.148**	-0.158	0.034	-0.100	**0.170**	0.217	0.046	0.006
	t-stat	1.045	-0.396	-2.346	-1.254	0.548	-0.227	2.157	1.205	0.569	0.975
	Model F-stat	*0.564*	*0.570*	***5.287***	*0.006*	*0.164*	*0.849*				
HMLB	Coefficient	0.496	-0.078	**-0.128**	-0.195	-0.034	-0.029	**0.122**	0.067	-0.010	0.004
	t-stat	0.434	-0.357	-1.976	-1.425	-0.564	-0.077	1.686	1.030	-0.103	0.870
	Model F-stat	*0.066*	*0.936*	***7.199***	*0.001*	*0.159*	*0.853*				

Table 2.7 Testing Over-Reaction to Market Conditions

This table reports the slope estimates and Newey-West (1987) t-statistics from regressions of *OFMKT*, *OFSMB*, and *OFHML* in month t+1 on month *t*-1 values of the default and term spreads (*DEF* and *TERM*), implied volatility (*VXO*), market liquidity (*LIQ*), the National Association of Purchasing Managers Index (*NAPM*), and investor sentiment (*SENT*) and the changes in these variables between months *t*-1 and *t*, controlling for month *t* market and own portfolio returns (*MKT.L* and *RETP.L*) and own order flow (*OFP.L*) and a deterministic time trend (*TIME*). The intercept terms are suppressed to save space. The difference between the contemporaneous coefficient (β, the coefficient on the change) and the true coefficient (β*, the coefficient on the level) is presented in the last row of every specification, together with a chi-square statistic from a Wald test for significance. The critical 5% (10%) $\chi^2(1)$ values are 5.024 (3.841). The details are in Section 5.3 and the notes for Tables 2 and 5. Coefficients significant at the 5% level are shown in boldface.

Panel A: Portfolio Order Flows																	
		ΔDEF	DEF.L2	ΔTERM	TERM.L2	ΔVXO	VXO.L2	ΔLIQ	LIQ.L2	ΔNAPM	NAPM.L2	ΔSENT	SENT.L2	MKT.L	RETP.L	OFP.L	TIME
OFMKT	Estimate	-2.695	-1.535	-0.005	**0.497**	-0.053	-0.005	-0.019	-0.031	0.098	**-0.079**	-0.216	-0.401	-0.043		**0.465**	**0.030**
	t-statistic	-1.335	-1.473	-0.020	2.797	-0.951	-0.131	-0.684	-0.615	0.931	-2.144	-1.298	-1.303	-0.868		3.729	3.501
	β - β*	*-1.160*	*(0.418)*	*-0.502*	*(3.761)*	*-0.048*	*(1.152)*	*1.220*	*(0.141)*	*0.177*	*(2.611)*	*0.185*	*(1.037)*				
OFSMB	Estimate	-3.758	**-1.821**	**0.640**	**0.553**	0.059	-0.011	0.005	0.014	0.019	0.021	-0.113	-0.114	**0.169**	-0.017	**0.343**	**0.007**
	t-statistic	-1.728	-2.209	2.677	3.236	0.976	-0.298	0.130	0.255	0.229	0.690	-0.570	-0.444	3.782	-0.248	5.061	2.248
	β - β*	*-1.937*	*(0.750)*	*0.087*	*(0.158)*	*0.070*	*(1.633)*	*-0.936*	*(0.098)*	*-0.002*	*(0.001)*	*0.001*	*(0.000)*				
OFHML	Estimate	-2.621	-1.091	0.241	**0.291**	0.044	**-0.073**	**-0.052**	**-0.091**	-0.042	**-0.069**	-0.104	-0.198	**0.164**	0.015	**0.160**	**-0.020**
	t-statistic	-1.474	-1.623	1.181	2.949	0.921	-2.957	-2.262	-2.737	-0.607	-2.372	-0.799	-1.056	4.499	0.228	2.316	-7.077
	β - β*	*-1.530*	*(1.025)*	*-0.050*	*(0.063)*	***0.117***	*(13.689)*	*3.889*	*(1.627)*	*0.027*	*(0.122)*	*0.094*	*(0.552)*				
Panel B: Book-to-Market Subcategories for OFSMB																	
		ΔDEF	DEF.L	ΔTERM	TERM.L	ΔVXO	VXO.L	ΔLIQ	LIQ.L	ΔNAPM	NAPM.L	ΔSENT	SENT.L	MKT.L	RETP.L	OFP.L	TIME
OFSMBL	Estimate	-4.037	0.171	0.596	**0.486**	0.079	-0.030	0.022	0.006	0.093	0.061	-0.117	-0.165	**0.173**	-0.061	**0.231**	**0.018**
	t-statistic	-1.458	0.161	1.412	2.390	1.143	-0.726	0.509	0.089	0.709	1.414	-0.623	-0.687	3.008	-0.825	3.091	4.418
	β - β*	*-4.208*	*(2.750)*	*0.110*	*(0.070)*	*0.109*	*(2.376)*	*1.597*	*(0.205)*	*0.032*	*(0.079)*	*0.048*	*(0.062)*				
OFSMBM	Estimate	-2.875	**-2.078**	**0.524**	**0.555**	0.058	-0.011	-0.017	-0.010	-0.029	-0.020	0.009	0.124	**0.180**	-0.015	**0.316**	0.004
	t-statistic	-1.388	-2.519	2.103	3.548	1.101	-0.337	-0.632	-0.197	-0.362	-0.595	0.040	0.469	4.328	-0.176	3.699	1.430
	β - β*	*-0.797*	*(0.168)*	*-0.031*	*(0.016)*	*0.069*	*(2.381)*	*-0.693*	*(0.043)*	*-0.009*	*(0.014)*	*-0.115*	*(1.202)*				
OFSMBH	Estimate	-5.402	**-4.164**	**0.784**	**0.791**	0.040	-0.004	0.011	0.057	0.027	0.041	-0.180	-0.25	**0.185**	-0.005	**0.237**	0.003
	t-statistic	-1.563	-2.986	2.436	2.802	0.421	-0.069	0.224	0.796	0.225	0.827	-0.738	-0.575	2.644	-0.054	3.842	0.699
	β - β*	*-1.238*	*(0.105)*	*-0.007*	*(0.001)*	*0.044*	*(0.323)*	*-4.534*	*(0.946)*	*-0.014*	*(0.018)*	*0.070*	*(0.072)*				
Panel C: Firm Size Subcatefories for OFHML																	
		ΔDEF	DEF.L	ΔTERM	TERM.L	ΔVXO	VXO.L	ΔLIQ	LIQ.L	ΔNAPM	NAPM.L	ΔSENT	SENT.L	MKT.L	RETP.L	OFP.L	TIME
OFHMLS	Estimate	-3.608	**-3.502**	0.332	**0.480**	**0.036**	-0.060	**-0.054**	-0.061	-0.066	-0.082	-0.162	-0.273	**0.185**	-0.035	**0.154**	**-0.028**
	t-statistic	-1.257	-2.811	0.847	2.382	0.501	-1.619	-2.224	-1.132	-0.633	-1.883	-1.185	-1.203	3.266	-0.408	2.551	-6.499
	β - β*	*-0.106*	*(0.002)*	*-0.148*	*(0.133)*	***0.096***	*(4.608)*	*0.700*	*(0.021)*	*0.016*	*(0.028)*	*0.111*	*(0.397)*				
OFHMLB	Estimate	-1.580	1.198	0.118	0.124	0.075	**-0.082**	-0.046	**-0.118**	-0.003	-0.053	-0.127	-0.22	**0.155**	-0.051	**0.226**	**-0.011**
	t-statistic	-0.765	1.691	0.425	0.959	1.340	-3.042	-1.540	-3.463	-0.028	-1.431	-0.735	-0.891	3.639	-0.769	3.850	-4.269
	β - β*	*-2.778*	*(1.769)*	*-0.006*	*(0.001)*	***0.157***	*(6.162)*	***7.249***	*(7.234)*	*0.050*	*(0.357)*	*0.093*	*(0.412)*				

Chapter 3

Forecasting Macroeconomic Fundamentals and Expected Stock Returns with Equity-Market Order Flows

3.1 Introduction

How much does the general direction of trading activity in the stock market tell us about future market conditions? What does it mean when the net purchasing activity for stocks with certain defining characteristics, such as size or liquidity, increase disproportionately with respect to others? In a setting where information is distributed heterogeneously across agents, net order flow for broad portfolios may aggregate dispersed information and provide a valuable signal about how investors bet on their expectations about fundamentals with their wallets. Indeed, a recent literature provides evidence that aggregate order flow in the foreign exchange and bond markets reveals information about macroeconomic fundamentals (Brandt and Kavajecz, 2004; Green, 2004; Pasquariello and Vega, 2006; Evans and Lyons, 2009; Beber, Brandt, and Kavajecz, 2008).[24] This paper adds to this literature by investigating the predictive power of equity market order flow for (a) economic growth and (b) stock returns.

[24] Green (2004) finds that intraday order flow in the U.S. Treasury market reveals fundamental information about riskless rates. The author shows that the informational role of order flow increases after public information releases, consistent with the notion that some investors are better than others in converting public information into private forecasts. Brandt and Kavajecz (2004) show that order flow explains about a quarter of the daily variation in yields on days with no economic announcements in the Treasury market, with the effect being stronger when liquidity is low. Pasquariello and Vega (2006) find that the unanticipated U.S. Treasury bond market order flow has a significant impact on daily bond yield changes on both announcement and non-announcement days, with the effect being stronger when the dispersion of beliefs is high and the announcements are noisy. Evans and Lyons (2009) show that foreign exchange order flows forecast macro fundamentals and foreign exchange returns.

Our analysis focuses on two distinct aggregate order flow measures. The first measure, *market order flow* (*OFM*), is the value-weighted cross-sectional average of individual stock order flows estimated from intraday trade and quote data using the Lee-Ready (1991) algorithm. *OFM* parallels the aggregate bond and foreign exchange market order flows studied elsewhere, and captures overall buying or selling pressure exerted by trade initiators who place market orders and demand immediacy. Provided that there is a class of investors who trade solely for liquidity reasons, we conjecture that *OFM* should reflect the exchange that take place between these liquidity traders and relatively more sophisticated portfolio optimizers that is brought about by the effect of changing consumption and investment opportunities on the optimal portfolio allocation to stocks.

The second measure, *order flow differential* (*OFD*), is novel and specific to the stock market. We define *OFD* as the difference between the average buying pressures generated by the active big and small stock traders in a given period.[25] We conjecture that this measure should capture the time variation in intertemporal hedging demand induced by the strategic behavior of investors who wish to hedge against adverse changes in future consumption and investment opportunities. That is, as small stock returns are more sensitive to marketwide fluctuations than big stock returns, deterioration of economic expectations (or an accompanying increase in risk aversion) should result in a disproportionate decline in the fraction of wealth allocated to small stocks in relation to the fraction of wealth allocated to big stocks.[26] This would lead to an exchange of securities between sophisticated hedgers and liquidity traders, which is eventually picked up by *OFD*.

[25] Note that both of these measures focus on the market orders that require immediate execution and omit the trades executed through the limit order book. The rationale here is that the trades by aggressive investors are much more likely to be based on proprietary information than the trades of passive investors, since proprietary information (on certain cash flow and macroeconomic forecasts) tends to have an "expiry date." Lastly, any persistent imbalance in order flows should be thought of as being accommodated either by offsetting trades from the limit order book (where less aggressive investors place their trades) or through market maker inventories.

[26] For risky assets with procyclical returns, the hedging demand of a risk-averse investor is shown to be negative and declining in the coefficient of relative risk aversion and in the covariance between asset returns and future consumption and investment opportunities. See Restoy (1992) for an initial derivation of optimal

Using stock-level order flows constructed from high frequency data, we compile the two order flow aggregates quarterly over the period January 1988 through December 2004. We start our analysis by examining the predictive power of *ODM* and *ODD* for future economic output growth, as measured by the quarterly growth rates of real GDP, industrial production, and corporate earnings (*QPG*, *QYG*, and *QEG*). Our results show that *OFM* is related positively to future growth rates for real GDP and industrial production, but not for corporate earnings: a one standard deviation increase in *OFM* forecasts an increase of about 0.21 to 0.38 standard deviations in *QPG* (0.23 to 0.41 percent) and about 0.18 to 0.35 standard deviations in *QYG* (0.10 to 0.18 percent) over the four subsequent quarters. *OFD*, on the other hand, is related negatively to all the three proxies for future economic growth and its predictive power is even stronger: a one standard deviation increase in *OFD* forecasts a decline of about 0.33 to 0.48 standard deviations in *QPG* (0.36 to 0.51 percent), 0.26 to 0.39 standard deviations in *QYG* (0.13 to 0.20 percent), and 0.21 to 0.31 standard deviations in *QEG* (1.08 to 1.64 percent). These relations are robust to the inclusion of the lagged economic growth rates and contemporaneous return factors from a four-factor model including the excess market return and the size, value, and momentum premiums.

The findings above parallel the evidence from the foreign exchange market reported in Evans and Lyons (2009) that the information in order flows is not captured by returns.[27] A potential explanation, suggested by the work of Chan (1993), is that market makers are unable to immediately extract the marketwide component of a noisy firm-level signal (embedded in order flow, in our case) and, instead, assimilate this information over time as they learn from the signals of

---

portfolio weights under time-dependent returns and Campbell and Viceira (1999) for a thorough treatment of the optimal consumption and portfolio choice problem of an infinitely-lived Epstein-Zin-Weil utility maximizer who faces a constant riskless interest rate and a time-varying and partially predictable equity premium.

[27] Evans and Lyons (2009) present a general equilibrium model where fundamental information that is first manifested at the firm-level and is not symmetrically observed by all agents provides foreign exchange (FX) market order flow with an important role in aggregating information. Based on their model, the authors conjecture that FX market order flows should forecast future macro fundamentals, do so significantly better than FX returns, and forecast FX returns. Their empirical tests support these conjectures.

other stocks in subsequent periods.[28] Note that the noise thus induced in returns may not wash away in aggregation if it is correlated across market-makers.[29] The hypothesis that follows from this reasoning is that, if the macroeconomic signal in the order flow measures is not impounded in prices in a timely manner, our aggregate order flow measures should predict stock market returns. To address this issue, we regress the future quarterly returns for ten size-sorted portfolios and the future realizations of the market, size, value, and momentum premiums on the quarterly changes in *OFM* and *OFD*. We expect a positive relation between *OFD* and expected returns as the hedging component of demand will be more pronounced when future outlook is dim and risk aversion is high. The relation between *OFM* and expected returns, on the other hand, can go either way: it may be positive if the information in market order flow is only partially incorporated into prices (the noisy macro signal story) or negative because of the negative link between realized and expected returns. In order to ensure that the information in the two order flow variables are unique, we extend the set of control variables with several business cycle indicators (default spread, term spread, forecasted earnings growth, and new equity additions: *DEF*, *TERM*, *FEG* and *NEQ*), proxies for liquidity and investor sentiment (Pastor-Stambaugh (2003) liquidity measure and the Baker-Wurgler (2006) sentiment index: *LIQ* and *ΔSENT*), and lagged portfolio returns.[30]

---

[28] While the model is devised to explain the positive cross-autocorrelations in stock returns, the driving idea that market makers cannot immediately assimilate the macro information embedded in the noisy micro signal they receive for their own stock proves useful in providing a framework within which to view our results.

[29] Additionally, Albuquerque, Francisco, and Marques (2008) develop a model of equity trading where private information can be firm-specific or marketwide and show that an industry-level measure of marketwide private information extracted from intraday trade and quote data forecasts industry and foreign exchange market returns.

[30] Fama and French (1989) and Chen (1991) show that *DEF* and *TERM* predict future stock and bond portfolio returns. Baker and Wurgler (2000) find that a greater share of equity in new debt and equity issues forecasts stock market returns. Note that our measure of new equity, *NEQ*, is the growth rate of the total market capitalization of the index less the value-weighted market return. Pastor and Stambaugh (2003) demonstrate that a *LIQ* explains cross-sectional variation in expected stock returns. Lastly, Baker and Wurgler (2006) study how investor sentiment affects the cross-section of stock returns and provide evidence that sentiment shocks should have a more pronounced effect on securities with more subjective valuations.

Our tests reveal that *OFM* and *OFD do* have significant predictive power for stock market returns. An increase in *OFM* in quarter *t* forecasts higher quarter *t*+1 returns for most size-sorted decile portfolios (with the exception of the three largest portfolios), but not for any of the four return premiums. Controlling for *OFD*, a one standard deviation change in *OFM* forecasts an increase of 0.21 to 1.41 percent (0.03 to 0.13 standard deviations) in the decile portfolio returns. This forecast power, however, is mostly subsumed when contemporaneous return factors are added to the model as controls, and disappears totally with the inclusion of the rest of the control variables. Unlike *OFM*, the forecast power of *OFD* is robust to the inclusion of the contemporaneous return factors, business-cycle indicators, marketwide liquidity, and investor sentiment. Keeping all else constant, a one percent increase in *OFD* forecasts an increase of 0.40 percent (0.15 standard deviations) in the excess market return, 0.44 percent (0.24 standard deviations) in *SMB* , and a rise between 0.76 and 2.79 percent (0.11 and 0.21 standard deviations) in the decile portfolio returns.

The positive relation between *OFM* and subsequent small stock returns shows that it takes time for small stock prices to fully reflect the signal embedded in marketwide order flow. This is (a) plausible as the macro signal would be easier to detect for market makers in big stocks since the noise is diversified to a certain extent due to the greater scale of such firms' operations and (b) consistent with the lead-lag relation between big and small stock returns first documented in Lo and MacKinlay (1990). The finding that the explanatory power of *OFM* is subsumed when we account for liquidity and other controls is in line with Albuquerque et al. (2008), who find that a simple statistical factor of equity-market order flows captures mostly liquidity. The strength and robustness of the forecast power of the order flow differential across size deciles, on the other hand, signals a more pervasive effect. The positive relation between *OFD* and subsequent returns is consistent with investors reallocating portfolios from more to less procyclical assets as risk aversion increases prior to economic downturns. Further investigation confirms that the observed effect is distinct from liquidity: a size-controlled order flow differential between liquid and illiquid stocks behaves much like *OFM* does, and fails to achieve the strong explanatory power displayed by *OFD*. Ultimately, the evidence that the information in our aggregate order flows is not incorporated into stock prices for extended periods is striking. In particular, it is intriguing that common return factors, including the excess market return and *SMB* —closely linked to *OFM* and *OFD*—do not subsume the signal in aggregate order flows. To the best of our knowledge, this paper is the first to analyze the predictive content of equity market order flows for fundamentals and expected stock returns.

A recent paper by Beber, Brandt, and Kavajecz (2008) analyzes order flow movements across sectors of the economy and shows that a portfolio based on cross-sector order flows dominates the market portfolio, particularly during economic downturns. We view the two papers as complementary, and our analysis differs from theirs in several respects. First, Beber et al. (2008) use sector order flows and returns to predict the Chicago FED National Activity Index and stock and five-year bond returns, while we use aggregate order flow to predict real GDP, production, and earnings growth. Second, we introduce *OFD* as a novel proxy that captures time-variation in intertemporal hedging demand and forecasts future fundamentals and stock returns. Third, by controlling for a host of economic indicators and return factors, we verify the uniqueness of the signal contained in our measures. The results in Beber et al. (2008) are mostly from univariate relations between sector order flows, macro fundamentals, and returns.

The rest of the paper is organized as follows. Section 2 details our data and variables. Section 3 presents a brief review of prior research relating to our study. Section 4 presents and discusses our results on the forecast power of order flows for macroeconomic growth. Section 5 reports our findings from predictive regressions for future stock returns. Section 6 distinguishes between liquidity effects and hedging behavior. Section 7 concludes.

3.2 Literature Review and Research Questions

A recently developing body of research suggests that aggregate order flow in financial markets contain information about macroeconomic fundamentals. Studying the government bond market, Green (2004) finds that aggregate order flow for U.S. Treasury bonds reveals fundamental information about riskless rates. The author shows that the informational role of order flow increases after public information releases, consistent with the notion that some investors are better than others in converting public information into private forecasts. Similarly, Brandt and Kavajecz (2004) find that about a quarter of the daily variation in U.S. Treasury yields on non-announcement days is explained by order flow, with the effect being permanent and strongest when liquidity is low. Pasquariello and Vega (2006) find that the unanticipated U.S. Treasury bond market order flow has a significant and permanent impact on daily yield changes during both announcement and non-announcement days, with the effect being stronger when the dispersion of beliefs among investors is high and the announcements are noisy. In the foreign exchange (FX) market, Evans and Lyons (2009) show that FX order flows forecast macroeconomic fundamentals (such as the output growth, money growth, and inflation) and future exchange rates.

Despite the importance of the information aggregating role of order flow, evidence from the stock market is scarce. Albuquerque, Francisco, and Marques (2008) develop a model of equity trading where private information can be firm-specific or marketwide. The authors demonstrate that (a) a measure of marketwide private information (MPI) estimated from intraday order flows is shown to forecast FX and industry-level stock returns, (b) market order flow displays little correlation with MPI, and (c) the comovement in order flow is mostly liquidity-related. In a recent study that draws a close parallel with our paper, Beber, Brandt, and Kavajecz (2008) analyze the order flow movements across sectors of the economy and show that an order flow portfolio based on cross-sector movements dominates the market portfolio particularly during economic downturns. This paper, on the other hand, studies the predictive power of the market order flow and, critically, the average order flow difference between big and small stocks, which we refer to as the *order flow differential*.

The relationship between market order flow and future changes in economic growth should be positive since investors would demand more stock if their expectations about future fundamentals are favorable. The relation between market order flow and expected returns, on the other hand, is uncertain. The returns could be higher if the information in market order flow is only partially incorporated into the prices at the time when it is realized, since future prices will then reflect the information in current order flow. The returns could be lower if the information is fully incorporated at the time when order flow is realized because of the negative relation between realized return and expected return.

How does the order flow differential relate to future economic growth and stock returns? Evidence from the literature on the optimal consumption and portfolio choice provides us with useful insights. Restoy (1992) solves the \portfolio allocation problem of an infinitely-lived Epstein-Zin-Weil utility-maximizer facing state-dependent returns. The author shows that the single-period portfolio allocation for a risky asset is the sum of a myopic single period demand and an intertemporal hedging demand. Campbell and Viceira (1999) extend these results by solving the consumption and portfolio allocation problems analytically and demonstrate that the hedging component comprises a significant part (between 20 and 50 percent) of the demand for stocks by long-lived risk-averse investors. The intuition from both studies is that the demand for a risky asset is decreasing in both the coefficient of relative risk aversion and the covariance between the risky asset return and future investment opportunities.[31]

[31] We note that most of the models in this line of literature are representative agent models where portfolios are adjusted through price changes, without the need for trading. While we do not devise a structural model to link the trading process to portfolio choice, we argue that the existence of heterogeneously informed agents (for instance, informed hedgers versus uninformed liquidity traders) would necessitate trading.

Although the aforementioned results are for the case of a single risky asset and a risk-free security, we argue that the hedging component of the demand should be more negative for risky assets whose current returns covary more strongly with expected returns.[32] Since big stocks tend to deliver more wealth than small stocks (in form of more stable dividend streams and less procyclical returns) when wealth is most needed, , we expect the intertemporal hedging demand to be less negative for big stocks compared to small stocks.[33] Further, changes in risk aversion would affect the hedging demand for small stocks more, simply because the same difference in the coefficient of risk aversion is factored by a greater scaled covariance term. Hence, we argue that the difference between the average buying demand for small stocks would decline (increase) disproportionately more than that for big stocks when (a) marketwide risk aversion increases (declines) and/or (b) the covariance between current returns and expected returns increases (declines) disproportionately more for small stocks than for big stocks. Extant research hints that risk aversion increases (e.g. Rosenberg and Engle, 2002) and the stock returns become more procyclical (e.g. Yogo, 2006) as the economy nears a trough.[34] In light of these, we hypothesize that the order flow differential between big and small stocks should be related negatively to future economic growth and positively to expected stock returns.

How may the information in aggregate order flows not be subsumed by marketwide returns? Chan (1993) develops a model to explain the cross-autocorrelations in stock returns, which proves useful in answering this question. Subject to a noisy signal comprised of a macroeconomic and an idiosyncratic part, the market-maker cannot simultaneously assimilate other stocks' signals. The

---

[32] Using quarterly returns, dividends, and prices from CRSP over the period 1947 through 1995, Campbell and Viceira (1999) estimate this correlation as -0.74 for the U.S. stocks.

[33] Yogo (2006) shows that the returns on small stocks and value stocks are more procyclical and that the covariance of durable consumption with stock returns is higher at business cycle troughs than at peaks.

[34] For instance, the utility functions in habit persistence models (e.g. Abel, 1990; Constantinides, 1990; and Campbell and Cochrane, 1999) exhibits time-varying relative risk aversion, where relative risk aversion is a declining function of the difference between the current consumption and habit. The empirical evidence reported in Rosenberg and Engle (2002), risk aversion is countercyclical, high prior to/during recessions and low prior to/during expansions.

idiosyncratic component gets diversified when the signals are aggregated across stocks at a later date and the macroeconomic signal, which is now precise, is fully incorporated into the prices. The author uses this model to explain the positive cross-autocorrelations between stock returns. Departing from the same idea, we extend the intuition provided by this model as follows. The noisy macro signal that the market maker receives may be public or private. In the case of a noisy public signal, market makers can confirm the signals for other stocks through the financial media as well as from lagged returns. Hence, adjustment may be expected to occur relatively faster. A private macro signal embedded in order flow would be harder to decipher. As market makers do not have ready access to the each other's order flow data, the adjustment may be expected to take place over a relatively longer time period, during which aggregate order flow may contain superior information in comparison to the market return.

There is also ample evidence that certain other variables forecast economic output growth and stock returns. Fama and French (1989) find that the default and term spreads (*DEF* and *TERM*) track economic fluctuations and are useful in explaining expected stock returns. Chen (1991) shows that (a) future output growth is related negatively to *DEF* and positively to *TERM* and the excess market return (*MKT*) and (b) future market return is related positively both *DEF* and *TERM*.[35] Using data from ten countries, Liew and Vassalou (2000) study the link between future GDP growth and the returns on the market, *SMB*, *HML*, and *WML* portfolios. The authors document a positive relation between the excess market return, *SMB*, and *HML* and the GDP growth rate over the subsequent year for five of the ten countries studied. For the U.S. market, they find that the excess market return and *SMB* contain information about future GDP growth over and above the Treasury bill rate, dividend yield, term spread, and lagged production growth. Baker and Wurgler (2000) find a positive link between the equity share in new debt and equity issues and future stock returns.

[35] Fama and French (1989) demonstrate that the default spread and the dividend yield captures similar information related to security returns. We omit the dividend yield and include the default spread as a control in our analysis as the default spread is free of the price-in-denominator concern.

Pastor and Stambaugh (2003) show that a liquidity factor (*LIQ*) based on order flow-related return reversals explains the cross-sectional variation in expected stock returns in the U.S. market. Lastly, Baker and Wurgler (2006) document a significant relation between an investor sentiment index constructed from several sentiment proxies (*SENT*) and the future returns for securities whose valuations are more subjective and hard to arbitrage.

3.3 Data and Variables

Our sample comprises all common shares listed on the New York Stock Exchange (NYSE) with available data at the intersection of the Center for Research in Security Prices (CRSP) monthly return files, COMPUSTAT Industrial Annual files, and the NYSE Trades and Quotes (TAQ) or Institute for the Study of Security Markets (ISSM) databases.[36] Following Fama and French (1993), the sample is divided into two firm size categories (small or big: S or L), based on the median NYSE size, and three book-to-market (BM) categories (high, medium, or low: H, M, or L), based on the 30th and 70th BM percentiles.[37] [38] At each June-end, six portfolios are formed from the stocks at the intersection of these size and BM categories (S/L, S/M, S/H, B/L, B/M, and B/H) and the returns and order flows for these portfolios are computed for the subsequent twelve months. Hence, in order to be included in the sample in a given period (from July of year t to June of year t+1), a stock should have price and shares outstanding data from CRSP for June of year t and relevant accounting data (book value of equity as defined in Fama and French (1993)) for year t-1.

[36] We restrict our sample to common stocks trading on the NYSE in order to ensure that our results are not influenced by differences in trading protocols across venues or in trading characteristics across asset classes.

[37] Firm size and book-to-market ratio are computed as defined in Fama and French (1993). Specifically, firm size is the market capitalization (price times shares outstanding) at the beginning of the measurement period (end of June), while book-to-market is the book value of equity (shareholders' equity plus balance sheet deferred taxes and investment credit minus the book value of preferred equity) reported in the previous year's financial statement divided by the market capitalization at the beginning of the measurement period.

[38] The cutoff points for the size-BM portfolios are obtained from the personal website of Kenneth French.

The intraday trade and quote data used in the estimation of order flows come from ISSM for the period January 1988 through December 1992 and from NYSE TAQ for the period January 1993 through December 2004. Trades for all NYSE common shares are classified as either buyer- or seller-initiated using the Lee and Ready (1991) algorithm.[39] For each stock, order flow is computed quarterly as the share volume generated in buyer-initiated trades less the share volume generated in seller-initiated trades divided by the total volume over the period. The market order flow (*OFM*) is the value-weighted cross-sectional average of the individual stock order flows in a given quarter, while the order flow differential between big and small stocks (*OFD*) is the difference between the arithmetic averages of the small and big stock order flows. Both of the aggregate order flow measures are de-trended and corrected for a quarterly seasonal before being included as explanatory variables in our formal regression models.

The quarterly returns and market capitalizations for the value-weighted market and capitalization decile indices are obtained from CRSP. The market premium (*MKT*) is computed as the excess return on the value-weighted market index over the one-month Treasury bill rate obtained from Ibbotson Associates. The new equity added to the index (*NEQ*) is computed as the percentage increase (from quarter t-1 to quarter t) in the total market capitalization of the firms comprising the value-weighted market index, less the value-weighted market return. The size, value, and momentum (*SMB*, *HML*, and *WML*) premiums as well as the Pastor and Stambaugh (2003) marketwide liquidity measure (*LIQ*) are obtained from the Fama-French, Momentum, and Liquidity database provided by the Wharton Research Data Services. *SMB* is the difference between the average return for the three small stock portfolios and the average return for the three big stock portfolios. *HML* is the difference between the average return for the two high BM stock portfolios and the average return for the two low BM portfolios.

[39] Each trade is matched with the first quote occurring at least five seconds prior to the trade. The trade is classified as a buy (sell) if it occurs above (below) the prevailing quote midpoint. If the trade occurs exactly at the quote midpoint, the tick-test is applied, and the trade is classified as a buy (sell) if it results in a positive (negative) price change.

WML is the difference between the average return for the two "winner" stock portfolios and the average return for the two "loser" stock portfolios.[40] *LIQ* is the cross-sectional average of the individual stock liquidity measure (multiplied by 10^2) in Pastor and Stambaugh (2003), which measures the strength of the return reversal in month *t*+1 associated with the signed trading volume in month *t*. *SMB*, *HML*, and *WML* are defined as the geometric average of the monthly values within a quarter. For *LIQ*, this transformation is done by taking the arithmetic average of the monthly values in a given quarter.

We obtain experts' earnings per share (EPS) forecasts and actual figures reported by firms from the Institutional Brokers Estimates System (I/B/E/S) summary file. The cross-sectional averages of the one-quarter-ahead forecasts and actual values of EPS are computed at the end of each quarter. The forecasted corporate earnings growth (*FEG*) is computed as the percentage change in average EPS implied by the one-quarter-ahead forecasts. The quarterly corporate earnings growth (*QEG*), on the other hand, is the actual percentage increase in EPS from quarter *t*-1 to quarter *t*. The quarterly industrial production and real GDP growth rates (*QPG* and *QYG*) are defined as the percentage changes in the industrial production index and per-capita real GDP, respectively. The data for industrial production, real GDP, and interest rates (for both corporate and government securities) come from the St. Louis Fed database (FRED). The default spread (*DEF*) is defined as the difference between Moody's seasoned Baa-and Aaa-grade corporate bond portfolio yields, while the term spread (*TERM*) is the difference between the 10-year Treasury constant maturity rate and the 3-month Treasury bill rate. Lastly, the investor sentiment index (*SENT*) is from the personal website of Jeffrey Wurgler.[41] Our tests reveal that *SENT* is a unit-root process and, thus, it is first-differenced to rid the analysis of the econometric issues that may arise.

[40] At the beginning of each month *t*, stocks are sorted based on the total return between *t*-2 and *t*-12. A winner (loser) is a stock that belongs to the top (bottom) 30 percent of this return distribution.

[41] This composite sentiment index is estimated by Baker and Wurgler (2006) as the first principal component of six sentiment proxies: the closed-end fund discount, NYSE share turnover, the number of IPOs and their

3.3.1 Descriptive Statistics

The means, medians, standard deviations, and the minimum and maximum values of the variables under study are reported in Table 2. Over the sample period, *OFM* ranges between -0.9 percent observed in the third quarter of 1990 (1990/3) and 11.4 percent observed in the first quarter of 1998 (1998/1). The mean (median) for *OFM* is 7.2 (7.7) percent with a standard deviation of 2.9 percent.[42] This significantly positive average mean mostly reflects the high net buying pressure for big stocks and is documented in other studies (e.g. Chordia, Roll, and Subrahmanyam, 2002). Indeed, the difference between big stock and small stock portfolios (*OFD*) varies between 0.9 percent (2002/2) and 11.6 percent (1990/4) with a significantly positive mean (median) of 5.6 (5.3) percent and a standard deviation of 2.4 percent.

Looking at the four return factors, we observe that *MKT* ranges between -6.0 percent (1998/4) and 6.3 percent (2002/2), with a mean (median) of 0.6 (1.0) percent and a standard deviation of 2.7 percent.[43] *SMB* ranges varies between -3.6 percent (1993/3) and 4.1 percent (2001/4), with a mean/median of 0.1 percent and a standard deviation of 1.8 percent. *HML* ranges between -6.8 percent (1999/4) and 7.9 percent (2000/4), with a mean (median) of 0.3 (0.2) and a standard deviation of 2.2 percent. *WML* varies between -7.2 percent (2003/2) and 8.0 percent (1999/4), with a mean (median) of 0.8 (0.7) percent and a standard deviation of 2.6 percent.

The mean (median) values for *DEF* and *TERM* are 0.9 (0.8) percent and 1.7 (1.6) percent over the sample period. *DEF* ranges between 0.6 (1999/4) and 1.4

---

average first day returns, the equity share in new issues, and the dividend premium. Each sentiment proxy is orthogonalized with respect to business cycle effects.

[42] The most viable explanation is that the excess buying pressure observed for market orders is absorbed by an offsetting selling pressure in the limit orders. Consistent with this, the average portfolio order flow is 0.80 percent for small stocks where the limit order books are presumably thinner. The fact that our sample period largely corresponds with the extended bull market of 90s may also add to the explanation.

[43] This corresponds to an annual equity premium of about 7.7 percent.

(1990/4) percent, with a standard deviation of 0.2. *TERM* varies between -0.6 (2000/3) and 3.7 (1992/3) percent, with a standard deviation of 1.1 percent. The mean (median) forecasted corporate earnings growth is 0.6 (0.9) percent. *FEG* ranges between -10.1 (1998/1) and 8.5 (1999/2) percent, with a standard deviation of 4.73 percent. The average and median values for new issues, *NEQ*, are both 0.4 percent, with a standard deviation of 0.6 percent. The greatest expansion (contraction) in new issues is a 1.8 (1.7) percent increase (decline) observed in the second quarter of 2002 (fourth quarter of 1988). *LIQ* ranges between -0.15 (2002/4) and 0.03 (1992/2), with a standard deviation of 0.03 and a mean (median) of -0.02 (-0.01). *ΔSENT* has a mean of zero and a standard deviation of 0.4. Finally, U.S. industrial production and real GDP both grow at a quarterly rate of 0.7 percent over our sample period. *QPG* ranges between -2.3 (1990/4) and 3.0 percent (1997/3), with a standard deviation of 1.1 percent, while *QYG* varies between -0.8 (1990/3) and 1.8 percent (2003/2), with a standard deviation of 0.5 percent.

3.3.2 Correlations

The contemporaneous correlations between the explanatory variables are presented in Table 3. Starting with the order flow variables, we see that *OFM* and *OFD* are negatively correlated with a correlation coefficient (ρ) of -0.12. As would be expected (since the first-differences are defined from quarter *t*-1 to quarter *t*), the correlation between the levels and first-differences of both of these variables is in the order of 0.50. The facts that *OFM* is correlated positively and significantly with *MKT* (ρ= 0.36) and negatively and significantly with *HML* (ρ= -0.19), *TERM* (ρ= -0.30), and *NEQ* (ρ= -0.18) indicate that buying pressure in the stock market tends to be high when (a) the excess market return is high, (b) growth stocks yield higher returns relative to value stocks, (c) the yield curve is steeper, and (d) more new equity is added to the market. *OFD*, on the other hand, is correlated positively and significantly with *WML* (ρ= 0.23) and *DEF* (ρ= 0.28) and negatively and significantly with *MKT* (ρ= -0.29), *SMB* (ρ= -0.32), *TERM* (ρ= -0.44), *NEQ* (ρ= -0.30) and *LIQ* (ρ= -0.31). These suggest that the buying pressure for big stocks increases relative to that for small stocks when (a) the excess market return is low, (b) small stocks yield lower returns relative to big stocks, (c) winner stocks yield higher returns relative to loser stocks, (d) default risk is higher, (e) the yield curve is steeper, (f) less new equity is added to the index, and (g) liquidity is low.

Turning to the return factors, we see that *MKT* is correlated positively and significantly with *SMB*, while both *MKT* and *SMB* are correlated negatively and significantly with *HML* and *WML*. The excess market return tends to be low when default risk increases and high when liquidity improves. Periods with higher liquidity also tend to have lower value stock and greater small stock returns, as indicated by the negative (positive) correlation between *HML* (*SMB*) and *LIQ*. The positive (negative) correlation between *SMB* (*WML*) and *TERM*, on the other hand, tells us that small stocks and loser stocks perform better in periods that see an increase in the slope of the yield curve. Lastly, an increase in investor sentiment appears to be associated with greater returns on value stocks, as seen in the positive correlation between *ΔSENT* and *HML*, which is in line with the behavioral connotations attached to the value premium (e.g. Lakonishok, Shleifer, and Vishny, 1994).

As for the remaining explanatory variables, we observe a positive correlation between *DEF* and *TERM* on the order of 0.29. Marketwide liquidity tends to be high when default risk is low and forecasted corporate earnings growth and new equity additions are high. The positive correlations between *NEQ* and *TERM*, *LIQ*, and *ΔSENT* tells us that new equity is more likely to be added to the index when the yield curve is steep, market is liquid, and investor sentiment is bullish, while the negative correlation between *DEF* and *NEQ* indicates that high default risk deters new equity infusions.

3.4 Aggregate Order Flows and Future Economic Output Growth

3.4.1 Univariate Regressions

Panels A to C of Table 4 report the coefficient estimates, t-statistics, and R^2 values from the univariate regressions of future quarterly growth rates for real GDP, industrial production, and corporate earnings (for quarters t+1 through t+4) on each of the explanatory variables. Starting with the order flow aggregates, we see that an increase in average buying pressure in the stock market signals an increase in the industrial production and real GDP growth, but has no association with future corporate earnings growth. In economic terms, a one standard deviation increase in *OFM* predicts an increase of 0.18 to 0.35 (0.21 to 0.38) standard deviations in *QYG* (*QPG*) over the four subsequent quarters, explaining 3 to 13 percent (5 to 15 percent) of the variation in this variable. A higher-than-average *OFD*, on the other hand, signals a significant decline in the real GDP, industrial production, and corporate earnings growth, and does so strongly. A one standard deviation increase in *OFD* predicts a decline of 0.26 to 0.39 standard deviations in *QYG*, 0.33 to 0.48 standard deviations in *QPG*, and 0.21 to 0.31 standard deviations in *QEG*. The total variation explained by *OFD* ranges between 11 and15 percent for industrial production, 7 and15 percent for real GDP, and 10 and 21 percent for corporate earnings.

Looking at the common return factors, we see that a higher-than-average excess market return tends to forecast above-average economic growth, consistent with the findings reported in Chen (1991). A one standard deviation increase in *MKT* predicts an increase of 0.35 standard deviations in *QYG* in the following quarter and increases of 0.17 to 0.36 standard deviations in *QPG* and of 0.04 to 0.16 standard deviations in *QEG* in the subsequent quarters. *SMB* is related positively to only corporate earnings growth, with a one standard deviation increase in this variable forecasting a 0.16 standard deviation increase in *QEG* in the quarter that immediately follows the order flow observation. *HML* is related negatively to industrial production growth in the subsequent quarter, though not to *QYG* or

QEG, while *WML* does not seem to contain any information about future economic growth. The maximum amount of variation in future economic growth explained by any of the return factors is 12 percent for industrial production and real GDP and 6 percent for corporate earnings, significantly less than the explanatory power of the order flow aggregates.

As for the to the remaining controls, future industrial production growth is related positively to *NEQ* and *LIQ* and negatively to *DEF*, indicating that an increase in new equity, a more liquid market, and lower default risk are precursors to rapid production growth. Future real GDP growth is related positively to *NEQ* and negatively to *DEF* as well, though not to *LIQ*. Future values of *QEG* are related positively to *FEG*, *NEQ*, and *TERM*, suggesting that corporate earnings grow faster when forecasted earnings growth is high, when more new equity is added, and when the yield curve is steeper. In economic terms, a one standard deviation increase in *NEQ* forecasts an increase of 0.24 to 0.39 standard deviations in *QPG*, of 0.14 to 0.28 standard deviations in *QYG*, and of 0.09 to 0.27 standard deviations in *QEG*. A one standard deviation increase in *DEF* predicts a decline of 0.21 to 0.34 standard deviations in *QPG* and of 0.05 to 0.27 standard deviations in *QYG*, while a one standard deviation increase in *TERM* forecasts an increase of 0.20 to 0.35 standard deviations in *QEG*. Lastly, a one standard deviation increase in *LIQ* sees industrial production growth increase by 0.29 to 0.35 standard deviations in the two subsequent quarters.

3.4.2 Multivariate Regressions

Panels A to C of Table 5 report the slope estimates and Newey-West (1987) adjusted t-statistics for *OFM* and *OFD*, as well as the model R^2 values from predictive regressions of the three economic growth proxies on the order flow measures and controls. The first column in each panel shows the results for models where the only explanatory variables are the two order factors, while the second column adds the common return factors (*MKT*, *SMB*, *HML*, and *WML*) and lagged economic growth rate as controls.

The first thing to note is that *OFM* and *OFD* do not subsume each other's predictive power. In models where the two factors are used together the model R^2 roughly equals the sum of the model R^2 values from the univariate regressions, indicating that the order flow differential provides information that is different from what is contained in the market order flow. The coefficient estimates for *OFM* and *OFD* decline slightly, but retain their signs and statistical significance; therefore, our conclusions from the previous section are maintained: future output growth is related positively to market order flow and negatively to the order flow differential between big and small stock portfolios. The amount of variation explained by the two order flow aggregates ranges from 14 to 30 percent for industrial production growth, 10 to 24 percent for real GDP growth, and 11 to 24 percent for corporate earnings growth over the four subsequent quarters. Recall, however, that these are models in which the information in returns is not accounted for. One might argue that the excess market return should capture all the information relevant to the stock market, including future expected changes in industrial production, real GDP, and corporate earnings growth. Hence, in the second stage of our multivariate analysis, we include the excess market return (*MKT*) in order to take this argument into account and observe whether and to what extent the predictive power of the two order flow aggregates is subsumed. Rolling the dice against ourselves, we also include the contemporaneous realizations of three empirical return factors related to size, value, and momentum (*SMB*, *HML*, and *WML*).

The results from the full model are reported in the second column of each panel in Table 5. The slope coefficients for the two order flow variables decline in certain quarters and increase in others. The amount of variation explained by the models increases to about 28 to 35 percent for industrial production growth, 14 to 29 percent for real GDP growth, and 18 to 27 percent for corporate earnings growth.

However, we do not observe any change in the significances of *OFM* and *OFD* when we control for the own lag of the dependent variable and the four return factors. The finding of a strong relation between the order flow aggregates and future output growth in the presence of controls for returns extends evidence from the bond and foreign exchange that marketwide measures of order flow contain information that is private in the sense that it is only incorporated into prices after a protracted lag.

Collectively, the evidence reported in this section is consistent with an information aggregation role for marketwide measures of order flow. The positive relation between market order flow and future economic growth parallels evidence from other markets and is plausible if investors allocate funds to stocks as a class on the basis of their expectations regarding the future performance of the economy. The negative relation between the order flow differential and future economic growth supports our thesis that *OFD* captures time variation in investors' intertemporal hedging demand, which is induced by the strategic behavior of informed investors who wish to hedge against adverse changes in investment opportunities.[44] That is, informed investors who wish to hedge their exposure to adverse wealth shocks would rationally want to tilt their portfolios towards big stocks, whose returns are much less procyclical than those of small stocks, when they expect economic conditions to worsen.[45] In the presence of such behavior, the order flow differential between big stocks and small stocks will be negatively related to future macroeconomic fundamentals.

The finding that the signal in the two order flow aggregates is subsumed by neither the excess market return nor the three empirical return factors (*SMB*, *HML*, and *WML*) is intriguing. This result can be explained by the existence of (i) investors who are endowed with superior information about macroeconomic

[44] A significant fraction (20% to 50%) of the demand for stocks by long-lived risk-averse investors is due to intertemporal hedging motives (Campbell and Viceira, 1999) and this demand can be shown to be greater for higher levels of risk aversion and lower for securities whose returns are more procyclical (Restoy, 1992).

[45] See Yogo (2006), for instance, for formal research evidence supporting this argument.

fundamentals or (ii) investors that are endowed with private information at the firm level that is correlated across firms. In either case, our evidence suggests that the macroeconomic information in aggregate order flows is not incorporated into stock returns immediately. This interpretation, in turn, suggests the hypothesis that the order flow aggregates should predict stock returns. The next section addresses this hypothesis.

3.5 Aggregate Order Flows and Expected Stock Returns

In the previous section, we provided evidence that the aggregate market order flow and the order flow differential contain a signal about future economic output growth that is incremental to the information contained in the return factors. A potential explanation for why returns may not capture all the information in the order flow aggregates is suggested by the work of Chan (1993). The author presents a model in which market-makers receive a noisy macroeconomic signal (blurred by firm-specific noise), which cannot be incorporated into prices immediately. The noise thus induced in returns may not get washed away when returns are averaged across the market if this noise is correlated across market-makers. Over time, however, the signal becomes precise as market-makers observe the signals for other stocks in the market, and prices are adjusted accordingly. The results in the previous section, therefore, motivate the hypothesis that, if aggregate order flows contain a macro signal that is not impounded into the prices immediately, the order flow aggregates may predict stock returns.

Our initial test of this hypothesis relates the returns for the size-sorted decile portfolios, the market, size, value, and momentum premiums (*MKT*, *SMB*, *HML*, and *WML*), and the change in the 3-month Treasury bill rate (ΔTB) in quarter t+1 to changes in *OFM* and *OFD* from quarter t-1 to quarter t, which we term ΔOFM

and *ΔOFD*.[46] The slope estimates and Newey-West (1987) adjusted t-statistics for *ΔOFM* and *ΔOFD*, and the model R^2 values are reported in Tables 6 and 7. As before, we introduce our control variables in stages. The first column reports the results from models where *ΔOFM* and *ΔOFD* are the only regressors. The return factors are added as controls in the second column, while the third column presents estimates from the full model, which includes the default and term spreads (*DEF* and *TERM*), new equity (*NEQ*), forecasted earnings growth (*FEG*), Pastor-Stambaugh (2003) liquidity measure (*LIQ*), quarterly change in the Baker and Wurgler (2006) investor sentiment index, and lagged portfolio return.

The results from the regressions of the decile portfolio returns are given in Table 6. In models where only the order flow aggregates are used as predictors (first column), *ΔOFM* has significant forecast power for six out of ten portfolio returns, with the exception of the smallest and the three largest portfolios, while *ΔOFD* is related strongly to the future returns of all decile portfolios. A one percent increase in *ΔOFM* (*ΔOFD*) predicts the decile portfolio returns to be 0.21 to 1.37 (1.10 to 2.09) percent higher in the subsequent quarter. The slope estimates for both *ΔOFM* and *ΔOFD* are consistently positive for all decile portfolios and tend to decline as we go from small cap portfolios to large cap portfolios. The amount of variation in decile portfolio returns explained ranges from 10 to 19 percent in these models.

The amount of variation explained increases to between 15 and 21 percent when the quarter t realizations of the return factors are added to the model (second column) and to between 29 and 39 percent when the default and term spreads, new equity, earnings growth forecasts, marketwide liquidity, and investor sentiment are added (third column). The coefficient estimates for *ΔOFM* decline in magnitude and become insignificant for all decile portfolios except one when the return factors are added as controls and all predictive power is lost in the full

[46] Any relevant information that is to be extracted should be extracted from the unexpected component of trades. We use the first-differences of the two order flow aggregates as a rough proxy for the innovations in these variables. Our results do not change significantly when a first-order autoregressive model is used to strip the order flows from their expected component.

model. The predictive power of *ΔOFD*, on the other hand, is robust to the inclusion of all the controls. The coefficient on *ΔOFD* increases in magnitude for the smaller deciles and declines slightly for larger deciles when the return factors are controlled for, and increases for almost all the deciles when the rest of the controls are added to the model.

Table 7 presents the results from the regressions of the quarter t+1 values of *MKT*, *SMB*, *HML*, *WML*, and *ΔTB* on *ΔOFM* and *ΔOFD*. We observe that, regardless of the set of control variables used in out models, *ΔOFD* is related positively and significantly to the following quarter's excess market return and *SMB*, while the relation between *ΔOFM* and the future return premiums is flat. The model that contains only the two order flow aggregates explains 10 percent of the variation in the excess market return and 8 percent of the variation in the size premium for the subsequent quarter. The amount of variation explained increases to 14 percent for *MKT* and 10 percent for *SMB* when the four return factors are added to the model and to 30 percent for *MKT* and 33 percent for *SMB* when the remaining controls are included. According to the coefficients from the full model, a one percent increase in *ΔOFD* predicts the monthly excess market return (monthly size premium) to be 0.40 percent (0.44 percent) higher on average in the subsequent quarter. The order flow aggregates do not predict the future realizations of either *HML* or *WML*.

The evidence in this section is consistent with a world where marketwide order flows reflect information about economic fundamentals that is dispersed across agents. As in Evans and Lyons (2009), our results suggest that this information is *private* in the sense that it is not incorporated into asset prices instantaneously. The positive relation between market order flow and decile portfolio returns is consistent with the noisy macroeconomic signal story along the lines of Chan (1993). The finding that market order flow does not have significant predictive power for the returns of larger stocks is plausible if the noise in the macro signal is diversified to a certain extent because of the greater scale of such firms' operations. As in Albuquerque et al. (2008), we find that the explanatory power of *OFM* is subsumed when liquidity and other controls are added to the model. This is consistent with the conclusion of the authors that a simple statistical factor constructed as the average of stock order flows captures mostly liquidity. The strong and robust forecast power of the order flow differential for both economic growth and stock returns supports our thesis that this variable captures time-variation in market's risk aversion.

3.6 Is It Just Liquidity?

As a robustness test, we want to examine whether and how much the results that we obtained are related to time-variation in liquidity. In order to do this, we run a two-way sort of the sample on size and liquidity (as measured by the annually-estimated Amihud (2001) illiquidity measure), classifying stocks as small or big (S or B) and liquid or illiquid (L or I) at the end of every June. We measure the quarterly order flows for the resulting portfolios over our sample period and form two order flow differentials. The first is the difference between the two big stock portfolios (B/L and B/I) and the two big stock portfolios (S/L and S/I). The second is the difference between the two liquid stock portfolios (S/L and B/L) and the two illiquid stock portfolios (S/I and B/I). We label these two measures as *OFDS* and *OFDL* respectively and repeat the analysis that we conducted for *OFM* and *OFD* once again, after substituting *OFDS* and *OFDL* in place of *OFD*, controlling for the same set of variables.[47]

The results from the regressions of the economic growth proxies on this extended set of order flow aggregates are given in Table 8. Controlling for the information in returns, both *OFDS* and *OFDL* display significant forecast powers for the real GDP and industrial production growth rates over the subsequent four quarters.

[47] *OFDS* and *OFDM* are both negatively correlated with *OFM*, with correlation coefficients of -0.22 and -0.16. The correlation between *OFDS* and *OFDL* is -0.26, indicating that our two-way sort is successful in disentangling size and liquidity effects.

OFM is insignificant. In economic terms, a one standard deviation increase in *OFDS* predicts a decline of about 0.28 to 0.52 percent (0.26 to 0.48 standard deviations) in *QPG* and 0.09 to 0.25 percent (0.17 to 0.49 standard deviations) in *QYG*. Similarly, a one standard deviation increase in *OFDL* signals to a decline of about 0.26 to 0.33 percent (0.25 to 0.32 standard deviations) in *QPG* and 0.07 to 0.17 percent (0.14 to 0.33 standard deviations) in *QYG*. For corporate earnings growth, *OFDS* dominates both *OFM* and *OFDL*. The coefficient estimates for *OFDL* are insignificant except for the quarter that immediately follows the order flow observation. A one standard deviation increase in *OFDS* forecasts a statistically significant decline between 1.35 and 1.84 percent (0.38 and 0.52 standard deviations) in *QEG* in the next four quarters, while a similar increase in *OFDL* predicts a decline of 0.91 percent (0.25 standard deviations) in the subsequent quarter.

Table 9 presents our results from the regressions of the decile portfolio returns on *OFDS* and *OFDL*, where the full set of control variables is employed. We see that *ΔOFDS* forecasts the subsequent quarter's returns for all ten deciles, while *ΔOFDL* is only significant for the small stock portfolios. The effects are economically significant: a one standard deviation increment to *ΔOFDL* forecasts the next quarter's return to be 0.23-0.29 standard deviations higher for the smaller five of the ten decile portfolios. The positive relation between *ΔOFDL* and expected returns of small stocks is consistent with the existence of an illiquidity premium for these stocks during periods of flight-to-liquidity. A one standard deviation increment to *ΔOFDS*, on the other hand, predicts the returns to be higher by 0.26-0.50 standard deviations for all decile portfolios in the subsequent quarter. The positive relation between *ΔOFDS* and expected returns is not specific to small stocks and is supportive of the hedging/risk aversion-related role that we assigned to this variable.

3.7 Concluding Remarks

In this paper, we showed that the current values of two aggregate equity market order flow variables, the market order flow and the order flow differential between big stock and small stock portfolios, are related significantly to future economic output growth, as measured by the quarterly growth rates in U.S. industrial production, real GDP, and corporate earnings growth over the period January 1988 through December 2004. The first of these relations, the positive link between the market order flow and future economic growth, parallels the evidence from foreign exchange (e.g. Evans and Lyons, 2009) and government bond markets (e.g. Green, 2004). The second relation, the negative link between order flow differential and future economic growth, points to a role for *OFD* based on intertemporal hedging demand. We argue that, in a market with heterogeneously informed investors, the buying pressure for large stocks will be greater than the buying pressure for small stocks when investors expect economic conditions to deteriorate, since then the informed hedgers are going to increase their portfolio allocations for stocks that are better hedges (i.e. whose returns are less procyclical) in order to hedge their exposure to adverse wealth shocks. Since big stocks have less procyclical returns compared to small stocks, the order flow differential is going to be higher prior to economic downturns. This argument is reinforced by the considerable anecdotal evidence that big stocks are perceived as safe havens that investors flock to when the economy turns down.

The finding that the macro signal in aggregate order flows is not subsumed by common return factors –in particular, the excess market return— is striking. A plausible explanation, suggested by the work of Chan (1993), is that market-makers cannot separate the macro and stock-specific signals conveyed by order flow for their own stocks in a timely manner. The macro signal becomes precise over time as market-makers observe the signals for other stocks. This is the stage at which market-makers adjust their prices to reflect macroeconomic information. Our thesis is that, if market-makers fail as a group to read through the noise in their own firm-specific signals, prices may not contain all the information embedded in aggregate order flow. If this is the case, aggregate order flow should predict stock market returns.

Our evidence from the regressions of future stock returns on aggregate order flows and controls supports this argument. The positive relation between market order flow and future small cap decile portfolio returns is in line with the noisy macro signal story. The fact that the returns for large cap portfolios cannot be predicted using market order flow is plausible if the sheer scale of such firms' operations render the micro signal more precise. Consistent with the conclusion of Albuquerque et al. (2008) that a simple statistical factor constructed from order flows mainly captures liquidity, this relation is subsumed when business cycle indicators and market liquidity are added to the model. The positive relation between the order flow differential and expected stock returns is strong and intuitively appealing. In particular, intertemporal hedging demand is greater when risk aversion is higher and the effect is more pronounced for assets with more procyclical returns. Hence, an increase in the portfolio allocation for big stocks prior to economic downturns can be viewed as a consequence of an increase in marketwide risk aversion during such states. An increase in risk aversion also implies an increase in expected returns. This is precisely what our evidence linking returns and the order flow differential seems to tell us.

Lastly, it is intriguing that the return factors, including the excess market return and *SMB* —which are closely related to *OFM* and *OFD*—do not subsume the macroeconomic signal contained in the order flow measures. Future research may focus on the relation between *OFD* and *SMB*, as these two variables are the average order flow and the average return (with an inverted sign) on essentially the same portfolio. Our argument regarding the correspondence between marketwide risk aversion and the order flow differential may prove useful in shedding light on the origins of the size premium.

3.8 References

Amihud, Yakov (2002), “Illiquidity and Stock Returns: Cross-Section and Time-Series Effects,” *Journal of Financial Markets 5*, 31-56

Albuquerque, Rui, Eva de Francisco, and Luis Marques (2008), “Marketwide Private Information in Stocks: Forecasting Currency Returns,” *Journal of Finance 63*, 2297-2343

Baker, Malcolm and Jeffrey Wurgler (2000), “The Equity Share in New Issues and Aggregate Stock Returns,” *Journal of Finance* 55, 2219-2257

Baker, Malcolm and Jeffrey Wurgler (2006), “Investor Sentiment and the Cross-Section of Stock Returns,” *Journal of Finance 61*, 1645-1680

Beber, Alessandro, Michael W. Brandt, and Kenneth A. Kavajecz (2009), “What Can Equity Orderflow Tell Us About the Economy?” *Working Paper*

Brandt, Michael W. and Kenneth A. Kavajecz (2004), “Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve,” *Journal of Finance 59*, 2623-2654

Campbell, John Y. And John H. Cochrane (1999), “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy 107*, 205-251

Campbell, John Y. and Luis M. Viceira (1999), “Consumption and Portfolio Decisions when Expected Returns are Time-Varying,” *Quarterly Journal of Economics 114*, 433-495

Chan, Kalok (1993), “Imperfect Information and Cross-Autocorrelation among Stock Prices,” *Journal of Finance 48*, 1211-1230

Chen, Nai-Fu (1991), “Financial Investment Opportunities and the Macroeconomy,” *Journal of Finance 46*, 529-554

Chordia, Tarun, Richard Roll, Avanidhar Subrahmanyam (2002), “Order Imbalance, Liquidity, and Market Returns,” *Journal of Financial Economic 65*, 111-130

Epstein, Larry G. and Stanley E. Zin (1989), “Substitution, Risk Aversion, and Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica* 57, 937-969

Evans, Martin D. D., and Richard K. Lyons (2009), “Exchange Rate Fundamentals and Order Flow,” *Working Paper*, University of California Berkeley, Berkeley, California

Fama, Eugene A. and Kenneth R. French (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics 25*, 23-49

Fama, Eugene A. and Kenneth R. French (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics 33*, 3-56

Green, T. Clifton (2004), "Economic News and the Impact on Bond Prices," *Journal of Finance 56*, 1201-1233

Kyle, Albert S., 1985, Continuous Auctions and Insider Trading, *Econometrica 53*, 1315-1336

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny (1994), "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance 49*, 1541-1578

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

Liew, Jimmy and Maria Vassalou (2000), "Can Book-to-Market, Size, and Momentum Be Risk Factors That Predict Economic Growth?" *Journal of Financial Economics 57*, 221-245

Lo, Andrew W. and A. Craig MacKinlay (1990), "When Are Contratian Profits Due to Stock Market Overreaction," *Review of Financial Studies 3*, 175-205

Newey, Whitney K. and Kenneth D. West (1987) "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica 55*, 703-708

Pasquariello, Paolo and Clara Vega (2007), "Informed and Strategic Order Flow in the Bond Markets," *Review of Financial Studies 20*, 1975-2019

P**á**stor, Ľuboš and Robert F. Stambaugh (2003), "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy 111*, 642-685

Rosenberg, Joshua V. and Robert F. Engle (2002), "Empirical Pricing Kernels," *Journal of Financial Economics 64*, 341-372

Spiegel, Matthew and Avanidhar Subrahmanyam (1992), "Informed Speculation and Hedging in a Noncompetitive Securities Market," *Review of Financial Studies 5*, 307-329

Yogo, Motohiro (2006), "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance 61*, 539-580

3.9 Tables

Table 3-1 Variable Definitions

Variable	Notation	Definition
Marketwide Order Flow	*OFM*	The value-weighted cross-sectional average of the quarterly order flows for individual stocks
Order Flow Differential	*OFD*	The difference between the average order flows for the three big stock portfolios and three small stock portfolios
Excess Market Return	*MKT*	The excess return on the value-weighted market portfolio over the one-month Treasury bill rate
Size Premium	*SMB*	The difference between the average returns for the three small stock portfolios and the three big stock portfolios
Value Premium	*HML*	The difference between the average returns for the three value portfolios and the three growth portfolios
Momentum Premium	*WML*	The difference between the average returns for the three winner portfolios and the three loser portfolios
Default Spread	*DEF*	The difference between the yields on Moody's Baa and Aaa Grade Seasoned Bond Portfolios
Term Spread	*TERM*	The difference between the 10-year Treasury Constant Maturity rate and 3-month Treasury bill rate
Forecasted Earnings Growth	*FEG*	The cross-sectional average of quarterly earnings growth forecast by financial analysts
New Equity Additions	*NEQ*	The growth rate of the total market capitalization of stocks included in the market index less the market return
Marketwide Liquidity	*LIQ*	Pastor and Stambaugh (2003) Liquidity Measure
Investor Sentiment	$\Delta SENT$	Baker and Wurgler (2006) Investor Sentiment Measure

Table 3-2 Descriptive Statistics

This table presents the time-series means, medians, standard deviations and the minimum and maximum values for the quarterly levels and changes of the market order flow and the order flow differential (*OFM*, *OFD*, *ΔOFM*, and *ΔOFD*), the excess market return (*MKT*), the returns for the small-minus-big, high-minus-low, and winners-minus-losers portfolios (*SMB*, *HML*, and *WML*), the default and term spreads (*DEF* and *TERM*), forecasted corporate earnings growth rate (*FEG*), new equity additions to the market index (*FEG*), market liquidity (*LIQ*), the quarterly change in the investor sentiment index (*ΔSENT*), and the quarterly growth rates in industrial production, real GDP, and corporate earnings (*QPG*, *QYG*, and *QEG*). The variables are as defined in Table 1.

	Mean	Median	Std Dev	Min	Max
OFM	0.0716	0.0771	0.0292	-0.0087	0.1135
OFD	0.0556	0.0534	0.0237	0.0087	0.1158
ΔOFM	0.0009	0.0019	0.0216	-0.0541	0.0604
ΔOFD	-0.0003	-0.0002	0.0221	-0.0648	0.0420
MKT	0.0062	0.0102	0.0269	-0.0603	0.0631
SMB	0.0012	0.0009	0.0182	-0.0363	0.0407
HML	0.0032	0.0016	0.0216	-0.0678	0.0786
WML	0.0080	0.0070	0.0264	-0.0721	0.0800
DEF	0.0085	0.0081	0.0022	0.0055	0.0141
TERM	0.0174	0.0162	0.0119	-0.0055	0.0366
FEG	0.0063	0.0092	0.0473	-0.1009	0.0853
NEQ	0.0039	0.0042	0.0063	-0.0175	0.0178
LIQ	-0.0209	-0.0116	0.0344	-0.1456	0.0333
ΔSENT	-0.0010	0.0000	0.4175	-1.0800	0.9800
QPG	0.0068	0.0078	0.0108	-0.0235	0.0304
QYG	0.0073	0.0073	0.0052	-0.0076	0.0182
QEG	0.0026	0.0089	0.0527	-0.1231	0.0971

Table 3-3 Correlation Matrix

This table presents the contemporaneous correlations between the quarterly levels and changes of the market order flow and the order flow differential (*OFM*, *OFD*, *ΔOFM*, and *ΔOFD*), the excess market return (*MKT*), the returns for the small-minus-big, high-minus-low, and winners-minus-losers portfolios (*SMB*, *HML*, and *WML*), the default and term spreads (*DEF* and *TERM*), forecasted corporate earnings growth rate (*FEG*), new equity additions to the market index (*FEG*), market liquidity (*LIQ*), the quarterly change in the investor sentiment index (*ΔSENT*), and the quarterly growth rates in industrial production, real GDP, and corporate earnings (*QPG*, *QYG*, and *QEG*). Bold entries signify statistical significance at the 5 percent level. The variables are as defined in Table 1.

	OFM	OFD	ΔOFM	ΔOFD	MKT	SMB	HML	WML	DEF	TERM	FEG	NEQ	LIQ
OFM	1.000	-0.120	**0.524**	-0.008									
OFD	-0.120	1.000	0.066	**0.486**									
ΔOFM	**0.524**	0.066	1.000	0.047									
ΔOFD	-0.008	**0.486**	0.047	1.000									
MKT	**0.357**	**-0.287**	**0.249**	**-0.236**	1.000	**0.412**	**-0.508**	**-0.303**					
SMB	0.032	**-0.321**	0.052	**-0.549**	**0.412**	1.000	**-0.194**	**-0.363**					
HML	**-0.192**	-0.109	-0.001	-0.033	**-0.508**	**-0.194**	1.000	-0.085					
WML	-0.128	**0.230**	**-0.257**	**0.318**	**-0.303**	**-0.363**	-0.085	1.000					
DEF	-0.089	**0.279**	0.082	-0.009	**-0.150**	0.118	-0.034	-0.115	1.000	**0.285**	-0.133	**-0.208**	**-0.214**
TERM	**-0.302**	**-0.440**	-0.031	-0.137	0.022	**0.355**	0.079	**-0.203**	**0.285**	1.000	-0.009	**0.400**	0.117
FEG	0.019	-0.145	0.037	-0.060	-0.052	-0.113	-0.027	0.126	-0.133	-0.009	1.000	0.101	**0.203**
NEQ	**-0.182**	**-0.302**	**-0.241**	-0.146	0.027	**0.163**	0.054	0.125	**-0.208**	**0.400**	0.101	1.000	**0.217**
LIQ	0.065	**-0.308**	-0.017	-0.112	**0.379**	**0.205**	**-0.278**	0.043	**-0.214**	0.117	**0.203**	**0.217**	1.000
ΔSENT	-0.074	-0.084	-0.121	**-0.170**	-0.029	0.117	**0.234**	-0.092	-0.112	-0.047	-0.036	**0.183**	-0.008

Table 3-4 Univariate Predictive Regressions for Economic Growth

This table presents the coefficient estimates, t-statistics (in parentheses below), and R^2 values (in smaller font underneath) from the univariate regressions of quarterly industrial production, real GDP, and corporate earnings growth (*QPG*, *QYG*, and *QEG*) up to four quarters ahead on the market order flow and the order flow differential (*OFM* and *OFD*), the excess market return (*MKT*), the returns for the small-minus-big, high-minus-low, and winners-minus-losers portfolios (*SMB*, *HML*, and *WML*), forecasted corporate earnings growth rate (*FEG*), new equity additions to the market index (*NEQ*), the default and term spreads (*DEF* and *TERM*), and market liquidity (*LIQ*). The variables are as defined in Table 1. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero. Panels A, B, and C report our results where the dependent variable is the quarterly growth rate in (a) U.S. industrial production, (b) real GDP, and (c) corporate earnings.

Panel A: Quarterly Real Gross Domestic Product Growth Rate, QYG(t+k)

k	OFM(t)	OFD(t)	MKT(t)	SMB(t)	HML(t)	WML(t)	FEG(t)	NEQ(t)	DEF(t)	TERM(t)	LIQ(t)
1	0.081*	-0.065*	0.068**	0.002	-0.030	-0.015	0.014	0.184..	-0.631*	0.051	0.021
	(2.268)	(-1.952)	(3.167)	(0.067)	(-0.947)	(-0.748)	(1.028)	(1.782)	(-2.408)	(0.874)	(1.166)
	0.090	0.065	0.123	0.000	0.016	0.006	0.016	0.050	0.073	0.014	0.020
2	0.089**	-0.095**	0.040	-0.010	-0.017	-0.018	0.000	0.113	-0.144	0.059	-0.009
	(2.899)	(-2.668)	(1.598)	(-0.283)	(-0.379)	(-0.986)	(0.001)	(1.171)	(-0.493)	(0.799)	(-0.461)
	0.108	0.137	0.042	0.001	0.005	0.009	0.000	0.019	0.004	0.018	0.003
3	0.095**	-0.098**	0.029	0.019	-0.020	-0.003	0.018	0.228**	-0.189	0.061	0.008
	(3.160)	(-2.798)	(1.100)	(0.496)	(-0.608)	(-0.108)	(1.338)	(3.083)	(-0.574)	(0.734)	(0.452)
	0.125	0.145	0.023	0.004	0.007	0.000	0.027	0.078	0.007	0.020	0.003
4	0.050	-0.074..	-0.002	-0.010	0.025	-0.008	-0.002	0.222..	-0.111	0.083	-0.013
	(1.468)	(-1.706)	(-0.068)	(-0.229)	(0.833)	(-0.305)	(-0.120)	(1.782)	(-0.326)	(1.037)	(-0.701)
	0.034	0.082	0.000	0.001	0.011	0.001	0.000	0.074	0.002	0.036	0.008

Table 3-4 Univariate Predictive Regressions for Economic Growth (continued)

Panel B: Quarterly Industrial Production Growth Rate, QPG(t+k)

k	OFM(t)	OFD(t)	MKT(t)	SMB(t)	HML(t)	WML(t)	FEG(t)	NEQ(t)	DEF(t)	TERM(t)	LIQ(t)
1	0.122	-0.174*	0.130*	0.069	-0.130*	0.007	1.629	0.556**	-1.623**	0.160	0.109**
	(1.424)	(-2.041)	(2.268)	(0.729)	(-2.098)	(0.138)	(1.672)	(2.796)	(-2.633)	(1.155)	(3.028)
	0.047	0.110	0.106	0.014	0.068	0.000	0.039	0.106	0.112	0.031	0.122
2	0.215**	-0.201*	0.142**	0.016	-0.098	-0.027	0.026	0.575**	-1.074*	0.202	0.091*
	(2.945)	(-2.351)	(3.827)	(0.326)	(-0.903)	(-0.657)	(0.946)	(2.880)	(-2.090)	(1.170)	(2.451)
	0.145	0.141	0.124	0.001	0.038	0.005	0.014	0.113	0.049	0.050	0.085
3	0.120	-0.207*	0.034	-0.096	-0.013	0.002	0.023	0.408..	-1.003..	0.174	0.062
	(1.144)	(-2.273)	(0.623)	(-1.290)	(-0.099)	(0.044)	(0.829)	(1.967)	(-1.953)	(0.882)	(1.619)
	0.045	0.149	0.007	0.026	0.001	0.000	0.011	0.057	0.042	0.037	0.039
4	0.200*	-0.252**	0.067**	0.028	-0.025	-0.056	-0.004	0.661**	-1.238*	0.133	0.056
	(2.340)	(-3.357)	(2.608)	(0.323)	(-0.432)	(-1.298)	(-0.138)	(3.319)	(-2.086)	(0.653)	(1.433)
	0.125	0.214	0.028	0.002	0.002	0.019	0.000	0.149	0.065	0.021	0.032

Table 3-4 **Univariate Predictive Regressions for Economic Growth (continued)**

Panel C: Quarterly Corporate Earnings Growth Rate, QEG(t+k)											
k	OFM(t)	OFD(t)	MKT(t)	SMB(t)	HML(t)	WML(t)	FEG(t)	NEQ(t)	DEF(t)	TERM(t)	LIQ(t)
1	0.041	-0.800**	0.264..	0.469**	0.077	-0.120	0.333**	0.761	-2.765	0.890*	0.147
	(0.178)	(-4.187)	(1.650)	(2.007)	(0.377)	(-0.724)	(4.002)	(1.101)	(-1.421)	(2.518)	(0.784)
	0.000	0.210	0.040	0.058	0.002	0.008	0.195	0.018	0.030	0.088	(0.009)
2	-0.015	-0.531**	0.242	0.174	-0.170	0.014	-0.006	1.459*	-1.997	1.160**	0.185
	(-0.070)	(-1.989)	(1.604)	(0.772)	(-0.898)	(0.092)	(-0.070)	(2.322)	(-1.084)	(3.680)	(0.984)
	0.000	0.103	0.038	0.009	0.012	0.000	0.000	0.077	0.018	0.172	0.015
3	0.046	-0.804**	0.311*	0.320	-0.215	-0.116	-0.001	1.620*	-1.052	1.479**	0.086
	(0.210)	(-3.295)	(2.072)	(1.426)	(-1.133)	(-0.743)	(-0.012)	(2.590)	(-0.563)	(4.994)	(0.453)
	0.001	0.237	0.063	0.031	0.020	0.009	0.000	0.095	0.005	0.280	0.003
4	-0.048	-0.710**	0.077	0.324	-0.187	-0.161	0.022	2.219**	0.525	1.540**	-0.212
	(-0.217)	(-2.720)	(0.495)	(1.434)	(-0.971)	(-1.029)	(0.248)	(3.683)	(0.278)	(5.171)	(-1.109)
	0.001	0.180	0.004	0.032	0.015	0.017	0.001	0.177	0.001	0.298	0.019

Table 3-5 **Multivariate Predictive Regressions for Economic Growth**

This table presents the coefficient estimates, t-statistics (in smaller font below), and R^2 values from multivariate models where the quarterly industrial production, real GDP, and corporate earnings growth (*QPG*, *QYG*, and *QEG*) up to four quarters ahead are regressed on the market order flow and the order flow differential (*OFM* and *OFD*). The first column in each panel reports the results from models where *OFM* and *OFD* are the only two explanatory variables. The second column contains the results after controlling for the excess market return (*MKT*), the returns for the small-minus-big, high-minus-low, and winners-minus-losers portfolios (*SMB*, *HML*, and *WML*), and the first lag of the dependent variable. The variables are as defined in Table 1. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero. Panels A, B, and C report our results where the dependent variable is the quarterly growth rate in (a) U.S. industrial production, (b) real GDP, and (c) corporate earnings.

		Panel A: Real GDP Growth, QYG(t+k)				Panel B: Industrial Production Growth, QPG(t+k)				Panel C: Corporate Earnings Growth, QPG(t+k)			
k		OFM(t)	OFD(t)	OFM(t)	OFD(t)	OFM(t)	OFD(t)	OFM(t)	OFD(t)	OFM(t)	OFD(t)	OFM(t)	OFD(t)
1	Estimate	0.074*	-0.056..	0.041	-0.041	0.102	-0.163*	0.083	-0.092..	-0.063	-0.807**	-0.141	-0.509*
	t-statistic	(2.187)	(-1.886)	(1.360)	(-1.456)	(1.353)	(-2.146)	(1.237)	(-1.769)	(-0.233)	(-2.664)	(-0.269)	(-2.641)
	R-squared		0.139	ΔRsq	0.038		0.142	ΔRsq	0.042		0.211	ΔRsq	0.074
2	Estimate	0.077**	-0.085**	0.061*	-0.083**	0.189**	-0.177*	0.146*	-0.186**	-0.113	-0.583	-0.228	-0.496*
	t-statistic	(2.611)	(-2.765)	(2.356)	(-3.465)	(3.316)	(-2.318)	(2.448)	(-2.779)	(-0.361)	(-1.540)	(-1.422)	(-2.499)
	R-squared		0.216	ΔRsq	0.122		0.252	ΔRsq	0.148		0.106	ΔRsq	0.075
3	Estimate	0.082**	-0.087**	0.091**	-0.112**	0.091	-0.195**	0.085	-0.252**	-0.050	-0.756*	-0.200	-0.820**
	t-statistic	(3.135)	(-2.695)	(3.373)	(-3.672)	(1.209)	(-2.686)	(1.374)	(-3.193)	(-0.131)	(-1.970)	(-1.029)	(-4.625)
	R-squared		0.236	ΔRsq	0.243		0.174	ΔRsq	0.169		0.238	ΔRsq	0.180
4	Estimate	0.039	-0.069..	0.050..	-0.084..	0.164**	-0.227**	0.154*	-0.303**	-0.129	-0.672	-0.184	-0.803*
	t-statistic	(1.311)	(-1.609)	(1.864)	(-1.841)	(2.614)	(-3.953)	(2.428)	(-4.658)	(-0.326)	(-1.857)	(-0.750)	(-2.652)
	R-squared		0.102	ΔRsq	0.108		0.296	ΔRsq	0.287		0.189	ΔRsq	0.169
Controls		No		Yes		No		Yes		No		Yes	

Table 3-6 Predictive Regressions for Decile Portfolio Returns

This table presents the coefficient estimates, Newey-West (1987) adjusted t-statistics (in parentheses), and R^2 values from the regressions of quarter t+1 decile portfolio returns on the quarterly change in the market order flow and order flow differential (*ΔOFM* and *ΔOFD*) from quarter t-1 to t. The first column in each panel reports the results from models where *OFM* and *OFD* are the only two explanatory variables. The second column controls for only the return factors (*MKT*, *SMB*, *HML*, and *WML*). The last column contains the results from the full model where *DEF*, *TERM*, *NEQ*, *FEG*, *LIQ*, and *ΔSENT* are included. The variables are as defined in Table 1. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero.

Size Decile		ΔOFM	ΔOFD	ΔOFM	ΔOFD	ΔOFM	ΔOFD
1	Estimate	1.351	2.087**	1.251	2.491**	0.379	2.793**
	t-statistic	(1.435)	(2.835)	(1.578)	(2.986)	(0.816)	(3.728)
	R-squared		0.139		0.175		0.353
2	Estimate	1.368**	1.751**	1.211*	1.884**	0.833	1.957**
	t-statistic	(2.308)	(2.959)	(2.048)	(2.721)	(1.488)	(2.667)
	R-squared		0.190		0.208		0.390
3	Estimate	0.729*	1.385**	0.571	1.704**	0.074	1.869**
	t-statistic	(2.048)	(3.032)	(1.528)	(3.133)	(0.176)	(3.120)
	R-squared		0.148		0.164		0.325
4	Estimate	0.802*	1.204**	0.650	1.497**	0.311	1.688**
	t-statistic	(2.643)	(3.132)	(1.572)	(2.987)	(0.656)	(3.690)
	R-squared		0.135		0.162		0.310
5	Estimate	0.701*	1.484**	0.578	1.707**	0.227	1.785**
	t-statistic	(2.545)	(3.989)	(1.473)	(3.551)	(0.432)	(3.424)
	R-squared		0.171		0.206		0.337
6	Estimate	0.580**	1.352**	0.444	1.603**	0.377	1.799**
	t-statistic	(3.153)	(3.760)	(1.370)	(3.104)	(0.977)	(3.840)
	R-squared		0.139		0.189		0.325
7	Estimate	0.512**	1.401**	0.428	1.523**	0.072	1.744**
	t-statistic	(2.739)	(3.941)	(1.321)	(3.066)	(0.177)	(3.403)
	R-squared		0.127		0.155		0.303
8	Estimate	0.406	1.253**	0.387	1.153*	0.206	1.408**
	t-statistic	(1.547)	(3.177)	(1.288)	(2.216)	(0.555)	(3.371)
	R-squared		0.121		0.146		0.293
9	Estimate	0.376..	1.376**	0.382	1.277*	0.147	1.554**
	t-statistic	(1.699)	(3.618)	(1.220)	(2.599)	(0.388)	(3.454)
	R-squared		0.133		0.160		0.324
10	Estimate	0.212	1.100**	0.286	0.611	0.479	0.758..
	t-statistic	(0.613)	(3.396)	(0.992)	(1.341)	(1.584)	(1.986)
	R-squared		0.097		0.166		0.349
Control Variables							
Return Factors			No		Yes		Yes
Other Controls			No		No		Yes

Table 3-7 Predictive Regressions for Return Premiums

This table presents the coefficient estimates, Newey-West (1987) adjusted t-statistics (in parentheses), and R^2 values from the regressions of quarter *t*+1 realizations of the four common return factors (*MKT*, *SMB*, *HML*, and *WML*) on the quarterly change in the market order flow and order flow differential (*ΔOFM* and *ΔOFD*) from quarter *t*-1 to *t*. The first column in each panel reports the results from models where *OFM* and *OFD* are the only two explanatory variables. The second column controls for only the return factors (*MKT*, *SMB*, *HML*, and *WML*). The last column contains the results from the full model where *DEF*, *TERM*, *NEQ*, *FEG*, *LIQ*, and *ΔSENT* are included. The variables are as defined in Table 1. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero.

Portfolio		ΔOFM	ΔOFD	ΔOFM	ΔOFD	ΔOFM	ΔOFD
MKT	Estimate	0.084	0.417**	0.065	0.325*	0.108	0.400*
	t-statistic	(0.861)	(3.859)	(0.662)	(2.003)	(0.868)	(2.309)
	R-squared		0.096		0.137		0.295
SMB	Estimate	0.083	0.240*	0.056	0.277**	0.004	0.436**
	t-statistic	(0.908)	(2.407)	(0.496)	(2.770)	(0.335)	(3.469)
	R-squared		0.078		0.101		0.334
HML	Estimate	0.029	-0.198*	-0.024	-0.119*	-0.164..	-0.090
	t-statistic	(0.323)	(-2.380)	-0.195	(-0.710)	(-1.969)	(-0.683)
	R-squared		0.032		0.051		0.274
WML	Estimate	0.085	-0.245..	0.028	-0.035	0.108	0.035
	t-statistic	(0.562)	(-1.807)	(0.165)	(-0.213)	(0.839)	(0.196)
	R-squared		0.036		0.139		0.303
Control Variables							
Return Factors			No		Yes		Yes
Other Controls			No		No		Yes

Table 3-8 Order Flow Differentials for Size and Liquidity

This table presents the coefficient estimates and Newey-West (1987) adjusted t-statistics (in smaller font below) from multivariate models where the quarterly industrial production, real GDP, and corporate earnings growth (*QPG*, *QYG*, and *QEG*) up to four quarters ahead are regressed on the market order flow and the order flow differentials in terms of size and liquidity (*OFM*, *OFDS*, and *OFDL*), controlling for *MKT*, *SMB*, *HML*, *WML*, and the first lag of the dependent variable. To obtain *OFDS* and OFDL, we sort the sample by market capitalization and illiquidity (using the Amihud (2002) measure) at the beginning of each July and allocate the stocks into four portfolios formed using median size and liquidity as cutoff points. The rest of the variables are as defined in Table 1. The change in model R^2 when each of *OFM*, *OFDS*, or *OFDL* is added to a model comprised of *MKT*, *SMB*, *HML*, and *WML* is reported as ΔR^2. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero. Panels A, B, and C report our results where the dependent variable is the quarterly growth rate in (a) real GDP, (b) U.S. industrial production, and (c) corporate earnings.

		Panel A: Real GDP Growth			Panel B: Industrial Production Growth			Panel C: Corporate Earnings Growth		
k		OFM(t)	OFDS(t)	OFDL(t)	OFM(t)	OFDS(t)	OFDL(t)	OFM(t)	OFDS(t)	OFDL(t)
1	Estimate	0.030	-0.039	-0.037	0.008	-0.125*	-0.133*	-0.169	-0.600**	-0.457*
	t-statistic	(1.039)	(-1.515)	(-1.424)	(0.170)	(-2.514)	(-2.684)	(-1.058)	(-4.829)	(-2.662)
	ΔR^2	0.041	0.018	0.010	0.009	0.025	0.020	0.001	0.057	0.007
2	Estimate	0.027	-0.093**	-0.071**	0.045	-0.129	-0.165*	-0.060	-0.700**	-0.101
	t-statistic	(1.048)	(-3.353)	(-2.951)	(0.997)	(-1.411)	(-2.155)	(-0.294)	(-3.581)	(-0.401)
	ΔR^2	0.061	0.089	0.017	0.041	0.025	0.042	0.001	0.167	0.015
3	Estimate	0.037	-0.113**	-0.043..	0.015	-0.231**	-0.171*	-0.099	-0.819**	-0.399..
	t-statistic	(1.788)	(-3.100)	(-1.801)	(0.354)	(-3.423)	(-2.640)	(-0.654)	(-4.387)	(-1.855)
	ΔR^2	0.083	0.177	0.000	0.021	0.108	0.015	0.000	0.166	0.000
4	Estimate	0.025	-0.084	-0.087*	0.046	-0.223**	-0.146**	-0.099	-0.754*	-0.524..
	t-statistic	(1.683)	(-1.612)	(-2.272)	(0.967)	(-3.230)	(-2.853)	(-0.677)	(-2.236)	(-1.707)
	ΔR^2	0.049	0.045	0.035	0.043	0.106	0.007	0.000	0.104	0.006
Controls		MKT, SMB, HML, WML			MKT, SMB, HML, WML			MKT, SMB, HML, WML		

Table 3-9 Predicting Decile Portfolio Returns Using OFDS and OFDL

This table presents the coefficient estimates, Newey-West (1987) adjusted t-statistics (in parentheses), and R^2 values from the regressions of quarter *t*+1 decile portfolio returns on the quarterly change in the order flow differentials in terms of size and liquidity (*ΔOFDS* and *ΔOFDL*) from quarter *t*-1 to *t*, controlling for contemporaneous realizations of *MKT*, *SMB*, *HML*, *WML*, *DEF*, *TERM*, *NEQ*, *FEG*, *LIQ*, and *ΔSENT*. The change in model R^2 when *ΔOFDS* or *ΔOFDL* is added to a model comprised of the listed control variables is reported as ΔR^2. *ΔOFDS* or *ΔOFDL* are defined in the notes for Table 8. The rest of the variables are as defined in Table 1. '**', '*', and '..' signify statistical significance at 1%, 5%, and 10% levels as indicated by a two-tailed test of the null hypothesis that the coefficient estimate is equal to zero.

	Decile	ΔOFDS	ΔOFDL	Decile	ΔOFDS	ΔOFDL
Estimate	1	2.783	1.978	6	1.487	0.563
t-statistic		(2.697)	(2.470)		(3.082)	(1.547)
ΔR-squared		0.060	0.008		0.081	0.007
R-squared			0.139			0.096
Estimate	2	2.136	1.464	7	1.355	0.462
t-statistic		(2.556)	(2.471)		(2.679)	(1.297)
ΔR-squared		0.062	0.004		0.068	0.009
R-squared			0.130			0.080
Estimate	3	1.823	1.045	8	1.160	0.277
t-statistic		(2.616)	(1.890)		(2.670)	(1.255)
ΔR-squared		0.078	0.000		0.075	0.018
R-squared			0.120			0.080
Estimate	4	1.666	0.912	9	1.306	0.177
t-statistic		(2.801)	(1.758)		(2.946)	(0.615)
ΔR-squared		0.070	0.000		0.104	0.032
R-squared			0.104			0.105
Estimate	5	1.775	1.005	10	0.806	-0.177
t-statistic		(3.683)	(2.139)		(2.314)	(-0.755)
ΔR-squared		0.079	0.001		0.102	0.073
R-squared			0.122			0.107
Controls		ALL			ALL	

Chapter 4

Trading Activity, Price Informativeness, and the Business Cycle

4.1 Introduction

In their seminal paper, Grossman and Stiglitz (1980) argue that securities markets are characterized by an "equilibrium degree of disequilibrium, where security prices reflect the information of informed individuals, but only partially, so that those who expend resources to obtain information receive compensation." This equilibrium degree of disequilibrium is characterized by a trade-off between the informativeness of the price system and the incentives that the individuals in the system have for acquiring private information. Clearly, in a market state where prices are perfectly informative, there is no room for arbitrageurs to function, and in a market state where arbitrageurs do not function, it is paradoxical to have perfectly informative prices. Based on their model, Grossman and Stiglitz (1980) make three propositions relevant to our paper. They posit (a) as the proportion of informed individuals increases, the price system becomes more informative, (b) a greater level of noise would render the price system less informative for uninformed individuals and lead to an increase in the proportion of individuals who are informed, and (c) as the cost of obtaining information increases, the equilibrium proportion of individuals who are informed will be smaller.

Do higher trading costs hinder arbitrage activity and reduce the amount of private information generated by arbitrageurs? Does greater private information generation increase trading costs through alleviating the adverse selection problem faced by market makers? What is the role of liquidity in determining the nature of these interrelationships? How do liquidity, trading costs, and private information generation vary across the business cycle? Are there real effects associated with reductions in private information generation? This paper tries to address these questions through investigating the interrelationship between trading costs, trading activity, and the share of firm-specific information in price movements over an 83-year period from 1926 to 2008, focusing on business-cycle patterns in these variables. Trading activity and trading costs are proxied, respectively, by share turnover (STO) and the average price impact of trading (PIM), defined as the change in price implied by a $1 million trade (in 2008 dollars) following Amihud (2002). Consumer sentiment (SEN) is used as an instrument to capture the general mood of individuals in the economy. Following the insights in Roll (1988) and Morck et al. (2000), the share of firm-specific information in price movements (FSI), computed as one minus the market model R^2, is used as a measure of the informational efficiency of the pricing system. Following Chen, Goldstein, and Jiang (2007), the informativeness of prices for corporate managers is measured as the sensitivity of corporate investment in a given year to the normalized price (Tobin's Q) at the end of the previous year. I also use data on financial analysts' forecasts to construct additional measures of firms' informational environment. These measures are the number of analysts providing earnings-per-share estimates for a given firm (NUM) and the dispersion of these analysts' forecasts (FDISP).

My main results are as follows. I first show that both price impact and FSI display discernible business-cycle patterns, with PIM increasing, and FSI declining significantly during recessions. Two-way causality tests in the spirit of Granger (1969) and Sims (1972) reveal, at the market-level, FSI is caused by consumer sentiment and price impact, and price impact is in turn caused by sentiment and share turnover. Time-series tests of the interrelationship between these variables indicate that FSI is related negatively to contemporaneous and lagged changes in sentiment and PIM, while PIM itself is related negatively to STO and sentiment. Next, I study PIM, the probability of informed trading (PIN), and FSI at the firm level and investigate possible sources of cross-sectional variability in these variables. Here, I find that PIM and PIN are lower for big firms, growth firms, firms whose stock is more liquid, firms with extensive analyst coverage, and firms with lower analyst forecast dispersion. Note that these are stocks that potentially grab the attention of liquidity traders (e.g. Barber and Odean, 2008). In line with this observation, I show that such stocks experience the greatest increase in the probability of informed trading as a recession hits the economy. FSI, on the other hand, is greater for small stocks, value stocks, less liquid stocks, and for stocks with higher trading costs, more disperse analyst forecasts, or little or no analyst coverage.

Collectively, these findings are consistent with a world where a decline in uninformed investor activity aggravates the adverse selection problem faced by market-makers–an effect that is distinct from an increase in the adverse selection problem due to greater informed trading activity. The market-makers rationally respond by adjusting their pricing functions, driving up trading costs. The increase in trading costs reduces the amount of firm-specific information that is incorporated into stock prices through informed trading, since a certain fraction of the signals that used to be profitable in low trading cost regimes will not be worth trading on based on the new, and worsened, terms of trade. In the end, we face a strategic interaction where market-makers know that a trade, if executed, is more likely to come from an informed trader and informed traders know that a trade, if executed, will be less profitable since the market maker knows that the trade is more likely to be information-based. The end-result of this interaction between informed traders and the market-maker is an equilibrium which is optimal for both parties playing this game, but potentially suboptimal for the informational efficiency of the market since part of the relevant firm-specific information is left unincorporated into stock prices. This reduction in information-based trading appears to have a material effect on the informativeness of prices for corporate managers. [48] The sensitivity of corporate investment to the information in prices is 62.5% lower in recessions, underlining the importance of well-functioning, informationally-efficient securities markets.

In a related study, Chordia, Roll, and Subrahmanyam (2001) conduct a higher frequency analysis of aggregate market spreads, depths, and trading activity for U.S. stocks over the period 1988 through 1998. The authors demonstrate that the increase in spreads in down markets is greater in magnitude than the decline in spreads during up markets, while the effect of up and down markets on trading activity is roughly symmetric. This asymmetric relation between the marketwide averages for spreads and returns is consistent with the notion in our paper that the greater price impacts during recessions come about as a result of a decline in uninformed trading activity instead of an increase in informed arbitrage activity.

To the best of my knowledge, this is the first paper to analyze the interrelationship among trading costs, trading activity, and firm-specific information dissemination over a sample period as long as ours. The results of this analysis are important for several reasons. First, delineating the link between informed trading and adverse selection problem in financial markets is beneficial for future research. My results imply that the activity of uninformed traders has an important influence on measures of adverse selection such as PIN and PIM, an influence distinct from that of the intensity of information-based trading. Second, by characterizing the business-cycle patterns in trading costs and private information generation, we provide perspective on the deadweight costs of recessions for the functioning of financial markets. In doing so, we quantify the approximate effect of the high trading cost – low informed arbitrage activity market regimes observed during

[48] Theoretical evidence, exemplified by Dow and Gorton (1997) and Subrahmanyam and Titman (1999), holds managers learn from the information in prices about the prospects of their own firms as prices incorporate information from many informed investors, some of whom may have no channels but the trading process to communicate with the firm.

recessions on the informativeness of prices in guiding corporations' investment decisions. The significant decline in private information generation and the accompanying reduction in price informativeness suggest that promoting trading activity by households and other non-arbitrageurs at all times is critical for the informational efficiency of financial markets. In this sense, the role played by financial analysts in encouraging their clienteles to trade more actively may paradoxically be benefiting the market as a whole. Policy-makers may, hence, find it worthwhile to promote trading activity in financial markets through improving shareholder property rights and providing unsophisticated investors with easier-access investment vehicles in order to attract greater non-arbitrage demand.[49]

The next section presents a brief summary of the relevant literature. Section 3 describes our data and variables. Section 4.1 inspects business-cycle patterns in price impact, share turnover, and arbitrage activity. Section 4.2 studies the interrelationship between these variables at the market level. Sections 4.3 and 4.4 analyze cross-sectional determinants of PIM, PIN, and FSI. Section 4.5 investigates the effect of business-cycle fluctuations on trading costs, arbitrage activity, and price informativeness. Section 5 concludes.

4.2 Literature Review and Research Questions

In their seminal paper, Grossman and Stiglitz (1980) present a model "in which there is an equilibrium degree of disequilibrium, where prices reflect the information of informed individuals (arbitrageurs) but only partially, so that those who expend resources to obtain information do receive compensation." Among

[49] This is consistent with the findings in Morck, Yeung, and Yu (2000) that (a) countries with greater shareholder protection tend to have more informationally efficient markets and (b) there is a steady improvement in the informational efficiency of the U.S. market in the last twenty years, given the 1980-1990s seen a booming interest in public's attention in mutual funds.

other things, the author posit (i) as the proportion of informed individuals increases, the price system becomes more informative, (ii) a greater level of noise would render the price system less informative for uninformed individuals and lead to an increase in the proportion of individuals who are informed, and (iii) as the cost of obtaining information increases, the equilibrium proportion of individuals who are informed will be smaller. These three propositions guide us in our empirical analysis in this paper.

We focus on trading activity, trading costs, and private information dissemination over an eighty-two year period of the U.S. economy. Trading activity is proxied by share turnover (STO), defined as the volume of trading in a given period as a percentage of a firm's market value. In general, we expect share turnover to be high when market conditions are favorable, but, given STO is an unsigned measure, extreme negative events may also trigger increased trading activity. Our proxy for trading costs is the price impact of trading (PIM), which is estimated using daily price data as in Amihud (2002). Despite being a more crude measure compared to finer microstructure variables estimated from trading data (such as Kyle's λ or the liquidity and adverse selection components of the bid-ask spread), PIM has the advantage of being estimable over a much longer sample horizon. In addition, it is also shown to correlate strongly with the mentioned finer measures and contain information about stock returns (Amihud, 2002).[50]

For private information dissemination, we use the share of price movements that is due to firm-specific information (FSI) as our proxy. Roll (1988) shows that the relative amounts of firm-specific and marketwide information disseminated into the economy determine the extent to which stocks in a market move together. In support of this, Morck, Yeung, and Yu (2000) demonstrate that cross-country differences in the systematic component of return variation are well explained by measures of property rights and argue that strong property rights promote

[50] Amihud (2002) shows that expected the excess stock return is related positively to expected component of price impact and negatively to its unexpected component. Price impact is often used as a measure of either illiquidity and/or adverse selection.

informed arbitrage and increase the amount of firm-specific information that is ultimately incorporated into stock prices. Following such insights, we define FSI as one minus the R^2 from quarterly regressions of daily individual stock returns on contemporaneous and, to account for stale price effects, one-day-lagged value-weighted market return. We expect FSI to be high when private information is generated more efficiently in the economy.

Does greater private information generation increase trading costs through alleviating the adverse selection costs faced by market makers? Do higher trading costs hinder arbitrage activity? What is the role of liquidity in determining the nature of these interrelationships? Are there real effects implied by higher trading costs or lower informational efficiency? This paper tries to address these questions. As a starting point, Kyle (1985) model tells us that price impact should be related negatively to the variability of uninformed order flow and positively to the magnitude of the signal that is yet to be incorporated into stock prices. We start by testing this prediction using consumer sentiment and share turnover as two instruments that should, arguably, correlate positively with uninformed trading activity. We first establish the direction of causality among PIM, FSI, STO, and consumer sentiment. Then, time-series tests are run to establish and quantify the interrelationships between these variables. FSI is included in this test as an informed trading proxy, but it is not clear that a greater fraction of firm-specific information in price movements necessarily implies that the private signal is large in magnitude due to the scaled nature of the FSI measure (we introduce a better proxy in our cross-sectional tests). The results from this test, presented in Section 4.2, are consistent with a world where negative shocks to uninformed trading activity aggravates the adverse selection problem faced by market-makers. In this world, market-makers rationally respond by increasing the price impact of trading and the hike in price impact results in a decline in the amount of private information disseminated into prices through trading, as some signals are rendered unworthy to trade on due to the new terms of trade.

In our cross-sectional tests, we expand on our findings from the market-level time-series tests, and determine how price impact, FSI, and the probability of informed trading (PIN) is related to stock characteristics such as market capitalization (MV), book-to-market equity (BM), and liquidity, and to informational proxies such as the extent of financial analyst coverage (NUM) and the dispersion of these analysts' forecasts (FDISP). Here, we hypothesize that the presence and the richness of analyst coverage might lead to a decline in trading costs (and PIN) for two possible reasons, both precipitated by a reduction in the adverse selection problem. First, detailed reports made public by analysts may reduce information asymmetry between investors. Second, analyst coverage might induce greater liquidity trader activity, as in the "attention-grabbing" stocks argument of Barber and Odean (2008), resulting in a dilution of the trader pool from more informed towards less informed. Note here that, while the first story does not have a clear impact on FSI, the second should result in a decline in this variable. Controlling for analyst coverage, we also argue that the magnitude of private signals for a firm might be higher if analysts' views on the firm's outlook are more disperse, implying a positive relationship between PIM and FDISP and, potentially, between FSI and FDISP. Our findings are broadly consistent with these arguments and are presented in Sections 4.3 and 4.4.

Another dimension of the issue is the informativeness of stock prices in guiding the investment decisions of corporate managers. Extant research in corporate finance holds that firm managers can learn from the information in stock prices as prices aggregate information from many different investors, most of whom may not have direct channels for communication with the firm apart from the trading process (Dow and Gorton, 1997; Subrahmanyam and Titman, 1999). Chen, Goldstein, and Jiang (2007) argue that this learning process is expected to manifest itself in the sensitivity of corporate investment to stock prices, with the sensitivity being higher for stocks whose prices convey more information that is not already known by firm managers. Chen et al. show that the sensitivity of corporate investment to the information in stock prices is higher and the ex-post operating performances (as measured by return on assets, sales growth, and asset turnover rate) are better for firms whose shares are subject to more information-based trading. Consistent with this, Fang, Noe, and Tice (2008) find that firms with more liquid stocks tend to perform better and demonstrate that this superior performance is due to the increased information content of prices and enhanced incentive effects of performance-based compensation contracts. We show earlier in Section 4.1 that recessions are characterized by lower trading activity, higher trading costs, and a smaller share for firm-specific information in price movements. Therefore, as a final test, we investigate whether the sensitivity of corporate investment to the information in prices is also conditioned by the state of the economy. The results reported in Section 4.5 confirm our prediction that prices should be less guiding for corporate investment during recessions.

In a study that complements ours at a daily frequency with finer microstructure data, but for a shorter sample period, Chordia, Roll, and Subrahmanyam (2001) study aggregate market spreads, depths, and trading activity for U.S. stocks over the period from 1988 to 1998 and show that (i) spreads and market depth respond asymmetrically to equity market returns with down markets seeing a greater increase (decline) in quoted and effective spreads (market depth) compared to the decline (increase) in up markets, while the effect on trading volume is roughly symmetric, (ii) high market volatility in recent periods leads to a reduction in trading activity and spreads in the current period, (iii) increases in the short-term interest rate and the term spread lead to a widening of the quoted spread and a reduction of the market depth and trading activity, and (iv) market depth declines and trading activity increases prior to/during important macroeconomic announcements. The finding that (compared to up markets) down markets see an asymmetric increase in quoted and effective spreads but not in trading volume is consistent with the notion in our paper that the greater price impacts during recessionary periods come about as a result of the increase in adverse selection costs faced by market-makers when uninformed trading activity declines.

4.3 Data and Variables

Our main variables of interest are the share of firm-specific information in price movements (FSI), price impact of trading (PIM), and share turnover (STO). FSI is defined as one minus the R-square quarterly time-series regressions of daily individual stock returns on the contemporaneous and one-day-lagged value-weighted market return. Following Amihud (2002), PIM is estimated as absolute daily returns divided by daily dollar volume (in December 2008 dollars) averaged over each quarter. STO is the quarterly share volume divided by shares outstanding. The date that is used in the estimation of these variables is from the Centre for Research in Security Prices (CRSP) and is available over the eighty-three year period from January 1926 to December 2008.

Other than this, we also make use of a combination of CRSP, COMPUSTAT, and I/B/E/S databases to compute a set of variables that we use in our cross-sectional regressions. This set of variables include the market value of equity (MV), book-to-market ratio (BM), analyst forecast dispersion (FDISP), and the number of analysts following a firm (NUM) in our cross-sectional regressions. MV is computed at the beginning of each quarter as the natural logarithm of the product of closing share price of the previous quarter and number of shares outstanding. BM is computed at the beginning of each year as the market value of equity at the previous year end divided by its book value as reported in the previous year's financial statement. NUM is the number of analysts providing one-year-ahead earnings-per-share forecasts for a given firm and FDISP is the standard deviation of these forecasts scaled by the mean estimate. Lastly, we use the University of Michigan Consumer Sentiment Index (SEN) obtained from the St. Louis FED database as an instrument that captures the general mood of individuals.

Table 1 provides the means, standard deviations, and the minimum, median, and maximum values for the variables described above, as well as the correlations among them. We report time-series correlations among the marketwide aggregates for FSI, PIM, and STO as we later investigate the relation between these variables over time. By contrast, for MV, BM, FDISP, and NUM, which are later used in cross-sectional regressions, time-series averages of the cross-sectional correlation coefficients are presented. FSI ranges between 0.48 and 0.97 with a mean (median) of 0.88 (0.91). For PIM, the mean (median) is 1.47 (0.56) and the minimum and maximum values are 0.01 and 38.1. For the median firm, the mean (median) quarterly share turnover is 0.10 (0.06) with a minimum of almost zero and a maximum of 0.42. The mean (median) values for firm size and BM are $35 million ($3 million) and 0.9 (0.9). Last, over the period from 1977 to 2006, the mean and median dispersion in analysts' forecasts is roughly equal to 5% of the reported estimate and the mean and median number of analysts covering a firm is 2.

Turning to the correlations between our variables (lower panel), we see that, over time, marketwide FSI is related negatively to aggregate PIM (ρ= -0.22) and positively to average turnover (ρ=0.12), while STO and PIM are negatively associated (ρ= -0.25). Thus, firms with lower trading costs and higher turnovers tend to have a greater fraction of their price movements due to private information, and firms with greater trading activity tend to be those with lower trading costs. There does not appear to be a significant relation between stock market returns and contemporaneous price impact or share turnover, but we note that the time-series association between FDISP and RET is reliably positive (not tabulated). Over the cross-section, FSI is declining in firm size (ρ = -0.57) and number of analysts following the firm (ρ = -0.49) and increasing in analyst forecast dispersion (ρ = 0.10). The exact same pattern holds between PIM and MV (ρ = -0.25), NUM (ρ = -0.11), and FDISP (ρ = 0.04). Thus, smaller firms, firms that are not covered by analysts, and firms whose valuations are more uncertain also tend to have greater firm-specific information in their prices and higher trading costs. As might be expected, STO is increasing in firm size (ρ = 0.13) and number of analysts (ρ = 0.13) and declining in BM (ρ = -0.18). These findings are consistent with the evidence in Barber and Odean (2008) that noise traders tend to focus almost exclusively on "attention-grabbing" stocks, i.e. large-cap stocks, growth stocks, stocks covered by analysts. Last, the cross-sectional relation is negative between FDISP and firm size (ρ = -0.56) and FDISP and NUM (ρ = -0.13) and positive between MV and NUM (ρ = 0.38).

Finally, in our analyses of the determinants of the probability of informed trading (PIN), we merge estimates provided by Soeren Hvidkjaer with our data on MV, BM, STO, and FDISP. In testing the effect of high trading cost regimes on the informational quality of the market, we use an unbalanced panel of non-financial (SIC code 6000-6999) and non-utility (SIC code 4200) firms for the period from 1963 to 2006 to test the change in the sensitivity of corporate investment to the information in prices during recessions. Our three investment proxies are CAPX (capital expenditures, item 128), CAPXRND (capital expenditures plus R&D expenses, item 128 + item 46), and CHGASSET (the yearly change in the book value of assets, item 6), all scaled by the beginning-of-the year book assets. The main variables of interest are the normalized price (Q), FSI, RECD, and interaction terms between Q and FSI and Q and RECD. Q is defined as the market value of equity (item 24 times item 25) plus book value of assets minus the book value of equity (item 60), scaled by book value of assets. The control variables are the cash ratio (CFX, cash holdings divided by total assets), inverse book assets (INVA), and the market-adjusted three-year cumulative return beginning from the end of the investment year (RET3). INVA is the reciprocal of book assets. All explanatory variables are measured at the end of the fiscal year immediately preceding the investment period. This test is presented in Section 4.5. The next section provides an analysis of the business-cycle patterns in our main variables.

4.4 Results and Discussion

4.4.1 Business-Cycle Effects in Trading and Information

Our proxies for trading activity and trading costs are the quarterly share turnover (STO) and the Amihud (2002) price impact measure (PIM), estimated each quarter from daily return and volume data and reported in December 2008 dollars. The extent of private information disseminated into the market is captured, following Morck, Yeung, and Yu (2000), by the share of firm-specific information in price movements (FSI). FSI is computed as one minus the market model R^2 from regressions of daily stock returns on contemporaneous and lagged daily market return. We start our analysis by examining the behavior of the marketwide averages for PIM, STO, and FSI over the period from January 1926 to December 2008. The time-series patterns in these variables are depicted in Figure 1, where recessions, as identified by the National Bureau of Economic Research (NBER) is shaded in gray.

Starting with PIM, we observe that price impact demonstrates cyclical behavior throughout the eighty-year period. The highest levels of price impact are attained in 1932, 1975, and 1990, all these years lying within major recessions experienced in the U.S. We also see that influential financial crises, such as the stock market crash of October 1987 or events triggered by the Russian Financial Crisis and the fall of Long Term Capital Management, do push trading costs up, and the effects tend to last for a while before things revert back to normal. In addition to these, there is a clearly visible downward trend in PIM during the last two decades, with trading costs reaching historical lows in 2006 and 2007. The recession of 2008, however, does appear to be pushing price impacts back up.

In tandem with the decline in trading costs, trading activity appears to have boomed during the last two decades, increasing almost exponentially in 2000s. The highest values for STO are, therefore, reached within this period. The only historical episode where share turnover compares to that in 2000s is the brief expansionary period between the recessions of 1926-27 and 1929-33. Business-cycle fluctuations appear to affect share turnover as well. Paralleling the increase in price impacts during recessions, there seem to be sizable declines in share turnover in such periods. Last, while the share of firm-specific information does not display patterns as strong as those for PIM and STO, there seems to be declines in FSI following recessionary periods and stock market crashes. We will be investigating the time-series relationship between these variables more closely in Section 4.2, but given the evidence of strong cyclical patterns, a more thorough examination and quantification of business-cycle effects on our variables is in order. We compute the means, standard deviations, and the minimum, median, and maximum values for PIM, STO, FSI, and quarterly stock returns in the expansions and recessions within our sample period and report our results in Table 2.

The top panel reports the descriptive statistics for the quarterly measurements of our variables in expansionary and recessionary periods, while the bottom panel dissects the sample period into individual expansions and recessions and presents the variable means in each of these episodes. Starting first with trading activity, we observe that the share turnover (STO) ranges between 0.011 and 0.412 during recessions with a mean (median) of 0.0953 (0.0623), while the mean (median) is 0.1058 (0.0696) and the maximum and minimum turnover are 0.023 and 0.422 during expansions. These figures indicate that investors do trade less aggressively during economic downturns compared to expansionary periods. Indeed, as can be seen in the lower panel, share turnover almost always declines (with one exception) as the economy moves from an expansion to a recession.

Second, we observe that price impact (PIM) ranges between 0.0061 and 0.4795 with a mean (median) of 0.1404 (0.1071) during expansions. During recessions, the maximum and minimum values are 1.0935 and 0.0144 and the mean (median) is 0.2400 (0.1266). The difference in price impacts observed in the two different economic states is strikingly significant (both statistically and economically) and indicates that, in order to be able to trade, investors need to be willing to bear greater trading costs during economic downturns. In the lower panel of the table, it is seen that price impact always attains its local maxima during recessions, seeing average levels as high as 0.80, 0.56, and 0.35 during the recessions of 1990-91, 1973-75, and 1929-33.

Third, the share of firm-specific information in prices (FSI) ranges between 0.0292 and 0.9708 during expansions, with a mean (median) of 0.8881 (0.9131), while the mean (median) is 0.8521 (0.8805) and the minimum and maximum values are 0.041 and 0.9593 during recessions. The difference between the two means is statistically significant and suggests marketwide effects explain a greater fraction of price movements during economic downturns. Consistent with this, the period means in the lower panel of the table indicates that FSI tends to increase as the economy moves into recessions. This is consistent with the finding above that trading costs are greater in such periods, as greater trading costs may be expected to hinder arbitrage activity (more on this later).

Last, we report the descriptive statistics for quarterly returns. The time-series mean (median) of the cross-sectional aggregate return (RET) is 0.0156 (0.0072) in expansions and -0.0351 (-0.0444) during recessions. During the eighty-two-year period studied, aggregate market return has seen levels as low as -0.3943 and as high as 1.4767. Finally, looking at the individual period means reported in the lower panel of the table, we see that the highest average aggregate return (0.1220) is observed during the expansion of 1933 – 1937 and the lowest aggregate return (-0.1302) is observed during the recent recession of 2008.

4.4.2 The Relationship between Trading and Information

An interesting empirical regularity observed in the preliminary analysis conducted in the previous section is that the proportion of price changes due to firm-specific information declines significantly after local lows for share turnover and local highs for price impact. This might indicate that the activity of arbitrageurs is hindered by the lack of participation by liquidity traders and increased trading costs. In order to establish the direction of causality between these variables, and to quantify the effects more formally, this section presents a formal time-series analysis of the interrelationship among these variables.

As an additional proxy for changes in the activity of liquidity traders, we add consumer sentiment (SEN) into the set of variables under consideration. The conjecture here is that non-arbitrage traders would tend to decrease their presence in the market when the mood in the economy is bearish for at least two reasons. One, they might sell and go away or avoid trading completely, hoping to regain losses as the market recovers because they see investment in the stock market too risky during such times. Alternatively, unfavorable economic conditions might lead them to substitute consumption for investment and reduce the amount they allocate for investment in the stock market. In contrast, an arbitrageur would be willing to buy or sell shares regardless of the market mood, given there is a profitable arbitrage opportunity to exploit: whether the news are good or bad only determines which side of the trade the informed trader would be on.[51]

In order to set the stage, we run two-way bivariate vector autoregressions in the spirit of Granger (1969) and Sims (1972) for each possible pair of variables to establish the direction of causality between these four variables. The significance statistics and p-values from this analysis is presented in Table 2. These tests reveal that the marketwide averages for PIM, STO, and FSI are all Granger-caused by SEN. We also find that FSI is caused by PIM, which, in turn, is caused

[51] This instrumentation is supported by a correlation of 0.35 between consumer sentiment and share turnover and a correlation of 0.51 between consumer sentiment and stock market participation rate of households available from the U.S. Census Bureau database. We are aware that, despite our best efforts to convince the reader, the instrumentation described above would remain arguable. For the unconvinced, it would be best to interpret the relations that are to be investigated in the remainder of this section as being directly between consumer sentiment and the rest of the variables.

by STO. With these causalities in mind, we resume our analysis with time-series regressions of the changes in PIM and FSI (ΔPIM and ΔFSI) on their own lag and the contemporaneous and one-quarter lagged values of the remainder of the variables.[52] The coefficient estimates, Newey-West (1987) adjusted t-statistics, and adjusted R^2 values from these regressions are presented in Table 2.

Panel A shows that ΔPIM is related negatively to ΔSEN and ΔSTO, with the latter relation being statistically significant only at a 10% level.[53] Combined with our results from the earlier causality tests, this appears to tell us that greater activity by liquidity traders leads to a decline in trading costs. In economic terms, a one standard deviation increases in ΔSEN and ΔSTO lead to 0.22 and 0.25 standard deviation declines in ΔPIM and ΔFSI, controlling for the lagged price impact and the remaining explanatory variables. The full model explains roughly 16 percent of the variation in ΔPIM, which is significantly greater than the 5 percent explained by lagged price impact alone. The results in Panel B show that the ΔFSI is related positively to ΔSTO, ΔSEN, and negatively to ΔPIM. The negative relation between ΔFSI and ΔPIM is subsumed in the full model by ΔSEN and ΔSTO. Given our earlier finding that STO is caused by consumer sentiment, these relations underline the importance of uninformed trading activity on the intensity of information-based trading. One standard deviation increases in contemporaneous and lagged ΔSEN lead to increases of about 0.12 and 0.20 standard deviation increases in ΔFSI. The full model explains about 51 percent of the variation in ΔFSI, compared to 45 percent explained by lagged FSI alone.

---

[52] The marketwide aggregates for all of our variables except price nonsynchronicity are highly persistent. To rid our analysis from econometric issues that may arise, we conduct the analysis using first-differences.

[53] PIM, STO, and FSI are inextricably linked. For both PIM and FSI, we test specifications that include only lagged variables. This, however, does not quite get at the issue we are trying to address: for instance, is arbitrage activity higher or lower *when* share turnover is high? In order to address such questions, we also include contemporaneous realizations of the variables. In doing so, we admit that there may be simultaneity issues: the objective is to quantify the contemporaneous associations and interpret these associations based on the insights that we have obtained in the causality analysis previously presented.

These results are consistent with a world where a lack of participation by liquidity investors (as captured by low share turnover and consumer sentiment) precipitates an increase in trading costs (as captured by greater price impact of trading). This increase may be seen as a rational response by the market makers who realize that they are facing a more informed pool of investors due to the non-participation of liquidity traders. Effectively, the increase in price impact reduces the amount of private information that is incorporated into stock prices through informed trading, as a certain fraction of signals that may be profitable in lower trading cost regimes will no longer be worth the effort.[54] Put differently, our results follow from a game-theoretic setting where market makers know that a trade, if executed, is more likely to come from an arbitrageur and arbitrageurs know that a trade, if executed, will be less profitable since the market maker knows that the trade is more likely to come from an informed trader. This is an equilibrium which is optimal for both parties playing this game, but potentially suboptimal for the economy as a whole, as some information will not be incorporated into prices.

4.4.3 Determinants of Price Impact and Firm-Specific Information

To expand on the results from the analysis in the previous section, we also run Fama-MacBeth (1973) cross-sectional regressions of price impact and the share of firm specific information in price movements on several proxies that relate to differences in the information environment of individual firms as well as on natural logarithms of firm size (MV) and book-to-market equity (BM) and one-year lagged values of share turnover and price impact (LSTO and LPIM). The

[54] This hypothesis can be reconciled with the Kyle (1985) model by introducing a fixed cost for acquiring information in the profit function of the insider. Introducing such an information acquisition cost and imposing the insider profits to be greater than zero has the effect of rendering some signals unprofitable. In the single-auction case, for instance, the insider will not trade unless the signal exceeds the initial price of the security by at least two times the fixed information acquisition cost.

informational environment proxies are the number of analysts following the firm (NUM), dispersion of these analysts' one-year-ahead forecasts (FDISP), and an analyst coverage dummy that is equal to one if the firm is followed by at least one analyst and zero otherwise. Table 5 presents the coefficient estimates and t-statistics from these regressions for the sample period between 1976 and 2006 (determined by the availability of the I/B/E/S data) as well as for the recessions and expansions within this period.

Starting with the left panel, we first observe that trading costs are persistent, as indicated by the positive and highly significant coefficient on LPIM. Price impact is related negatively and significantly to firm size and BM, which is consistent with large-cap stocks and growth stocks receiving greater attention from liquidity traders (Barber and Odean, 2008). Perhaps in line with this argument, stocks with higher trading activity have lower trading costs. In economic terms, one standard deviation increases in firm size, BM, LSTO, respectively, are associated with a decline of 6.1, 0.7, and 1.0 in PIM. Controlling for all of the above variables, price impact is, on average, 0.8 lower for a stock that is covered by at least one analyst compared to one that has no analyst coverage. Recall that these figures represent the percentage increase in price in response to trading $1 million in December 2008 dollars, so the economic effect is quite significant. Last, the business-cycle decomposition of these coefficients suggest that cross-sectional differences in trading costs related to firm size, BM, and analyst coverage are much more pronounced during recessions compared to expansions. By contrast, whether share turnover is high or low is not a significant determinant of price impact during recessions. The right panel of the table contains the coefficient estimates from the regressions of PIM on our informational environment proxies without controlling for size or BM. These results indicate, in line with the above, that price impact is significantly greater or firms that are not covered. For stocks that are covered by at least one analyst, trading costs are declining in the number of analysts following the firm and increasing in the dispersion of these analysts' forecasts. In economic terms, a one standard deviation increase in FDISP (NUM) is associated with an increase (decline) of 0.2 (0.4) in PIM. The former finding highlights the importance of analysts in reducing the information asymmetry between investors, thereby leading to a decline in trading costs. The latter suggests that trading is costlier in the stock of firms whose future outlook is more uncertain, consistent with adverse selection problem increasing trading costs.

The business-cycle decomposition of these coefficients indicates that analyst coverage and its extent is a more significant determinant of trading costs during recessions than in expansions. As before, the relation between share turnover and price impact is flat during recessions. The dispersion of analysts' forecasts has a bigger positive coefficient (although less significant) in expansions, which is hard to interpret. It might well be the case that investors no longer care about how much analysts disagree when they are in the middle of a recession. The figure underneath Table 5 depicts the time-series behavior of the coefficient estimates from yearly cross-sectional regressions. Indeed, the coefficient for forecast dispersion tends to be less than or close to zero during recessions and peak in periods immediately before and after recessions. In general, the coefficient for analyst coverage starts to gets lower and price impact becomes more persistent about a year or two before the recession hits the economy.

We replicate the preceding analysis for the share of firm-specific information in price movements (FSI) and present our results in Table 6. The results indicate that FSI is related negatively and significantly to firm size, BM, and share turnover, consistent with the greater presence of liquidity traders in "attention-grabbing" stocks. Controlling for these variables, the fraction of price movements that is due to private signals is also lower for stocks with higher trading costs, which is consistent with our earlier evidence on high trading costs hindering arbitrage activity. For firms with similar size, BM, and trading activity, FSI is higher for stocks covered by at least one analyst. Over the business-cycle, cross-sectional differences in private information dissemination related to firm size and analyst coverage are more pronounced during recessions and those related to BM and share turnover are more pronounced during expansions, while price impact appears equally important in both economic states.

The results for the informational environment proxies, given in the right panel of the table, show that FSI is higher for firms with more dispersion in analyst forecasts and lower for stocks with analyst coverage, the more so the greater the number of analysts. The former finding is intuitively appealing, as a greater uncertainty of firm prospects would, potentially, give informed investors more rewarding profit opportunities, inducing private information generation. Although the latter finding goes against our results in the left panel, it is plausible if greater coverage by analysts is associated with higher noise trader activity when size and BM is not controlled for. We also note that, without these controls, PIM is positively related to FSI. This indicates that stocks with greater trading costs also tend to be the ones about which more private information is generated, which would follow from a positive relationship between adverse selection and trading costs. Lastly, while most variables display similar coefficients across the business-cycle, the effect of forecast dispersion on FSI appears to be greater during expansions, which is in line with our earlier evidence on high trading cost regimes (i.e. recessions) hindering the activity of informed investors.

4.4.4 Determinants of the Probability of Informed Trading

Panels A and B of Table 5 report the results from cross-sectional regressions of the probability of informed trading (PIN) during the recession of 1990-91 and the expansion of 1993-94. These periods are chosen because they fall within the time interval for which this microstructure measure is available and because they are in close proximity to one another. The set of explanatory variables, as before, include the natural logarithms of firm size and book-to-market (MV and BM), a coverage dummy that equals 1 for stocks covered by at least one analyst and 0 otherwise (DCOV), analyst forecast dispersion (FDISP), and lagged share turnover (LSTO). Each test is run using one cross-section, with PIN values defined as the average probability of informed trading for the two years included in each period and the explanatory variables computed at the end of the year immediately preceding the test periods (1989 and 1992).

The univariate regressions show that stocks with greater market capitalizations and lower BM have lower likelihood of informed trading. These findings are consistent with lower adverse selection for categories of stocks that likely receive greater uninformed investor attention, i.e. big stocks and growth stocks (Barber and Odean, 2008). In the full model, the relation between PIN and BM is insignificant for the expansion period and negative for the recession period, indicating that, controlling for size, firms that derive a greater fraction of their value from future growth options (low BM firms) face a greater adverse selection problem during economic downturns. This is plausible given the value of future growth options is more likely to be a greater source of uncertainty when business conditions are unfavorable.

Stocks that are covered by at least one analyst have a lower probability of informed trading, as indicated by the negative relation between DCOV and PIN. During the 1990-1991 and 1993-1994 periods, the average PIN for stocks covered by analysts are, respectively, 3.3 and 3.8 percent lower than that for stocks that are not covered. This is consistent with a role for analysts in attracting uninformed investor demand to the stocks they cover and at odds with the often assumed link between informed trading and analyst coverage (e.g. Shores, 1990; Brennan and Hughes, 1991). Further, for stocks covered by more than one financial analyst, the dispersion of analysts' forecasts is related positively to the likelihood of informed trading, indicating that greater uncertainty regarding firm prospects aggravates the adverse selection problem. The difference in PIN between the stock with the highest forecast dispersion and the lowest forecast dispersion is about 6.1 percent during the recessionary period and 9.6 percent during the expansionary period. The likelihood of informed trading relates negatively to trading activity, with PIN being lower for stocks with high share turnover: a one standard deviation increase in turnover corresponds to a 2 percent decline in PIN during both expansions and recessions.

Panel C of Table 5 reports the results from the cross-sectional regressions of the difference in PIN during the recessionary 1990-1991 period relative to the expansionary 1993-1994 period. Only those stocks that exist in both periods are included in the sample and the difference in PIN is calculated on a stock-by-stock basis as the recession PIN minus the expansion PIN. The average likelihood of informed trading is higher during recessions than during expansions by about 2 percent, consistent with our prior evidence that the adverse selection problem is more severe during recessionary periods. We find that the difference in PIN is increasing in firm size, share turnover, and analyst coverage and decreasing in BM. Hence, the stock of firms that may be expected to attract greater uninformed demand (big firms, growth firms, firms covered by financial analysts) are the ones whose likelihood of informed trading is the most affected during economic downturns. In economic terms, the change in PIN is higher by more than 4 percent for stocks that belong to the upper quartile of the size distribution compared to stocks that belong to the lower quartile in the expansionary 1993-1994 period. The corresponding figure is 1 percent for BM and 0.7 percent for LSTO. The difference in PIN is about 1.7 percent less for firms that are not covered by financial analysts compared to firms that are followed by financial analysts, perhaps highlighting the importance of analyst coverage in inducing uninformed trading activity.

An important question emerges. How does the increase in adverse selection during recessions affect the informativeness of stock prices? Are prices more informative because the greater adverse selection implies a more informed pool of traders? Are they less informative because greater trading costs hinder the activity of informed arbitrageurs? The next section addresses these questions.

4.4.5 The Implications of Trading Cost Regimes for Market Quality

Although the decline in the amount of firm-specific information disseminated into the market is potentially detrimental for the informational efficiency of the market as a whole, whether the prices would be more or less informative for corporate managers is not clear ex-ante. The prices may be more informative as some uninformed investors avoid trading during down markets, leaving behind a relatively more informed pool of investors. The reduction in uninformed trading would also mean lower liquidity, increased adverse selection, and higher trading costs, which may limit the activity of informed arbitrageurs and make prices less informative. To address this issue, we investigate whether recessions see a decline in the sensitivity of corporate investment to the information in prices, given our earlier evidence on the significant decline in firm-specific information dissemination during such states.

By way of formal analysis, we run Fama-MacBeth (1973) cross-sectional regressions of three corporate investment measures on normalized price (Q), price nonsynchronicity (FSI), a recession dummy that is equal to one if the period under consideration belongs in a NBER recession (RECD), and the interaction terms between Q and FSI and Q and RECD. Following Chen, Goldstein, and Jiang (2007), we control for the cash flow ratio (CF), inverse book assets (INVA), and cumulative abnormal return over the three years following the investment period (RET3). All explanatory variables are measured at the end of year *t*-1. The corporate investment measures analyzed are capital expenditures (CAPX), capital expenditures plus R&D (CAPXRND), and the change in assets (CHGASSET) in year *t*, all scaled by beginning-of-the-year assets. If prices are less (more) informative during recessionary periods, corporate investment should relate negatively (positively) to the interaction term between Q and RECD.

Table 8 presents the coefficient estimates along with their standard errors and the average adjusted R^2 values from these regressions. The results for all three corporate investment proxies point to the same prediction: prices are less informative for managers during and in the immediate aftermath of recessionary periods. To save space, we use CAPX in the following paragraphs to illustrate the magnitude of the effect. The first column of Panel A shows that, when the recession dummy and the interaction terms are left out of the equation, CAPX is related positively and significantly to Q with a coefficient estimate of 1.34 and a t-statistic of 6.61. This is consistent with the findings in the standing literature that corporate investment is positively correlated with stock prices. The second column reveals that the explanatory power of Q is subsumed once the interaction term between stock price and arbitrage activity, Q:FSI, is added. The 25th percentile, median, and 75th percentile values of FSI are 0.66, 0.81, and 0.93 (results not tabulated). This indicates that the sensitivity of investment to price for the median stock is 1.3 (-0.72+2.49 x 0.81). For the 25th and 75th percentile stocks, the corresponding figures are 0.9 and 1.6, indicating a 78% increase in the sensitivity of corporate investment to price as we go from the 25th percentile value of the FSI distribution to the 75th percentile value. Finally, as seen in the third column, when the interaction term between Q and RECD is added to the model, it obtains a highly significant negative coefficient and increases the average R^2 for the model from 0.54 to 0.61. The sensitivity of corporate investment to price for the median stock is 0.8 during recessions, which is about 62.5% lower than the sensitivity during expansions, 1.3. As seen in Panels B and C, our tests for CAPXRND and CHGASSETS yield very similar results.

The coefficient estimates for some of the control variables are also of interest. The positive relation between CF and corporate investment is consistent with the findings in the prior literature that firms with greater cash holdings tend to invest more (e.g. Fazzari, Hubbard, and Petersen, 1988). The negative coefficient estimate for RET3 is in line with the market mispricing argument which holds that firms whose stocks are overpriced tend to invest more (e.g. Loughran and Ritter, 1995; Baker and Wurgler, 2002; and Baker, Stein, and Wurgler, 2003).

4.5 Concluding Remarks

In this paper, we provide a detailed investigation of the time-series and cross-sectional patterns in the price impact of trading and the share of firm-specific information in price movements for stocks in the U.S. economy over an eighty-two-year sample period from 1926 to 2008. In doing so, we elaborate on the interrelationship among these variables and liquidity, trading activity, and the business cycle.

The main result is that, over long sample periods such as ours, variation in the activity of liquidity traders, which is often negligible in high-frequency daily/intraday analyses, plays a dominant role in determining the price impact of trading. During economic downturns, trades have a greater impact on prices, which may be linked to lower liquidity trader participation in stock trading in periods when households tend to substitute consumption for investment. Consistent with this argument, we show that (a) at the market level, price impact is caused by, and is negatively related to, the level of trading activity and consumer sentiment, and (b) at the individual stock level, the price impact of trades and the probability of informed trading are both significantly lower for stocks that tend to attract a high level of demand from uninformed investors (big stocks, liquid stocks, stocks covered by analysts).

The non-participation of liquidity traders and the accompanying increase in the price impact of trading appears to be detrimental for the informational efficiency of financial markets. We find, at the market level, the share of firm-specific information in returns is caused by, and is negatively (positively) related to, the price impact of trading (consumer sentiment), and demonstrate that the share of private news in returns is significantly lower during recessions. This reduction in information generation appears to have a material impact on the informativeness of prices for corporate managers. We finalize our study by showing that the sensitivity of corporate investment to the information in prices is significantly lower in recessions, underlining the importance of well-functioning, informationally-efficient securities markets.

Collectively, our findings are consistent with a world where (i) a reduction in uninformed trading activity leaves the market-maker facing a more informed investor pool, (ii) the market-maker rationally responds by increasing the trading costs, driving up the price impact of trades, and (iii) the increase in price impact of trades deters informed trading activity by reducing the profitability of a given signal. In the end, we face a game-theoretic scenario where market makers know that a trade, if executed, is more likely to come from an arbitrageur and arbitrageurs know that a trade, if executed, will be less profitable since the market maker knows that the trade is more likely to come from an informed trader. The result of the strategic interaction between arbitrageurs and market-makers is an equilibrium which is optimal for both parties playing this game, but potentially suboptimal for the economy. Indeed, our results indicate that the sensitivity of corporate investment to prices is significantly lower during recessions (when adverse selection is high and informed arbitrage is low) compared to expansions.

4.6 References

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

Baker, Malcolm, and Jeffrey Wurgler, 2002, "Market Timing and Capital Structure," *Journal of Finance* 57, 1 – 32

Baker, Malcolm, Jeremy C. Stein, and Jeffrey Wurgler, 2003, "When Does the Market Matter? Stock Prices and the Investment of Equity-Dependent Firms," *Quarterly Journal of Economics* 118, 969 – 1005

Barber, Brad M. and Terrance Odean, 2008, "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies* 21, 787 – 818

Brennan, Michael J. and Patricia J, Hughes, 1991, Stock Prices and the Supply of Information, *Journal of Finance* 46, 1665 - 1691

Brennan, Michael J. and Avanidhar Subrahmanyam, 1995, "Investment Analysis and Price Formation in Securities Markets," *Journal of Financial Economics* 38, 361 – 381

Chen, Qui, Itay Goldstein, and Wei Jiang, 2007, "Price Informativeness and Investment Sensitivity to Stock Price," *Review of Financial Studies* 20, 619 - 650

Chordia, Tarun, Richard Roll, Avanidhar Subrahmanyam, 2001, "Market Liquidity and Trading Activity," *Journal of Finance* 56, 501 - 530

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, "Is Information Risk a Determinant of Asset Returns," *Journal of Finance* 57, 2185 - 2221

Dow, James and Gary Gorton, 1997, "Stock Market Efficiency and Economic Efficiency: Is There a Connection?" *Journal of Finance* 52, 1087 - 1129

Easley, David and Maureen O'Hara, 2004, "Information and the Cost of Capital," *Journal of Finance* 59, 1553 - 1583

Fama, Eugene F. and James D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* 81, 607 - 636

Fang, Vivian W., Thomas H. Noe, and Sheri Tice, 2008, "Stock Market Liquidity and Firm Performance: Wall Street Rule or Wall Street Rules?" *Journal of Financial Economics (Forthcoming)*

Fazzari, Steven M., Glenn R. Hubbard, and Bruce C. Petersen, 1988, "Investment, Financing Decisions, and Tax Policy," *American Economic Review* 78, 200 – 205

French, Kenneth R. and Richard Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics* 17, 5 - 26

Granger, Clive W., 1969, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica* 37, 424 - 438

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

Kyle, Albert S., 1985, "Continuous Auctions and Insider Trading," *Econometrica* 53, 1315 - 1336

Morck, Randall, Bernard Yeung, and Wayne Yu, 2000, "The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?" *Journal of Financial Economics* 58, 215 - 260

Newey, Whitney K. and Kenneth D. West, 1987, "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703 – 708

Roll, Richard, 1988, "R^2," *Journal of Finance* 43, 541 – 566

Shores, D., 1990, "The Association between Interim Information and Security Returns Surrounding Earnings Announcements," *Journal of Accounting Research* 28, 164 - 181

Sims, C. A., 1972, "Money, Income, and Causality," *American Economic Review* 62, 540-552

Subrahmanyam, Avanidhar and Sheridan Titman, 1999, "Feedback from Stock Prices to Cash Flows," *Journal of Finance* 56, 2389 - 2413

4.7 Tables

Table 4-1 Descriptive Statistics, Correlations, and Variable Definitions

The top panel of this table presents the sources and definitions of the variables used in this paper as well as the time period for which data is available for each variable. The bottom panel reports time-series means and standard deviations, as well as the minimum, median, and maximum values for cross-sectional medians of stock returns (RET), share turnover (STO), price impact of trading (PIM), firm-specific information (FSI), market value of equity (MV), book-to-market equity (BM), analyst forecast dispersion (FDISP) and the number of analysts following the firm (NUM), as well as the time-series correlations between these variables. STO is defined, for each stock, as the quarterly share volume divided by the number of shares outstanding at the beginning of the quarter. PIM is defined as the quarterly average of the absolute daily returns divided by daily dollar volume. FSI is defined as one-minus the market model R^2 from the regressions of daily individual stock returns on contemporaneous and one-day-lagged value-weighted market return. BM is measured annually.

	Name of Variable	Source	Period	Variable Definition
FSI	Firm-Specific Information	CRSP	1926 - 2008	Share of Firm Specific Information in Price Movements
PIM	Price Impact of Trading	CRSP	1926 - 2008	Percentage Change in Share Price Due to Trading $1 Million
STO	Share Turnover	CRSP	1926 - 2008	Share Volume divided by Number of Shares Oustanding
MV	Market Value of Equity	CRSP	1926 - 2008	Share Price Multiplied by Number of Shares Oustanding
BM	Book-to-Market Ratio	COMPUSTAT	1950 - 2007	Book Value of Equity divided by Market Value of Equity
FDISP	Analyst Forecast Dispersion	IBES	1976 - 2006	Standard Deviation of Analyst Forecasts Scaled by the Mean Forecast
NUM	Number of Analysts	IBES	1976 - 2006	Number of Analysts Providing Earnings Estimates for the Firm

	Descriptive Statistics					Correlations					
	Mean	**StdDev**	**Minimum**	**Median**	**Maximum**	**PIM**	**STO**	**MV**	**BM**	**FDISP**	**NUM**
FSI	0.8798	0.0866	0.4829	0.9080	0.9708	**-0.2194**	**0.1184**	**-0.5721**	0.0340	**0.0980**	**-0.4910**
PIM	1.4688	3.3574	0.0064	0.5558	38.0952		**-0.2538**	**-0.2513**	0.0214	**0.0352**	**-0.1050**
STO	0.0969	0.0826	0.0019	0.0646	0.4151			**0.1303**	**-0.1807**	0.0454	**0.1270**
MV	10.4577	0.9607	8.0402	10.3712	12.8977				0.1223	**-0.5619**	**0.3804**
BM	-0.1013	0.2153	-0.4346	-0.1367	0.3279					0.0464	-0.0470
FDISP	0.0526	0.0131	0.0262	0.0513	0.0816						**-0.1267**
NUM	2.0117	1.0126	0.0000	2.0000	5.0000						

Table 4-2 Descriptive Statistics across the Business Cycle

The top panel of this table presents the time-series means, standard deviations, and the minimum, median, and maximum values of the cross-sectional medians for share turnover (STO), price impact (PIM), the share of firm-specific information in price movements (FSI), and quarterly stock returns (RET) for the expansions (left panel) and recessions (right panel) observed in the period between January 1926 and December 2008. The last columns contain the differences between the expansion and recession means for each variable and the standard errors of these differences. The bottom panel reports the mean values for the same variables in each expansion or recession period separately. The variables are as defined in Table 1.

	Expansions							Recessions					
	Mean	StdErr	Min	Med	Max	μ(R - E)		Mean	StdErr	Min	Med	Max	SE(R-E)
STO	0.1058	0.0062	0.0118	0.0696	0.4122	-0.0105**	STO	0.0953	0.0122	0.0234	0.0623	0.4217	0.0013
PIM	0.1404	0.3141	0.0061	0.1071	0.4795	0.0996**	PIM	0.2400	0.4640	0.0144	0.1266	1.0935	0.0034
FSI	0.8881	0.0031	0.0292	0.9131	0.9708	-0.0359**	FSI	0.8521	0.0044	0.0407	0.8805	0.9593	0.0012
RET	0.0156	0.0073	-0.2464	0.0072	1.4767	-0.0507**	RET	-0.0351	0.0078	-0.3943	-0.0444	0.9539	0.0027
	Begin	End	PIM	RSQ	STO	RET		Begin	End	PIM	RSQ	STO	RET
Expansion	1926.1	1926.3	0.0374	0.2861	0.2043	-0.1449	Recession	1926.3	1927.4	0.0450	0.0886	0.1590	0.0278
Expansion	1927.4	1929.3	0.0285	0.1272	0.2553	0.0421	Recession	1929.3	1933.1	0.3481	0.1667	0.0855	-0.0856
Expansion	1933.1	1937.2	0.1531	0.1964	0.0909	0.1220	Recession	1937.2	1938.2	0.2901	0.3304	0.0488	-0.1042
Expansion	1938.2	1945.1	0.2059	0.1911	0.0364	0.0102	Recession	1945.1	1945.4	0.0912	0.2154	0.0554	0.0952
Expansion	1945.4	1948.4	0.1554	0.2668	0.0456	-0.0242	Recession	1948.4	1949.4	0.2413	0.1996	0.0300	-0.0112
Expansion	1949.4	1953.2	0.1013	0.1439	0.0415	0.0175	Recession	1953.2	1954.2	0.1156	0.1442	0.0327	-0.0099
Expansion	1954.2	1957.3	0.0568	0.1346	0.0463	0.0212	Recession	1957.3	1958.2	0.0906	0.1889	0.0330	-0.0423
Expansion	1958.2	1960.2	0.0457	0.0995	0.0415	0.0332	Recession	1960.2	1961.1	0.0531	0.0999	0.0299	-0.0040
Expansion	1961.1	1969.4	0.0798	0.0939	0.0573	0.0044	Recession	1969.4	1970.4	0.1101	0.1446	0.0382	0.0010
Expansion	1970.4	1973.4	0.1124	0.0919	0.0602	-0.0151	Recession	1973.4	1975.1	0.5574	0.0830	0.0380	-0.0443
Expansion	1975.1	1980.1	0.1650	0.0641	0.0600	0.0341	Recession	1980.1	1980.3	0.0917	0.1188	0.0634	-0.0195
Expansion	1980.3	1981.3	0.0617	0.0753	0.0890	0.0449	Recession	1981.3	1982.4	0.1216	0.0797	0.0781	0.0025
Expansion	1982.4	1990.3	0.2750	0.0483	0.1114	0.0083	Recession	1990.3	1991.1	0.7968	0.0554	0.1061	-0.0131
Expansion	1991.1	2001.1	0.1942	0.0441	0.1806	0.0040	Recession	2001.1	2001.4	0.1258	0.0758	0.1742	0.0379
Expansion	2001.4	2007.4	0.0369	0.1109	0.4055	0.0085	Recession	2007.4	2008.4	0.0297	0.2259	0.2740	-0.1302

Table 4-3 Causality between Trading and Private Information

The four panels of this table presents the sum of squared residuals (SSE), F-statistics, and the corresponding p-values from Granger (1969) – Sims (1972) type causality tests ran between marketwide aggregates of the share of firm-specific information in price movements (FSI), price impact (PIM), share turnover (STO), and consumer sentiment index (SEN). The variables are as defined in the notes for Table 1. The tests use quarterly measurements of the listed variables and second order vector autoregressive models of the form:

Unconstrained Model: $Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + c_1 X_{t-1} + c2 X_{t-2} + e_t$

Constrained Model: $Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + e_t$

The null hypothesis is that the addition of the variable X, given in the second column of each panel, in the unconstrained model does not lead to a statistically significant increase in the explanatory power of the constrained model on variable Y, which is given in the first column of each panel. This hypothesis is tested with an F-test performed on the sum of squared residuals from each model.

Direction of Causality		GRANGER TEST STATISTICS		
To	**From**	**SSE**	**F-Statistic**	**p-value**
FSI	FSI	0.177		
	PIM	0.170	4.03**	0.02
	TURN	0.175	0.76	0.47
	SENT	0.171	3.74**	0.03
PIM	FSI	15.705	0.11	0.90
	PIM	15.705		
	TURN	15.239	3.22**	0.04
	SENT	14.764	6.73**	0.00
STO	FSI	0.148	1.01	0.36
	PIM	0.149	0.19	0.83
	TURN	0.149		
	SENT	0.145	3.03**	0.05
SEN	FSI	48.103	0.20	0.82
	PIM	48.158	0.08	0.93
	TURN	47.267	2.07	0.13
	SENT	48.192		

Table 4-4 Time-Series Regressions of Price Impact and Firm-Specific Information

The top panel of this table presents the coefficient estimates and Newey-West (1987) adjusted t-statistics from the regressions of the quarterly difference in PIM on contemporaneous and lagged differences in SEN, STO, and FSI, controlling for the one-quarter lagged level of PIM. The bottom panel gives the results from the regressions of the quarterly difference in FSI on contemporaneous and lagged differences in SEN, STO, and PIM, controlling for the one-quarter lagged level of FSI. The last column presents the adjusted R^2 statistics for each of the specifications. The variables are as defined in Table 1. Statistical significance at the 1%, 5%, and 10% levels are denoted by '**', '*', and '.', respectively.

	ΔSEN(t)	ΔSEN(t-1)	ΔSTO(t)	ΔSTO(t-1)	ΔFSI(t)	ΔFSI(t-1)	PIM(t-1)	RSQ
		-0.0533		-0.2500		-0.2080	-0.1202**	0.0591
		-1.3517		-0.6170		-0.3380	-4.2074	
	-0.1477**	-0.0754 .					-0.1153**	0.1176
	-2.9826	-1.6868					-3.9754	
ΔPIM			-2.7513*	-0.8328			-0.1147**	0.1134
			-2.1710	-1.6071			-5.2170	
					-0.2286	-0.3209	-0.1159**	0.0463
					-0.4019	-0.4201	-4.2711	
	-0.1250**	-0.0921*	-2.6581 .	-0.2276	0.4127	-0.1427	-0.1139**	0.1604
	-3.0359	-1.9515	-1.8698	-0.4929	0.6860	-0.2173	-5.1236	

	ΔSEN(t)	ΔSEN(t-1)	ΔSTO(t)	ΔSTO(t-1)	ΔFSI(t)	ΔFSI(t-1)	FSI(t-1)	RSQ
		0.0078**		-0.0552		-0.0179 .	-0.9044**	0.4790
		2.2957		-0.6559		-1.6574	-14.0948	
	0.0010	0.0102**					-0.9157**	0.4669
	0.3253	3.1420					-16.9894	
ΔFSI			0.2768**	0.0507			-0.8968**	0.4859
			4.4250	0.6101			-14.4533	
					-0.0013	-0.0200*	-0.9183**	0.4717
					-0.1325	-2.1628	-18.2361	
	0.0001	0.0098**	0.3084**	-0.0134	0.0077	-0.0158	-0.8771**	0.5117
	0.0280	2.9109	5.8482	-0.1547	0.9944	-1.3638	-14.4216	

Table 4-5 Cross-Sectional Determinants of Price Impact

The left and right panels of this table report the coefficient estimates and t-statistics from Fama-MacBeth (1973) regressions of individual stock-level price impact (PIM) on the natural logarithms of firm size and book-to-market ratio (MV and BM), the number of analysts following the firm (NUM) and the dispersion of these analysts' one-year-ahead forecasts (FDISP) an analyst coverage dummy (DCOV) that is equal to 1 if the firm is covered by at least one analyst and 0 if it is not, and the one-year lagged values of share turnover and price impact (LSTO and LPIM). MV is natural logarithm of the firm's market value of equity (number of shares outstanding times the share price) at the end of the previous year and BM is the book value of equity divided by the market value of equity. FDISP is the standard deviation of analyst forecasts divided by the mean estimate. The time-series plot of the coefficients obtained from each cross-sectional regression for FDISP, NUM, DCOV, LSTO, and LPIM are given in the figure below. Shaded regions correspond to years containing recessions determined by the NBER.

	Full Sample	Expansions	Recessions		Full Sample	Expansions	Recessions
MV	-2.9894	-2.7833	-3.7252	**FDISP**	0.6418	0.7129	0.3881
	-4.54	-3.87	-2.26		2.16	1.88	3.12
BM	-0.9440	-0.7948	-1.4768	**NUM**	-0.0524	-0.0374	-0.1060
	-2.81	-2.14	-1.85		-3.38	-4.55	-1.66
DCOV	-0.7601	-0.2193	-2.6913	**DCOV**	-7.7394	-6.7094	-11.4180
	-1.87	-0.51	-3.98		-4.43	-3.77	-2.33
LSTO	-3.0488	-3.9397	0.1329	**LSTO**	-2.3292	-3.3126	1.1829
	-2.21	-2.70	0.04		-1.95	-3.07	0.31
LPIM	0.3946	0.3899	0.4114	**LPIM**	0.4218	0.4140	0.4499
	8.55	7.48	3.84		8.84	7.76	3.95

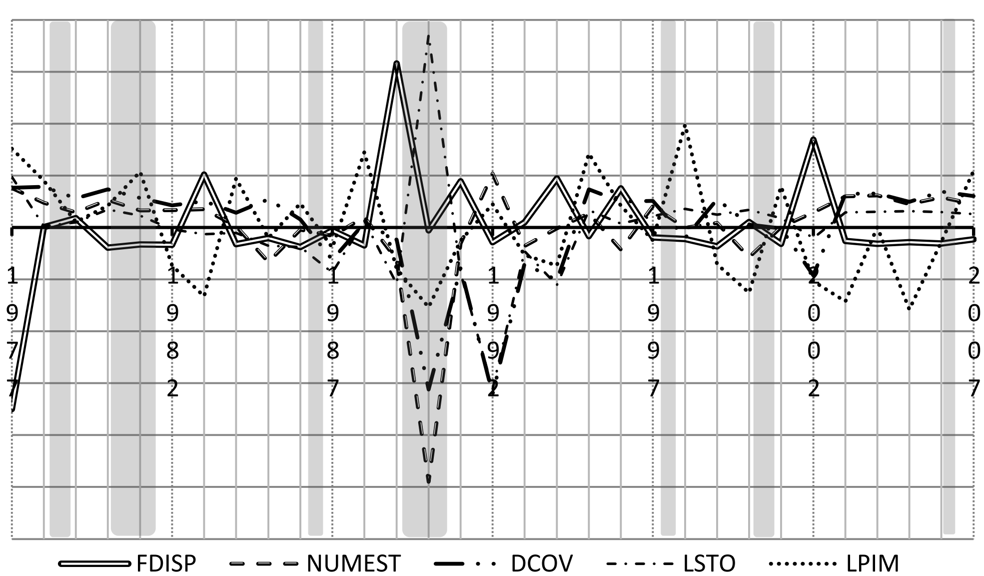

Table 4-6 Cross-Sectional Determinants of Private Information Dissemination

The left and right panels of this table report the coefficient estimates and t-statistics from Fama-MacBeth (1973) regressions of the share of firm-specific information in price movements (FSI) estimated at the individual stock-level on the natural logarithms of firm size and book-to-market ratio (MV and BM), the number of analysts following the firm (NUMEST) and the dispersion of these analysts' one-year-ahead forecasts (FDISP) an analyst coverage dummy (DCOV) that is equal to 1 if the firm is covered by at least one analyst and 0 if it is not, and the one-year lagged values of share turnover and price impact (LSTO and LPIM). The variables are as defined in the notes for Table 5. The time-series plot of the coefficients obtained from each cross-sectional regression for FDISP, NUM, DCOV, LSTO, and LPIM are given in the figure below. Shaded regions correspond to years containing the recession periods determined by the NBER.

	Full Sample	Expansions	Recessions		Full Sample	Expansions	Recessions
MV	-0.0379	-0.0357	-0.0461	**FDISP**	0.0574	0.0623	0.0400
	-14.20	-12.84	-6.93		5.89	5.23	3.22
BM	-0.0063	-0.0082	0.0004	**NUM**	-0.0073	-0.0068	-0.0090
	-1.85	-1.94	0.16		-13.64	-11.43	-8.90
DCOV	0.0073	0.0058	0.0129	**DCOV**	-0.0271	-0.0268	-0.0281
	2.42	1.61	2.48		-4.67	-4.04	-2.19
LSTO	-0.1915	-0.1970	-0.1716	**LSTO**	-0.1578	-0.1577	-0.1583
	-2.21	-1.79	-2.95		-2.31	-1.83	-2.67
LPIM	-0.0004	-0.0004	-0.0004	**LPIM**	0.0005	0.0004	0.0007
	-2.78	-2.22	-2.07		3.17	2.21	3.67

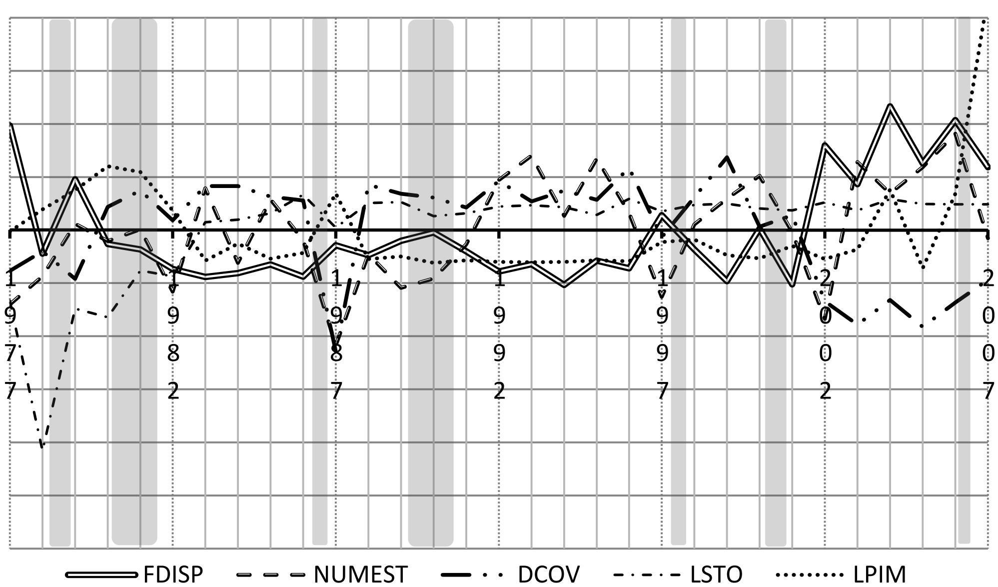

Table 4-7 Determinants of the Probability of Informed Trading

This table present the coefficient estimates and standard errors (in smaller font underneath from cross-sectional regressions of the probability of informed trading (PIN) estimates from Easley, Hvidkjaer, and O'Hara (2002) on the natural logarithms of market capitalization and book-to-market ratio (MV and BM), a coverage dummy that equals 1 if there is at least one analysts following the firm and 0 otherwise (DCOV), analyst forecast dispersion (FDISP), and lagged share turnover (LSTO). Panel A and B contain the results from the regressions of the average PIN for the 1990-1991 (recession) and 1993-1994 (expansion) periods on the set of explanatory variables measured at the end of 1989 and 1992, respectively. In Panel C, the difference in PIN between the recession period and the expansion period is regressed on the same set of variables measured at the end of 1989, controlling for the PIN for 1990-1991. The top row of each panel contains univariate regression results, while the bottom row presents the full model. Statistical significance at the 1%, 5%, and 10% levels are denoted by '**', '*', and '.', respectively.

Panel A: The Recession of 1990 - 1991					
	MV	**BM**	**DCOV**	**FDISP**	**LSTO**
Univariate Regressions	-0.0247**	0.0215**	-0.0333**	0.0021*	-0.6164**
	0.0007	0.0024	0.0041	0.0009	0.0626
	0.53	0.07	0.06	0.11	0.08
Full Model	-0.0239**	-0.0057**	-0.0103	0.0004	-0.1843**
	0.0009	0.0020	0.0064	0.0007	0.0595
Panel B: The Expansion of 1993 - 1994					
INT	**MV**	**BM**	**DCOV**	**FDISP**	**LSTO**
Univariate Regressions	-0.0234**	0.0238**	-0.0379**	0.0033**	-0.5923**
	0.0007	0.0023	0.0038	0.0009	0.0595
	0.52	0.09	0.09	0.13	0.08
Full Model	-0.0212**	-0.0002	-0.0187**	0.0018**	-0.1716**
	0.0009	0.0019	0.0061	0.0006	0.0570
Panel C: The Difference in PIN from 1990-1991 to 1993-1994					
	MV	**BM**	**DCOV**	**FDISP**	**LSTO**
Univariate Regressions	0.0147**	-0.0103**	0.0168**	-0.0019**	0.2011**
	0.0009	0.0018	0.0029	0.0006	0.0463
Full Model	0.0135**	-0.0017	0.0154.	-0.0016**	0.1121*
	0.0011	0.0018	0.0058	0.0006	0.0539

Table 4-8 Business Cycle Effects on Informativeness of Prices for Corporate Managers

Panels A, B, C of this table present the coefficient estimates, standard errors (in smaller font underneath), and adjusted R^2 statistics (in the last row) from Fama-MacBeth (1973) cross-sectional regressions of three corporate investment measures on normalized price (Q), price nonsynchronicity (INFO), cash flow ratio (CFX), inverse assets (INVA), future abnormal returns (RET3), a recession dummy that equals one in NBER identified recessions (RECD), and the interaction terms between Q and INFO and Q and RECD. The three corporate investment measures considered are capital expenditures (CAPX), capital expenditures plus R&D (CAPXRND), and the annual change in book assets (CHGASSET) in the investment year t, all scaled by the book value of assets as of the end of year t-1. Q is defined as the market value of equity plus the book value of debt scaled by the book value of assets at the end of year t-1. CF is the cash holdings divided by the book value of assets at the end of year t-1. INVA is the reciprocal of the book value of assets at the end of year t-1. RET3 is the three year market-adjusted return measured beginning from the end of year t. The remainder of the variables is computed as defined in the notes for Table 1. Statistical significance at the 1%, 5%, and 10% levels are denoted by '**', '*', and '.', respectively.

	PANEL A: *CAPX*			**PANEL B: *CAPXRND***			**PANEL C: *CHGASSETS***		
Q	1.3381**	-0.7206	-0.4609	1.5238**	-0.3581	-0.3302	5.5054**	2.7644	2.3483
	0.2023	0.6478	0.7396	0.2206	0.7342	0.8149	0.4920	2.2711	2.7198
INFO	-0.3763**	-2.1738**	-2.0213**	-0.3748*	-1.8841**	-1.9163**	0.4661	-1.4251	-1.9942
	0.1694	0.4706	0.5239	0.1962	0.5400	0.6058	0.5625	1.6114	1.8591
CFX	0.3203**	0.3165**	0.3198**	0.3921**	0.3851**	0.3898**	1.4941**	1.4867**	1.5074**
	0.0126	0.0127	0.0129	0.0138	0.0140	0.0141	0.0403	0.0407	0.0435
INVA	0.5802	0.8988	0.5911	-0.9455	-0.7861	-0.8798	7.7119**	8.7444**	9.6341**
	0.6221	0.6268	0.6505	0.6447	0.6514	0.6910	1.3375	1.3135	1.5053
RET3	-0.1298*	-0.1106	-0.1882**	-0.0588	-0.0444	-0.1246	-0.8537**	-0.8435**	-0.8372**
	0.0638	0.0655	0.0665	0.0705	0.0720	0.0737	0.1844	0.1884	0.2092
RECD			0.3368**			0.3859**			-0.3654
			0.0702			0.0789			0.2318
Q.INFO		2.4871**	2.1815**		2.2932**	2.1213**		2.7990	3.8969
		0.7349	0.8280		0.8220	0.9239		2.5791	3.0185
Q.RECD			-0.5418**			-0.5201**			-1.7739**
			0.1082			0.1186			0.3775
RSQ	0.5108	0.5444	0.6148	0.53845	0.56995	0.6356	0.5027	0.5324	0.5947

4.8 Figures

Figure 4-1 Time-Series Behavior of Price Impact, Share Turnover, and Firm Specific Information

This figure displays the time-series variation in marketwide aggregates of price impact (PIM), share turnover (STO), and the share of firm-specific information in price movements (FSI) over the period from January 1926 to December 2008. The gray-shaded regions mark recessions as reported by the National Bureau of Economic Research (NBER). On the price impact series, circular markers indicate peaks and triangular markers indicate troughs of the business-cycle.