Market Participant Behaviour
and
Equity Market Dynamics

By

Andrew Rhys Jackson

A dissertation submitted to the University of London

for the degree of Doctor of Philosophy

Department Of Finance

London Business School

University Of London

© September 2003

Abstract

This thesis consists of three studies examining the interaction between market participant behaviour and equity market dynamics.

Chapter 2 examines the relationship between sell-side analysts and investors in the presence of repeated interaction and trade generation incentives. I find that optimistic analysts generate more trade for their brokerage firms, as do high reputation analysts. Sell-side analysts face a conflict between telling the truth to build their reputation, and misleading investors to generate short-term increases in commission. When the investor is unsure of the analyst's type and when short sales constraints are present, I show using a simple model that equilibrium forecast optimism can exist, even when investment-banking affiliations are removed.

Chapter 3 examines the behaviour of individual investors. Behavioural models generally assume that investment decisions of irrational investors aggregate in a systematic way. Using a unique dataset of individual investor trades I find aggregate individual investor trades exhibit strong systematic patterns, including negative feedback trading and substantial persistence. In addition the trades of a large number of independent retail brokerage firms are highly contemporaneously correlated. However I do not find that the net trades of retail investors consistently predict future returns in a negative fashion. While small investors do act in a highly systematic fashion, their actions may not, at least in the short run, be classed as irrational.

Chapter 4 examines the behaviour of individual and institutional investors to assess whether there is evidence of noise trader risk in a broad equity market context. I find that increases in institutional participation are associated with future excess volatility and excess correlations between stocks. I also find evidence of a significant priced noise trader factor in returns. I conclude that there is evidence of noise trader risk, however this risk is not associated with individual investors, but instead with institutional investors. I argue that institutional frictions are a much more plausible source of noise trader risk than is individual investor sentiment.

Table Of Contents

Abstract 2

List of Tables 6

List of Figures 7

Acknowledgments 8

1. Introduction 9

2. Trade Generation, Reputation, and Sell-Side Analysts 13

2.1. Introduction 13

2.2. Related Literature 16

2.2.1 Sell-Side Analyst Behaviour 17

2.2.2 Reputation and Repeated Games 20

2.3. Data Description and Variable Definitions 21

2.3.1 Broker Market Share Data 22

2.3.2 Analyst Reputation Information 24

2.3.3 Investment Banking Affiliations 26

2.3.4 Analyst Forecast Information 26

2.3.5 Institutional Considerations 27

2.4. Empirical Analysis of Analyst Incentives 28

2.4.1 Analyst Reputation and Future Broker Market Share 28

2.4.2 Analyst Performance and Reputation 33

2.4.3 Analyst Optimism and Broker Market Share 37

2.5. Model of Reputation and Trade Generation 41

2.5.1 Description of Economy 41

2.5.2 Specification of Equilibrium 47

2.6. Extension: Tests of Model Predictions 55

2.7. Conclusion 59

2.8. Appendix 1: Analyst Variable Calculation 61

2.8.1 Measuring analyst accuracy 61

2.8.2 Measuring analyst optimism 63
2.8.3 Measuring Forecast Age 63
2.9. Appendix 2: Proof of Proposition 1 64
2.9.1 Investor's Optimal Beliefs 64
2.9.2 Investor's Desired Trade Size 67
2.9.3 Analyst's Optimal Strategy 68
2.9.4 Other Considerations 70
2.9.5 Derivation of Equilibrium Outcomes 70
2.10. Appendix 3: Proof of Proposition 2 71
2.11. Appendix 4: Proof of Proposition 3 74
2.12. Appendix 5: Cluster Adjusted t-statistics 76
3. The Aggregate Behaviour of Individual Investors 78
3.1. Introduction 78
3.2. Data Description 82
3.3. Systematic Patterns in Aggregate Individual Trading Behaviour 88
3.3.1 Market Level Flows 89
3.3.2 Cross Sectional Flows 94
3.4. Correlations Between Flows from Different Retail Brokerage Firms 103
3.5. Affect of Aggregate Systematic Patterns on Prices 111
3.5.1 Market Level Flows 112
3.5.2 Cross Sectional Flows 121
3.6. Conclusion 130
3.7. Appendix 1: Detrending of Market Level Flows 132
4. Noise Trader Risk Exists... but the noise traders are not who you think they are 134
4.1. Introduction 134
4.2. Simple Model of the Affect of Noise Traders on Market Dynamics 139
4.3. Data Description 147
4.4. Individual Participation and Stock Volatility 154

4.5. Individual Participation and Excess Correlation 159
4.6. Testing for a Noise Trader Risk Factor In Returns 165
4.7. Discussion and Interpretation 179
4.8. Conclusion 182
4.9. Appendix 1: Control Variable Discussion 183
4.10. Appendix 2: Derivation of Model Propositions 188
4.10.1 Equilibrium Price 188
4.10.2 Hypothesis 1-3 derivation 188
4.10.3 Volume Derivation 189
4.10.4 Lemma 1 Derivation 192
4.11. Appendix 3: Small Sample Distribution for Beta Test 192
5. References 194
5.1. Chapter 2: Trade Generation, Reputation, and Sell-Side Analysts 194
5.2. Chapter 3: The Aggregate Behaviour of Small Investors 199
5.3. Chapter 4: Noise Trader Risk 202

List of Tables

Table 2.1: Market Concentration in Trading Volume 22
Table 2.2: Summary Statistics 23
Table 2.3: Survey Data: Number of Rated Analysts per sector by year 25
Table 2.4: The Effect of Reputation on Future Broker Market Share 31
Table 2.5: The Effect of Accuracy on Analyst Reputation 36
Table 2.6: The Effect of Optimism on Broker Market Share 38
Table 2.7: Equilibrium Optimism Tests 58
Table 3.1: Summary Description Of Aggregate Data 84
Table 3.2: Market Share Of Retail Brokers By Size Group 87
Table 3.3: Descriptive Market Statistics For 2001 88
Table 3.4: Aggregate Weekly Flow Regressions 91
Table 3.5: Broker Type Regressions Of Aggregate Flows 92
Table 3.6: Cross Sectional Flow Regressions 98
Table 3.7: Broker Type Regressions of Cross Sectional Flows 99
Table 3.8: Broker-Pair Correlation Analysis 107
Table 3.9: Broker-Pair Correlation Analysis: Top 50 Stocks 108
Table 3.10: VAR For Market Returns And Flows 114
Table 3.11: VAR For Cross-Sectional Returns And Flows 123
Table 3.12: VAR For Cross-Sectional Returns And Flows by Size 124
Table 3.13: Portfolio Returns Based On Net Flows By Brokerage Type 128
Table 4.1: Summary Description Of Aggregate Data 148
Table 4.2.A: Equity Ownership By Investor Type 153
Table 4.2.B: International Comparison: Individual Investor Ownership 153
Table 4.3: Retail Participation and Future Stock Volatility 157
Table 4.4: Retail Participation and Future Stock Correlations 161
Table 4.5: Relationship Between Factor Time Series 166
Table 4.6: Test Results For Four-Factor Model 169
Table 4.7: Significance Tests For HRLR Beta 172
Table 4.8: Time Series Estimate Of Factor Risk Premia 175
Table 4.9: Two-Stage Estimates Of Factor Risk Premia 176
Table 4.10: Characteristics Explaining % Retail Share Of Turnover 187

List of Figures

Figure 2.1: Timeline Of Events In Model 46
Figure 2.2: Average Optimism Comparative Statics 52
Figure 2.3: Average Optimism Comparative Statics 54
Figure 3.1: Aggregate Market Share Over Time (By $ Value) 85
Figure 3.2: Aggregate Market Share Over Time (By # Trades) 86
Figure 3.3: Return Coefficients by Size Quintiles 101
Figure 3.4: Return Coefficients by Sub-Period 102
Figure 3.5: Correlation Map For Broker Net Flows – All Stocks 105
Figure 3.6: Correlation Map For Broker Net Flows – Top 50 Stocks 110
Figure 3.7: Market Level Net Flows For Full-Service and Internet Brokers 112
Figure 3.8: Market Accumulation Index 112
Figure 3.9: Annual Flow Coefficients 116
Figure 3.10: Flow Coefficients By Broker Type 116
Figure 3.11: IRF For Market Flows And Returns, K=0.0 119
Figure 3.12: IRF For Market Flows And Returns, K=0.01 120
Figure 3.13: Annual Flow Coefficients by Broker Type 125
Figure 3.14: IRF For Cross-Sectional Flows And Returns 129
Figure 4.1: Relationship Between RS, Sentiment Holdings and Mu. 146
Figure 4.2: Breakdown By Ownership For Listed Australian Equities 150
Figure 4.3: Aggregate Market Share Of Sampled Brokers Over Time 151
Figure 4.4: Individual Investor Participation For Average Stock 151
Figure 4.5: Empirical Distribution Betas And F-Statistics 174
Figure 4.6: Distribution Of Within-Group And Between-Group Trading 186

Acknowledgments

I owe my supervisor, Tim Johnson, a great deal of thanks for his valuable advice, encouragement, and support throughout my time at London Business School. Tim has been very generous with his time, has challenged (and hence greatly improved) my research, and overall has been an excellent adviser.

I also thank faculty who have been generous with their feedback and suggestions; notably Francisco Gomes, Viral Acharya, David Goldreich, Jan Mahrt-Smit, Henri Servaes, Art Kraft, and especially Elroy Dimson. Suleyman Basak, Raman Uppal, Narayan Naik, James Dow, and Anthony Neubeuger have also provided helpful encouragement and advice over the last few years.

My fellow PhD students have provided helpful feedback and friendship; notably Hongjun Yan, Stefano Rossi, Stefan Nagel, Subhra Tripathy, Purnendu Nath, Sergey Sanzhar, Daphne Lui, Sanjay Pareek, Sergei Daveydenko, and Ilya Strebulaev. I wish them all the best for the future.

From Australia, I would like to thank Les Balzer and Michael Sherris for their encouragement, and my industry contacts for generously providing me with the unique data used in this research, notably Nick Coles.

My family deserves a special mention for their help and encouragement throughout the years of study. I especially thank my parents and my grandparents for their continued belief in me, and their friendship over the years.

Last, but by no means least, I would like to thank my wife, Jenny. She has been a fantastic source of inspiration and motivation throughout my PhD. I thank her for small things, like volunteering to proofread my thesis and for making me smile when things sometimes didn't quite go to plan. I also thank her for larger things, such as her patience and understanding over the last few years, and her preparedness to accompany me to the other side of the world. I cannot imagine life without her, and look forward to the next stage in our life together.

1. Introduction

In an open capitalist economy financial markets, especially the equity market, serve the important role of channelling capital to its most productive use. In addition they enable investors to efficiently share risk and smooth consumption over time.

Equity markets comprise a large number of different participants including individual investors, institutions, regulators, brokerage firms, and company management. The behaviour of these participants can have important implications for the efficiency of financial markets. Changes in the relative importance of different groups over time, or changes in the incentives that different participants face can potentially lead to changes in market dynamics. Friedman (1995) notes that "if changing the cast of characters, or the way in which they perform their roles, alters the market's performance of its basic functions, then those changes can also alter the quantity and composition of the capital formation the economy undertakes, or the ways in which that capital is deployed".

Over time, there have been large shifts in the patterns of ownership in equity markets. Institutional investors have grown to control a substantial proportion of outstanding equities. In the US institutions controlled over 50% of the equity market in 2000 compared to only around 10% in 1950. Conversely the proportion of equities directly controlled by individual investors has fallen substantially. Recently however several developments have occurred that may again increase the proportion of equities directly controlled by individual investors. These include the rise of low-cost internet brokerage firms, the replacement of defined benefit with defined contribution pension funds, and reforms to social security systems around the world promoting individual choice. Direct individual participation in equity markets, as a percentage of the total population, has also grown dramatically over the last two decades in many developed economies. If individual investors behave differently to

institutions, then changes in the proportion of the market controlled by individuals may have important implications for market dynamics.

Another significant change over time relates to developments in the production and reporting of information on firm performance. Recently, following several large financial scandals, regulators have imposed numerous controls on sell-side brokerage analysts and on company information disclosures. Sell-side brokerage analysts occupy an important role in equity markets, acting as vital information intermediaries. They generate and disseminate private signals on firm performance allowing capital to flow to, or be removed from, different firms based on up-to-date and accurate information on their future prospects. If the incentives facing analysts lead these signals to be biased, then the flow of capital to its most productive source may be impeded, lowering economic growth.

This thesis examines the behaviour of three specific groups of economic agents to assess how their actions affect equity market dynamics. Chapter 2 examines the behaviour of sell-side broking analysts. Chapter 3 examines the behaviour of individual investors. Chapter 4 examines the interaction between individual investors and institutions.

In chapter 2 I examine the trade-generation and reputation-building incentives facing sell-side analysts. Using a unique dataset I find three highly robust empirical results. Firstly, optimistic analysts generate more trade for their brokerage firms. Secondly, high reputation analysts generate more trade for their firms. Thirdly, accurate analysts generate higher end-of-period reputations.

The sell-side analyst therefore faces a conflict between telling the truth to build her reputation versus misleading investors via optimistic forecasts to generate short-term increases in trading commissions.

Using a simple model I show that when the investor is unsure of the analyst's type and short sales constraints are present, then forecast optimism is likely to exist in equilibrium, even when investment-banking affiliations are removed. I also show

that investor irrationality is not required to generate this result. Subsequent empirical tests support two supplementary predictions from the model, namely that equilibrium optimism is higher when dispersion is higher, and when short-sales constraints are more likely to hold.

The conclusions of this chapter may have important policy implications given the recently proposed changes to the institutional structure of the brokerage industry. Namely, the separation of research and trading from investment banking is unlikely to eliminate optimistically biased forecasts in the longer term. One set of biased incentives (the incentive to generate investment banking business) may simply be replaced by another (the incentive to generate trading volume).

In chapter 3 I examine the aggregate behaviour of individual investors in the Australian equity market. Behavioural models in finance generally assume that the investment decisions of irrational investors aggregate in a systematic way. Using a unique dataset of individual investor trades I investigate the plausibility of this assumption.

I find that aggregate individual investor trades do indeed exhibit strong systematic patterns, including negative feedback trading and substantial persistence. These patterns hold for flows at both the cross-sectional and market levels. In addition, I find that the weekly cross-sectional net trades of a large number of independent retail brokerage firms are contemporaneously correlated to a remarkable extent. Thus the aggregation assumption appears plausible.

However, contrary to models that assume individual investors act in an irrational manner, I do *not* find that the net trades of retail investors consistently predict future returns in a negative fashion. In fact over the period 1991-2002 the net trades of full-service brokerage clients actually *positively* forecast future short-term market and cross-sectional returns. This result is consistent with aggregate price-pressure at the market level, and private information motivated trading at the cross-sectional level.

While small investors do act in a highly systematic fashion, their actions may not, at least in the short run, be classed as irrational.

Finally in chapter 4 I use the dataset of individual investor trades to examine whether there is evidence of noise trader risk in a broad equity market context.

I build a simple model of noise trader risk to motivate three empirical tests. Examining the data I find robust evidence of excess volatility and excess correlations between stocks, consistent with the presence of noise trader risk. Additionally factor-pricing tests reveal evidence of a significant priced noise trader factor in returns.

I find however that noise trader risk does *not* appear to be associated with individual investors as is commonly asserted, but seems instead to be associated with institutional investors. I argue that institutional frictions such as common investment strategies, performance related mutual fund flows, and manager career concerns are a much more plausible source of noise trader risk than is individual investor sentiment.

2. Trade Generation, Reputation, and Sell-Side Analysts

2.1. Introduction

Significant media, regulatory, and academic attention has recently been focussed on the conflicts of interest generated by the relationship between sell-side analysts and investment banking departments. One outcome of this controversy was a settlement between US regulators and ten top investment banks requiring "a clear separation of the research and investment banking divisions at [brokerage] firms"[1]. A key assumption underlying this settlement is that splitting analyst research and trading away from investment banking will significantly reduce conflicts of interest affecting analyst research, leading presumably to unbiased earnings estimates and better outcomes for brokerage clients.

However with the introduction of such an institutional structure, the incentive for sell-side analysts to generate investment-banking business is simply replaced by the incentive to generate trading commissions. Indeed such an incentive has already long-existed in the previous institutional structure, but has not been examined closely in the academic literature to date due to a lack of available data. This chapter provides strong empirical evidence that optimistic analysts generate higher trading volumes for their firm, providing an incentive for analysts to upwardly bias their forecasts and recommendations.

This incentive is limited in reality by analysts' concerns about their reputation. Since analysts interact with investors repeatedly, opportunistic behaviour may be

[1] For details of the US settlement see http://www.oag.state.ny.us/press/statements/global_resolution.html

curtailed by the threat of negative repercussions in the future. In this situation the analyst must trade off the short-term incentive to lie and generate more trade, against the long-term gains from building a good reputation. I present empirical evidence that analysts with better reputations generate significantly higher future trading volume for the brokers they work for. Thus analysts have a strong economic incentive to develop their reputations. I also empirically show that the market updates analyst reputations in a consistent way, with more accurate forecasters attaining higher end-of-period reputations.

A second aspect of the debate concerning analyst conflicts of interest is differing beliefs about the rationality of investors who receive analyst research. Several financial economists have questioned the appropriateness of significant regulatory intervention in analyst research production. For example Kent Womack states that "Just as consumers know to be somewhat sceptical of the commercials they see on television, so do professional investors know how to "de-bias" information they receive from analysts"[2]. For convenience throughout the chapter I will label this point of view as the "Neoclassical view".

An alternative view, which I will label the "Spitzer view"[3], is that investors (especially smaller investors) are unable to correctly de-bias analyst research and as a result are systematically misled by optimistic forecasts. For example the Wall Street Journal reports that brokers were "hyping stocks to win lucrative investment-banking work from corporate clients, and were misleading investors in the process"[4]. This view asserts that investors *are* likely to be fooled by analyst hype and therefore need some form of regulatory protection.

---

[2] Source: http://knowledge.wharton.upenn.edu/articles.cfm?catid=11&articleid=663 published 6-Nov-2002.

[3] Elliot Spitzer (New York Attorney General) led the high-profile investigations into conflicts of interest at Wall Street investment firms in 2001. These investigations led to the settlement between banks and regulators that required the separation of research and banking.

[4] Wall Street Journal, 15th April 2002, "Merrill Lynch Will Negotiate With Spitzer", Charles Gasparino.

At first glance, my empirical finding that optimistic analysts generate more trade appears to provide strong support for the Spitzer position. Indeed if analysts *were* fooling investors by issuing spurious buy recommendations and optimistic forecasts, this empirical result is exactly what we would expect to find (given the heavily right skewed distribution of broker recommendations). However, my second and third empirical findings, that investors pay more attention to high reputation analysts and update reputation based on past forecasting performance, are more consistent with the Neoclassical position.

To reconcile these apparently conflicting results, I develop a simple model describing the interaction between analysts and investors. I show using this model that my three empirical findings can be generated inside a fully rational framework after we allow for the presence of asymmetric information about an analyst's motivations. The model also generates two further comparative statics predictions, namely that consensus optimism will be higher when short-sales constraints are higher, and when dispersion is higher. These predictions are subsequently supported by the data.

Using this model I am also able to impose a particular type of investor irrationality consistent with the Spitzer view, namely that a proportion of investors naively believe that analysts' forecasts and recommendations are always truthful. Using this modification we can directly test the Neoclassical versus Spitzer views. The empirical results, showing that cross-sectional optimism levels do not increase as the percentage of small investors increases, support the predictions of the Neoclassical position, allowing us to distinguish somewhat between the two competing views.

This chapter adds to the empirical literature on the trade generation incentives facing sell-side analysts. Given the current settlement requiring brokerage firms to be split in two, these incentives are likely to become significantly more important going forward. In a policy sense, my empirical results suggest that optimistically biased forecasts are unlikely to disappear after the proposed changes. However, measures

that enable investors to cheaply monitor past analyst forecasting performance[5] may be an effective tool in ameliorating this effect. Such policies make the analyst's reputation more transparent and increase the implicit penalty for opportunistic behaviour.

In a broader sense, the chapter adds to the economics literature on the empirical importance of reputation formation. I show empirically that reputation matters in the market for professional advice. I also add to the behavioural / rationalist debate, by demonstrating how a simple model with rational investors and asymmetric information can be applied to generate apparently behavioural phenomenon. My empirical results do not support the conclusion that investors are stupid and need to be protected, but do suggest that moves to decrease the cost of acquiring information on analyst quality would be helpful.

The chapter is structured as follows. Section 2.2 discusses some related literature. Section 2.3 describes the data sample and defines variables used in the empirical analysis. Section 2.4 contains the central empirical results of this study, examining the interaction between trade generation, reputation, and optimism. Section 2.5 develops a simple model explaining how these results can be generated within a simple rational framework. Section 2.6 tests three supplementary predictions from the model. Section 2.7 concludes.

2.2. Related Literature

This chapter is related to existing research on institutional incentives facing sell-side analysts and also draws on literature examining reputation effects in repeated games. This section presents a brief review of the literature. The time-starved reader

[5] For example, the US settlement requires firms to make "certain analyses generated by an investment house... public within 90 days after the conclusion of each quarter. This will enable investors to compare and evaluate the performance of analysts from different firms and permit the market to generate objective rankings".

can however jump straight to the empirical results in Section 2.4 with little loss in continuity.

2.2.1 Sell-Side Analyst Behaviour

The issue of analyst optimism has been examined extensively, with many authors reporting that analysts' earnings forecasts are upwardly biased on average (eg O'Brien (1988), Lys and Sohn (1990), Brown (1993), Dugar and Nathan (1995)). Keane and Runkle (1998) challenge this result, arguing that correcting for write-downs and correlation between forecast errors eliminates the optimism bias. Gu and Wu (2003) also argue that apparent optimism can be generated by analysts' optimal reactions to skewed earnings distributions. While the latter papers dispute unconditional bias, in a number of conditional settings the bias appears to persist (eg when investment banking affiliations are present). Optimism bias is generally attributed to analysts' responses to institutional incentives including investment banking, management relations, and trade generation, which I discuss in turn below.

Investment Banking Biases: Past research demonstrates that stock recommendations made by affiliated brokers are significantly more positive than those from non-affiliates; eg Dugar and Nathan (1995), Lin and McNichols (1998) for SEOs, and Michaely and Womack (1999) for IPOs. Recently, analyst-research and investment-banking conflicts of interest have attracted intense regulatory scrutiny (Spitzer (2002)) leading to increased restrictions on analyst behaviour in the US (eg amendments to NYSE Rule 472 and NASD Rule 2210), and similar though less severe reactions in other countries such as the UK. Although evidence supporting such biases appears widespread, some authors have questioned whether investment-banking biases are generated merely by selection bias (Rajan and Servaes (1997)).

Management Relations: Other research attributes analyst optimism to attempts by analysts to cultivate management relations in order to improve access to private

information about the firm (Francis and Philbrick (1993)). Consistent with this theory, Das et al (1998) shows that optimism bias is higher for firms with harder to predict earnings. Also Lim (2001) shows that bias is worse for analysts with less experience, for whom management access is likely to be particularly important. Recent developments discouraging selective disclosure (eg Regulation FD in the US and equivalent rules in Australia and the UK) and preventing firms from 'barring' negative analysts may somewhat reduce the influence of the management relations effect on optimism going forward. Additionally Eames, Glover, and Kennedy (2002) dispute the importance of the management relations story, arguing that controlling for the level of earnings makes the previous research findings become insignificant.

Trade Generation: Several authors have recognised that analysts have incentives to generate trade. Hayes (1998) develops a model where analysts dedicate more resources to forecasting the earnings of stocks that are likely to perform well, since this increases future expected trading volume. Irvine (2003) tests this theory in event time using Canadian data in 1994 and shows that optimistic recommendations generate higher own-broker trading volume, but optimistic earnings forecasts do not. Aitken, Muthuswamy, and Wong (2001) find using Australian data that buy recommendations lead to higher market share in an event window around announcement, while sell recommendations actually lead to lower market share in the window. My research extends the empirical optimism/market-share result of these two papers by carefully controlling for both investment banking affiliations (which are highly important) and analyst reputation. I also demonstrate that the relationship holds for longer horizons. More importantly, my research extends this stream of literature by examining the trade-off that an analyst faces between generating more trading commission today versus generating a good reputation that enables the generation of more trade in the future.

Other related literature includes research on analyst reputation and on the forecasting ability of analysts.

Sell-side analyst reputation: Analyst reputation is typically measured via survey rankings. Stickel (1992) finds that the forecasts of top ranked (All-American) analysts are more accurate and have larger return impacts than other forecasts. Similarly, Chen, Francis, and Jiang (2003) find that return reactions are larger following forecast revisions from analysts with more accurate past forecasts and longer track records. Cooper, Day, and Lewis (2001) find larger return reactions following revisions from 'leader' analysts, defined as those who issue more timely forecasts. Leone and Wu (2002) show that higher rated analysts are less likely to be fired. Hong and Kubik (2003) find that accurate and optimistic analysts are most likely to get promoted, giving strong career-concern incentives for analysts to be optimistic on average. My work extends this research by demonstrating a clear empirical link between analyst reputation (ranking) and future own-broker trading volume. This link has, to my knowledge, not been established in previous research.

Usefulness of analyst forecasts: An extensive stream of literature exists examining the usefulness of analyst forecasts in predicting future returns. Dimson and Marsh (1984) present evidence of sell-side analyst forecasting ability together with a useful review of the earlier literature. They conclude that sell-side analysts exhibit a "small but potentially useful degree of forecasting ability". More recent evidence includes Womack (1996) who finds that changes in analyst recommendations have a significant immediate impact on stock returns when they are released. There is also evidence of post-event drift, especially following new sell recommendations where excess returns average -9.1% over the following six months. Barber, Lehavy, McNicols, and Trueman (2001) find that trading strategies based on consensus recommendations yield positive excess returns before costs. Conversely Barber et al (2003) find that during the period 2000-2001 stocks with the most favourable recommendations underperformed those with the least favourable by over 20% p.a. This chapter takes as given that some analysts have positive forecasting ability, consistent with the majority of the previous literature.

2.2.2 Reputation and Repeated Games

It is well known in game theory that the presence of repeated interaction between agents can act as an important disciplining device in preventing opportunistic behaviour[6]. Under certain conditions, agents have incentives to generate good reputations and take actions that would not be rational in a one-shot game. This concept has previously been applied to finance in a wide number of contexts.

Fama (1980) argues that agency problems between managers and shareholders are likely to be minimised due to managers' concerns for their reputation in the labour market. Eaton and Gersovitz (1981) show how reputational concerns can lead to repayment of sovereign debt. John and Nachman (1985) model how reputation can resolve under-investment in the presence of risky debt, while Diamond (1989) shows how reputation can resolve risk-shifting problems.

Leach (1992) models takeovers and finds that bidders who are likely to participate repeatedly try to develop a reputation for having low valuations. Benabou and Laroque (1992) model market manipulation by newsletter writers and show that the manipulator will be truthful in order to build a reputation but will lie when the incentive to do so is sufficiently large. Finally Gomes (2000) shows that owner-managers may develop a reputation for treating minority shareholders well, allowing outside equity to be raised in countries with poor explicit corporate governance mechanisms.

Reputation effects have also been examined in experimental settings that are quite relevant to the present setting. DeJong, Forsythe, Lundholm, and Uecker (1985) find evidence of reputation building over several periods of trade in simulated markets. They also find that the ability of the principal to audit ex-post the actions of

---

[6] For seminal papers on reputation in finitely repeated games see Kreps and Wilson (1982) and Milgrom and Roberts (1982).

the agent significantly reduces the level of cheating especially when the results are publicly disclosed. In our setting, the ability of investors to cheaply monitor the past recommendations of analysts is important. Several of the policy measures recently introduced encourage analysts to transparently report information on the success of past recommendations[7]. This is likely to strengthen the reputational considerations of analysts.

This chapter applies the basic insights from previous research on reputation effects to a sell-side analyst setting in order to assess how an analyst's desire to generate a reputation for forecasting ability can reduce her level of opportunistic behaviour.

2.3. Data Description and Variable Definitions

This study uses extensive data from the Australian equity market over the period January 1992 to December 2002. Only stocks included in the main benchmark index (the All Ordinaries index to 2000 and the ASX300 subsequently) are used, giving between 250 and 350 stocks in each weekly cross-section, and capturing around 88% of total market capitalisation. Non-index stocks are excluded due to extreme illiquidity and to ensure the results are not driven by an economically insignificant section of the market.

In the empirical analysis, two large unique datasets describing broker market shares of trading volume and analyst rankings are combined with the IBES and SDC Platinum databases. I briefly describe the four datasets below.

[7] See Boni and Womack (2002) for a description and discussion of the new measures introduced by the NYSE and Nasdaq.

2.3.1 Broker Market Share Data

Detailed data on the trades generated by 23 institutional brokerage firms was collected for the period January 1992 to December 2002 for all stocks in the Australian market. The dataset contains weekly information on the aggregate trades of all clients of each broker. The information available is the number of trades, value traded, and average price of all buy and sell trades executed by each broker during the week, for each stock. The data is proprietary and was supplied by an Australian broker. The information used in this study is observable in real-time on a trade-by-trade basis by some market participants (brokers) but is not available to investors in general.

Table 2.1 outlines the concentration of market share for the Australian brokerage industry. Market share is strongly concentrated in the top ten brokers, however within these ten the market shares are relatively evenly distributed. Table 2.2 presents summary statistics describing the market share dataset, showing the average market share over the full period, the number of years of data available, and the total number of trades for each of the 23 brokers.

Table 2.1: Market Concentration in Trading Volume

This table presents summary statistics describing market concentration in the Australian broking industry. The % of total volume accounted for by the top 5, the top 10, and all 23 broking firms is calculated annually.

Year	% Mkt Share by Top 5 Brokers	% Mkt Share by Top 10 Brokers	% Mkt Share by 23 Sample Brokers
1993	59.6%	81.4%	85.1%
1994	60.5%	81.3%	84.7%
1995	60.9%	81.0%	83.9%
1996	59.4%	86.8%	93.8%
1997	53.4%	86.0%	95.2%
1998	51.5%	86.3%	95.1%
1999	51.1%	85.6%	93.9%
2000	48.0%	82.1%	90.4%
2001	50.1%	82.6%	90.6%
2002	49.5%	82.1%	89.6%
Average	54.4%	83.5%	90.2%

Table 2.2: Summary Statistics

This table presents summary statistics for the 23 brokers in the sample. Broker names are excluded for confidentiality reasons. The first column contains the number of years of market share data available for the each broker, the second column the number of years for which the broker supplies estimates to IBES. The third column is the number of trades by that broker in the sample. The fourth column is the average market share of the broker over the full sample period. The fifth column counts the number of times the manager acted as a lead manager in an IPO for a stock that was included in the index during the sample period. The sixth column counts co-manager participation in IPOs. The last two columns count lead and co-manager participation for SEOs.

Broker #	Full Years in MktShare Sample	Full Years in IBES Sample	Number of Trades by Broker	Average Annual Market Share	Lead Manager IPO	Co-Manager IPO	Lead Manager SEO	Co-Manager SEO
1	10	10	10,448,204	8.3%	31	13	103	5
2	10	10	10,355,971	11.9%	26	8	86	5
3	10	10	6,588,775	7.9%	33	12	206	5
4	10	10	6,116,852	8.5%	22	6	48	3
5	10	10	5,947,764	13.5%	15	8	35	2
6	10	10	5,741,376	8.2%	23	18	31	4
7	10	10	5,449,647	8.8%	19	23	45	7
8	10	10	5,354,301	7.7%	12	8	51	5
9	6	10	4,898,021	7.3%	17	1	20	2
10	10	5	3,118,767	1.2%	13	7	22	3
11	4	4	2,048,482	1.0%	1	1	2	1
12	10	5	1,338,699	1.3%	8	2	21	2
13	8	8	1,321,021	2.0%	8	0	13	0
14	6	4	1,266,551	1.7%	6	5	28	1
15	10	4	1,257,281	0.3%	0	1	5	0
16	10	5	1,155,168	0.2%	5	4	15	0
17	10	6	1,079,777	0.2%	2	10	38	5
18	10	7	959,765	0.2%	2	6	6	0
19	10	10	824,243	2.2%	4	5	21	1
20	3	3	688,263	0.4%	0	2	0	0
21	10	5	337,776	0.1%	1	12	15	0
22	10	2	157,428	0.0%	0	0	0	0
23	6	3	145,870	0.1%	0	2	1	0

2.3.2 Analyst Reputation Information

Analyst reputation was measured using rankings from the leading Australian analyst survey, the East Coles Survey. This is the Australian equivalent of the US Institutional Investor survey and is conducted over a large number of buy-side institutions once each year[8]. Survey respondents rate analysts on the quality of research, level of service, and overall general satisfaction. Survey ranking data was available annually from 1994 to 2002.

The data available from the East Coles Survey has a significant advantage over that from the Institutional Investor surveys, in that it provides significantly more detailed ranking information on analysts. For each industry (and in some cases stocks) I am able to observe the ranking for *every* analyst in that sector (eg ranking analysts down to the 56th ranked small-cap analyst in 2000). This contrasts with the Institutional Investor survey where only the top 3 'All-American' analysts (and up to 4 runners up) are observable for each sector. The extra information allows significantly more precise measurement of reputation than has previously been available. Table 2.3 presents summary statistics showing the number of analysts rated in each sector or stock. Industry definitions change over time, and the same analyst may appear in several industry categories with a different rating. For the empirical tests if an analyst is ranked in two sectors I use the rating of the analyst appropriate to the stock in question.

[8] Results of the survey are published in July each year. Rankings in the survey are a weighted average of institutional responses, with the weights based on the assets under management of the respondent. Equal weighted rankings are also available, but are not used in this study. Analyst rankings from these surveys were hand-matched each year to the analyst and broker code in the IBES database using the confidential broker translation file kindly provided by IBES.

Table 2.3: Survey Data: Number of Rated Analysts per sector by year

This table contains summary statistics describing the analyst survey data. The table lists the number of rated analysts in each sector or stock category for each year of the sample.

Sector	1994	1995	1996	1997	1998	1999	2000	2001	2002	Total
Alcohol	13									13
Alcohol & Tobacco				13	16	19	11	17	18	94
Banks	14	15	15	16	19	19	20	21	24	163
Base Metals							24	25		49
BHP	13	11	12	14	11	12				73
Brewing	14	13								27
BTR	12	11								23
Building Materials	12	13	12	14	13	12	14	12	11	113
Chemicals	12	11	11	10	13	19				76
Chem & H'Care							18			18
Commodities	12	12	13	16	16	13	17	21	15	135
RIO	17	17	15	14	14	14				91
Developers & Contractors	9	11	9	11	8	9	15	14	14	100
Diversified Financials									15	15
Diversified Industrials				19	18	19	19	15	13	103
Diversified Resources								20	22	42
Energy				21	13	14	14	13	15	90
Engineering	9	12	10	9	17	16				73
Food & Household	14	17	18	17	19	15	15	14	13	142
Gold	20	27	38	34	25	21	20	16	17	218
H'Care & Biotech								15	17	32
Hotels									10	10
Infrastructure & Utilities					12	18	14	18	13	75
Insurance	10	9	9	14	15	16	17	15	16	121
Investment & Financials								18		18
Media	16	17	17	19	15	17	26	25	27	179
Metals & Mining									22	22
MIM	19	14	11	25	27	20				116
Oil & Gas	18	18	13							49
Paper & Packaging	16	19	15	15	16	16	13	11	13	134
PDP	11	14	14	11	14	14				78
Property Trusts	19	26	24	36	32	37	35	23	28	260
Retail	15	14	15	16	16	15	13	15	14	133
SRP	13	13	10	12	14	15				77
Small Cap	23	44	48	52	45	38	56	34	29	369
Solid Fuels	10	10	14							34
Technology							40	27		67
Telecoms		17	13	14	16	18	25	30	25	158
Tobacco		13	10							23
Tourism & Leisure		11	9	15	12	13	10	10		80
Transport	12	14	12	15	16	12	11	12	16	120
Wine		13	10							23
WMC	15	16	16	15	15	12				89
Total	368	452	413	467	467	463	447	441	407	3925

2.3.3 Investment Banking Affiliations

Data on investment banking affiliations was obtained from the SDC Platinum Global New Issues database. For each IPO and SEO deal, the lead and co-manager was identified and hand matched to the brokers in my sample. Summary statistics for lead and co-manager status for each of the brokers are given in Table 2.2. Investment banking affiliation status is a particularly important control variable in this research since the link between optimism and affiliation is well known from previous research (Michaely and Womack (1999) and Lin and McNichols (1998)).

2.3.4 Analyst Forecast Information

Analyst earnings forecasts and recommendations were obtained from the detail history IBES database from 1992 to 2002. The confidential broker translation files were used to link the IBES data to the turnover data for my sampled brokers. Earnings forecasts were available for 87.6% of the stocks in the average cross-section. The average level of analyst coverage was 14.3 for Top-20 stocks, 11.8 for stocks ranked 21-50, 8.4 for stocks ranked 51-100, and 4.1 for Ex-100 stocks. Recommendation data was also obtained from IBES, but was only available for a subset of the stocks in the sample and for a reduced time period (1994-2002). The IBES forecast data was used to calculate several measures of accuracy, optimism, and timeliness for each analyst-stock pair in the data. Appendix 1 details the construction of these measures. Each measure calculates the relative performance of the analyst versus the other analysts covering the same stock in order to control for changes in the information environment over time for a given stock.

2.3.5 Institutional Considerations

Finally, before proceeding to the empirical analysis, I briefly discuss two institutional considerations relevant to the study.

Firstly, sell-side analysts are employed by brokerage firms who provide analyst research to their clients. Clients do not pay directly for the research product, but pay indirectly via brokerage commissions. When the client receives a research report that induces them to trade, they are not obliged to deal through the broker that provided that research. Many institutional investors construct a panel of brokers at the start of a given period in deciding how brokerage will be allocated over the subsequent period, in return for access to analyst research. This arrangement may make it difficult to empirically associate brokerage firm volume with analyst actions for a given stock. Nonetheless Section 2.4 demonstrates that despite this effect, strong relationships do exist in the data. Informal discussions with several fund managers in the Australian market reveal that they often *do* trade with the broker whose analyst has provided them with an influential recent report on a stock, in order to maintain a good relationship with the analyst.

Another important consideration is that corporate finance incentives may be significantly larger than trade generation incentives over the period considered. For example Hayward and Boeker (1998) report that corporate finance revenues account for four times equity commission revenues in a typical bank at that time. To control for this effect, I include corporate finance controls in my analysis. In the future if corporate finance and equity trading departments are split then trade generation incentives are likely to significantly increase in importance.

2.4. Empirical Analysis of Analyst Incentives

In this section I empirically examine the relationship between reputation, optimism, and broker market share. Specifically I examine three empirical hypotheses. Firstly, do higher reputation analysts generate higher broker market share? Secondly, does better forecasting performance lead to improved analyst reputation? Thirdly, do optimistic analysts generate higher trading volume for their brokerage firms?

The first two hypotheses follow naturally from the Neoclassical view, where investors optimally update their beliefs about analyst ability based on past results, and also pay more attention to analysts with better reputations. The third hypothesis follows from the Spitzer view whereby investors are fooled by optimistic analysts and trade in response to biased positive forecasts. I now proceed to empirically examine these three hypotheses in turn[9].

2.4.1 Analyst Reputation and Future Broker Market Share

Firstly I examine the relationship between an analyst's reputation (the independent variable) and the *future* market share of the brokerage firm that employs this analyst (the dependent variable). The link between reputation and future market share is important to establish as it provides empirical evidence that reputation is a relevant objective for the analyst. Without this linkage there is little possibility that reputation could act as a disciplining mechanism against opportunistic behaviour by the analyst.

Analyst reputation is measured in two ways. The first set of regressions assign a percentile ranking to the analyst based on their survey ranking, with the top analyst

[9] The model subsequently developed in Section 2.5 derives these three hypotheses explicitly.

receiving a ranking of 1 and the lowest a ranking of 0. The second set of regressions uses a dummy variable for reputation which equals 1 if the analyst is a 'top 3' ranked analyst, and 0 if the ranking is outside the top three. In both cases, reputation is measured in year t-1. The relationship is analysed for each analyst-stock-year unit in the sample, subject to the conditions that the stock exists in the index for the full year, the analyst covers the stock over the year, and a survey ranking exists for the analyst[10].

To isolate the effects of reputation, two groups of control variables are also included. The first control variable is the size of the broker, since larger brokerage firms are likely to have higher market share due to higher resource levels (eg sales staff, research budgets, distribution infrastructure). Market-wide broker market share across all stocks in a given year is used as a proxy for broker size[11].

The second group of control variables captures investment-banking relationships between the broker and the stock. Ellis et al (2000, 2002) show that broker market share is influenced by investment banking relationships for three months after initial public offerings (IPOs) and secondary equity offerings (SEOs). To control for these effects I include six dummy variables. The first and second dummies capture whether the broker has acted as either the lead or co-manager in an SEO for the stock in the current year. The third and fourth dummy variables measure lead and co-manager status over the last three years. The fifth and sixth dummy variables measure lead and co-manager status for IPOs in the last three years[12].

---

[10] Where IBES data is available for an analyst, but rating information is missing, I drop the analyst from the analysis. Including these analysts and giving them a zero-rating slightly strengthens my conclusions.

[11] Previous research uses the number of analysts employed by the brokerage firm as a proxy for broker size (eg Irvine (2001), Hong and Kubik (2003)). Using this proxy does not affect the reputation result but results in a somewhat less effective broker size control for my data sample.

[12] There are no IPOs in the current year since a full year's market share is required for the firm to enter the sample.

Table 2.4 presents the results of the pooled regressions with future broker market share as the dependent variable. T-statistics are calculated firstly using White (1980) robust standard errors to control for heteroskedasticity, and secondly using the Williams (2000) robust variance estimate to adjust for within-cluster correlation in residuals induced by the same analyst covering several stocks in a given year. Appendix 5 discusses the calculation of the cluster-adjusted standard errors.

We can see from Table 2.4 that analyst reputation in the previous year has a highly significant *positive* impact on broker market share in the following year with a t-statistic of 7.8 using the White adjustment, and 5.4 using clustering. This confirms the first hypothesis that future trading volume is increasing in the reputation of the analyst. The significance of the reputation variable is higher for larger (Top 100) stocks as shown on the right-hand side of the table. This relationship between analyst reputation and future own-broker volume has, to my knowledge, not been previously established in the literature.

Table 2.4 also reveals that larger brokers have higher market shares, as we might expect. Additionally, acting as a lead manager for an IPO or an SEO significantly increases the market share of the broker in that stock over the following three years, by 6.3% for SEOs and 8.7% for IPOs. Conversely acting as a co-manager has no significant affect on market share in the following three years.

Table 2.4: The Effect of Reputation on Future Broker Market Share

This table analyses the impact of analyst survey ranking in year t-1 on broker market share in year t. 'Analyst Ranking' is a percentile ranking of the analyst relative to other analysts covering the stock (1=highest ranked, 0=lowest ranked). 'Top 3 Analyst' is a dummy variable equalling 1 if the analyst is ranked in the top 3 for that stock. 'Broker size' is the total market share of the broker in year t across all stocks. 'SEO Lead' is a dummy variable indicating whether the broker has acted as lead manager in that stock in the current year or in the previous 3 years respectively. 'SEO CoMgr' is a dummy variable indicating whether the manager acted as a co-manager. 'IPO Lead' and 'IPO CoMgr' are defined in the same way. Analyst-Stock-Years are used as the unit of analysis, with both robust and cluster adjusted t-statistics presented. The dependent variable is the market share of the broker in the stock over a given year. Only stocks where the broker has an analyst covering the stock are included (and where a ranking for the analyst exists). Results are presented for all stocks and for large (top 100) stocks separately.

	Broker Market Share (t)											
	All Stocks						Large Stocks (Top-100)					
	Coefficient	White t-stat	Cluster t-stat	Coefficient	White t-stat	Cluster t-stat	Coefficient	White t-stat	Cluster t-stat	Coefficient	White t-stat	Cluster t-stat
Constant	0.007	3.75	2.69	0.016	9.26	6.15	-0.005	-2.32	-1.76	0.005	2 85	2.17
Analyst Ranking (t-1)	0.020	7.78	5.43				0.028	10.23	8.17			
Top 3 Analyst (t-1)				0.015	9.23	5.81				0.018	9.93	8.03
Broker Size (t)	0.690	29.83	20.23	0.700	31.13	21.37	0.813	31.60	24.33	0.839	33.45	26.10
SEO1_Lead	0.037	3.44	3.52	0.037	3.45	3.53	0.042	3.32	3.37	0.042	3.31	3.37
SEO1_CoMgr	-0.003	-0.22	-0.22	-0.004	-0.30	-0.30	-0.005	-0.41	-0.41	-0.005	-0.42	-0.42
SEO3_Lead	0.063	9.25	9.14	0.063	9.22	9.13	0.044	6.16	5.91	0.044	6 20	5.97
SEO3_CoMgr	0.014	1.34	1.34	0.015	1.42	1.42	0.010	0.97	0.97	0.011	1.07	1.06
IPO3_Lead	0.087	6.66	6.68	0.087	6.66	6.68	0.036	2.90	2.91	0.035	2 88	2.89
IPO3_CoMgr	-0.009	-1.78	-1.75	-0.010	-1.77	-1.75	-0.007	-1.38	-1.36	-0.006	-1.23	-1.22
Observations	10,717			10,717			5,858			5,858		
Adjusted R2	18.74%			18.91%			27.33%			27.53%		

This result is consistent with Ellis et al (2000, 2002) and additionally demonstrates that the relationship persists for an even longer horizon than the 3-months that they document.

In terms of economic significance, Table 2.4 shows that a top 3 ranked analyst increases broker market share by 1.8% for a Top 100 stock. In this market the typical institutional trading commission rate is around 15bp and a typical analyst covers around 7.1 stocks in her given sector. For a large sector such as banks, the additional 1.8% market share equates to around $A6.88m extra commission revenue per annum for a top-3 analyst relative to a non-top-3 ranked analyst for turnover levels equal to those of the last year of the sample[13]. This revenue is *incremental* to the revenue generated by simply covering a given stock. Coverage alone generates additional market share of 2.4% in my data sample[14]. The effect of reputation on future trading commissions appears economically large.

To ensure the reputation-market share relationship is robust I run the following checks. Firstly, running the regression separately for each year from 1995 to 2002 yields a positive coefficient on the survey-ranking variable in every single year. The Fama-Macbeth coefficient for annual cross-sections on this variable is 0.027 with a standard error of 0.004 (t-statistic of 6.12), which is highly significant. Secondly, I redo the analysis using reputation in year t, rather than reputation in year t-1 as the independent variable of interest. This increases the number of observations by around 13% (due to the inclusion of an extra year), and yields coefficients slightly higher and more significant than using reputation measured at year t-1. Thirdly, I transform the market share variable using a log specification as in Khorana and

[13] To calculate this result I take total dollar trading volume (two-sided) over the last year of the sample in index bank stocks ($233bn) then multiply by 1.8% and by a commission rate of 15bp to get $6.88m. The equivalent result for some other sectors is media ($2.79m), healthcare ($0.94m), telcos ($1.98m), materials ($5.98m), listed property trusts ($1.61m).

[14] Irvine (2001) shows that using Canadian data for 1994, coverage alone generates 3.8% of incremental market share.

Servaes (2002). This significantly improves the normality of the residuals, and increases the fit of the regression (eg the R2 increases from 18% to 28% in the first regression). The transformation also significantly increases the t-statistics on the reputation variable however the economic interpretation of the coefficient is less transparent. Fourthly, I calculate broker-year clusters rather than analyst-year clusters to allow for common firm-wide shocks to market shares. I also adjust for firm-year clusters to account for any negative autocorrelation due to the adding up constraint of broker market shares. I then exclude negative earnings stocks to check that write-offs do not affect the result. Finally, I redo the analysis using only the largest 9 brokers in the sample. The reputation coefficient remains strongly significant for all of these last four variations.

2.4.2 Analyst Performance and Reputation

This section tests whether analyst reputation is increasing in the accuracy of recent forecasts. Demonstrating this link provides evidence that investors update their beliefs about analyst ability in a rational manner based on recent performance. It also provides an incentive for analysts to care about their accuracy, given we have just shown that reputations are economically valuable.

Of course in reality an analyst's reputation will be influenced by factors other than accuracy. For example, timeliness of forecasts is likely to be highly important to investors. Some analysts may trade off accuracy against timeliness in order to communicate private information rapidly to investors[15]. Other less quantifiable factors, such as an analyst's responsiveness to client requests or knowledge of the industry conveyed in written reports are also likely to influence reputation in a significant way.

---

[15] Cooper, Day, and Lewis (2001) find that timeliness can be used to identify high ability 'lead' analysts, whose forecasts have a greater impact on stock prices.

Several control variables are included to capture these factors. The average age of the analyst's forecasts relative to other analysts covering the stock is used to capture timeliness. The relative optimism of the analyst's forecasts is included, as are investment banking relationships where the firm acts as the lead broker in either an SEO or IPO in the past three years[16]. Finally the size of the broker is also included as a control[17].

The independent variable of most interest for our test is the relative accuracy of the analyst. The dependent variable in the regressions is analyst reputation, measured by the percentile rank of the analyst in the current end of year survey. Table 2.5 presents the results of the regression with analyst reputation as the dependent variable. Both univariate and multivariate regressions are presented using robust cluster adjusted t-statistics that control for correlated analyst-year effects.

We can see from the univariate results that the analyst's reputation is significantly positively related to relative accuracy in that year. Including the control variables does not affect this conclusion. Thus the data support the second hypothesis that accurate analysts receive higher survey ratings than non-accurate analysts, consistent with the market adjusting its beliefs about the analyst's ability in a rational manner.

The control variables reveal some interesting results. Firstly, reputation is strongly increasing in the size of the broker. This result is consistent with larger brokers employing better analysts and having more resources at their disposal. Secondly, reputation is significantly decreasing with the average age of the analyst's estimates. The coefficient on timeliness is 2.5 times the size of the coefficient on

---

[16] Investment banking relationships where the broker acts as a co-manager have no effect on the regression and are excluded for brevity.

[17] Interestingly Clement (1999) finds that analysts at larger brokers are more accurate.

accuracy[18], suggesting this variable may be significantly more important than accuracy in determining an analyst's ranking. Finally, the coefficients on investment banking and optimism variables are insignificant.

To ensure the accuracy-reputation linkage is robust I perform the following checks. Firstly, the analysis is repeated for large (Top 100) stocks only. This increases the size of the accuracy coefficient in the multivariate regression by 19%. Secondly, t-statistics are recalculated using broker-year clusters to account for common shocks to the reputation of analysts from the same brokerage firm. The accuracy variable remains significant. Thirdly, the analysis is repeated using the top 9 brokers only. Fourthly, I exclude negative earnings stocks to ensure write-offs are not influencing the result. Finally, I calculate t-statistics using stock-year clusters to allow for negative cross-correlations in rankings for the same stock. The empirical finding is robust to all these variations.

[18] Since both are percentile ranks of the analyst relative to other analysts covering the stock, it is meaningful to compare the coefficients on these two variables to see which is more economically important.

Table 2.5: The Effect of Accuracy on Analyst Reputation

This table analyses the impact of relative analyst accuracy in year t on analyst reputation in year t. 'FY1 accuracy' is a percentile ranking of the analyst's accuracy relative to other analysts covering the stock (1=most accurate, 0=least accurate) averaged over the full year. 'FY1 estimate age' is a percentile ranking of the average age of the analyst's estimates relative to other analysts covering the stock (1=least timely, 0=most timely). 'FY1 optimism' is a percentile ranking of the analysts relative optimism about FY1 earnings averaged over the year (1=most optimistic, 0=least optimistic). 'Broker size' is the total market share of the broker in year t across all stocks. 'SEO Lead' is a dummy variable indicating whether the broker has acted as lead manager in that stock in the current year or in the previous 3 years respectively. 'IPO Lead' is defined in the same way. Analyst-Stock-Years are used as the unit of analysis, with both robust and cluster adjusted t-statistics presented. The dependent variable is the percentile ranking of the analyst in year t.

	Analyst Reputation (t)					
	All Stocks			Large Stocks (Top-100)		
	Coefficient	White t-stat	Cluster t-stat	Coefficient	White t-stat	Cluster t-stat
Constant	0.611	52.09	32.58	0.568	40.02	26.83
FY1 Accuracy	0.021	2.36	2.28	0.025	2.22	2.15
FY1 Estimate Age	-0.052	-5.81	-4.47	-0.080	-7.08	-5.44
FY1 Optimism	-0.014	-1.53	-1.36	-0.017	-1.49	-1.32
Broker Size (t)	0.819	8.27	4.33	1.299	10.92	6.12
SEO1_Lead	0.023	0.95	0.93	0.014	0.43	0.43
SEO3_Lead	0.033	1.93	1.90	0.061	2.64	2.65
IPO3_Lead	0.021	0.87	0.88	0.065	2.12	2.12
Observations	9,208			5,748		
Adjusted R2	1.66%			4.19%		

2.4.3 Analyst Optimism and Broker Market Share

This section tests whether optimistic analysts generate higher trading volume. Establishing the link between optimism and market share establishes that analysts have an incentive to be optimistic on average, even in the absence of investment banking affiliations or management access concerns.

To proxy for optimism I use three separate measures. The first is the optimism of the analyst's FY1 earnings forecast relative to other analysts covering the same stock. This is measured as a percentile rank, with 1 being the most optimistic analyst on average throughout the year, and 0 being the least optimistic analyst[19]. The second proxy is the relative optimism of the analyst's FY2 earnings forecast. I include these two measures separately, since it is plausible that analysts may communicate optimism via longer-term forecasts, since these are harder for the investor to verify and the analyst may be able to 'get away' with excessive optimism more easily in distant forecasts. The final measure of optimism is the actual recommendation issued by the analyst (eg buy, hold, sell). Recommendation optimism is likely to be the least costly to a broker's reputation, since recommendations typically are not accompanied by a specific time horizon. Unfortunately recommendation information is only available on IBES for a subset of firms in my sample. This significantly reduces the sample size from around 14,000 to 3,000 analyst-stock-years.

The dependent variable is the broker market share in the current year. I include investment banking affiliations, broker size, and reputation as controls. The results of the analysis are presented in Table 2.6.

[19] See Appendix 1 for a detailed account of the construction of this variable.

Table 2.6: The Effect of Optimism on Broker Market Share

This table analyses the impact of relative optimism in year t on broker market share in year t. 'FY1 Optimism' is a percentile ranking of the analyst's FY1 EPS forecast relative to other analysts covering the stock (1=most optimistic, 0=least optimistic) averaged over the year. 'FY2 Optimism' and 'Rec Optimism' are constructed similarly using FY2 EPS forecasts and recommendations (buy, hold, sell) respectively. 'Analyst Ranking' is a percentile ranking of the analyst relative to other analysts covering the stock (1=highest ranked, 0=lowest ranked). 'Broker size' is the total market share of the broker in year t across all stocks. 'SEO Lead' is a dummy variable indicating whether the broker has acted as lead manager in that stock in the current year or in the previous 3 years respectively. 'IPO Lead' is defined in the same way. Analyst-Stock-Years are used as the unit of analysis and cluster adjusted t-statistics are presented. The dependent variable is the market share of the broker in that stock for that year. Only stocks where the broker has an analyst covering the stock are included (and where optimism and ranking information for that analyst exists – which significantly decreases the sample size for regressions using the recommendation data). Results are calculated using all stocks in the sample.

	Dependent Variable = Broker Market Share (t)															
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat	Coef	Cluster t-stat
Constant	0.002	5.99	0.006	3.42	-0.033	-6.47	-0.034	-6.53	0.007	1.99	0.001	0.18	-0.038	-5.11	-0.038	-5.20
FY1 Optimism (t)	0.002	3.52					-0.005	-1.42	0.005	2.08					-0.003	-0.69
FY2 Optimism (t)			0.015	7.82			0.010	2.65			0.017	6.58			0.007	1.73
Rec Optimism (t)					0.016	8.98	0.015	8.43					0.016	7.82	0.016	7.43
FY1 Estimate Age (t)	-0.010	-5.62	-0.009	-5.20	-0.006	-1.63	-0.005	-1.48	-0.006	-2.61	-0.006	-2.34	-0.003	-0.89	-0.002	-0.62
Analyst Ranking (t-1)									0.022	5.59	0.022	5.48	0.012	2.51	0.012	2.46
Broker Size (t)	0.845	49.32	0.850	48.00	0.913	14.29	0.912	25.02	0.706	18.38	0.712	18.33	0.856	14.24	0.853	13.87
SEO1_Lead	0.036	3.86	0.035	3.78	0.054	2.98	0.055	2.92	0.041	3.58	0.040	3.52	0.055	2.73	0.056	2.74
SEO3_Lead	0.057	10.51	0.057	10.37	0.028	2.83	0.028	2.78	0.057	8.17	0.056	8.10	0.029	2.65	0.028	2.6
IPO3_Lead	0.087	7.54	0.086	7.43	0.033	1.94	0.032	2.02	0.082	5.64	0.082	5.52	0.034	1.72	0.033	1.7
Observations	14216		13832		3375		3305		8286		8103		2569		2569	
Adjusted R2	29.2%		29.1%		35.0%		35.0%		18.9%		19.3%		24.8%		24.8%	

The first four regressions omit the analyst reputation variable (increasing the size of the sample by around 75% for the FY optimism regressions), while the second four include reputation. Standard errors are corrected for analyst-year clustering.

The results show that individually FY1, FY2, and recommendation optimism each positively influence broker market share in the current year, consistent with the hypothesis. When all three variables are included in the regression, the recommendation optimism variable dominates the other two variables, consistent with the interpretation that analysts are more likely to transmit optimism via softer messages rather than near-term earnings forecasts. However, it should be noted that this final result is based on a relatively small data sample, so it is still likely that EPS optimism is an important channel on its own.

As we saw in Section 2.4.1, investment banking affiliations are an important control variable in the regression and significantly influence the level of broker market share in a given stock. Broker size and analyst reputation both enter significantly, consistent with our previous results.

To ensure the optimism results are robust I perform the following checks. Firstly I recalculate the standard errors using both broker-level and firm-year level clustering rather than analyst-level clustering. This does not alter the economic findings; FY1, FY2, and recommendations each remain individually significant in regressions 1-3 and 5-7, and similar results apply for regressions 4 and 8 with recommendations dominating the other variables. Secondly, I repeat the regressions for each separate year from 1995 to 2002. The coefficients on FY2 and recommendation optimism in the univariate regressions are statistically significantly positive in every year. The coefficient on FY1 optimism is positive in every year though significant in only two of the years. Thirdly, I redo regressions 1-8 for the top 9 brokers only. The results are unchanged. Fourthly, I log-transform the Y-variable as in Section 2.4.1. This improves the fit of the regressions, strengthens the FY2 and recommendation optimism results, while the FY1 optimism result becomes weaker.

Next, I exclude all stocks that have had an IPO or SEO in the last three years. The optimism findings remain highly significant. Finally, I exclude stock-years with negative earnings, which does not change the results. In summary, the FY2 and recommendation optimism results appear quite robust while the FY1 optimism result is somewhat less so. Overall however, the results are strongly supportive of the third hypothesis that more optimistic analysts generate higher trading volume for their firms.

To conclude, the first two empirical results support the Neoclassical view that investors are acting rationally. Namely, high reputation analysts generate higher future trading volume, and better forecasting performance leads to a higher reputation. However the third result, that optimistic forecasts generate higher current trading volume, appears to support the Spitzer view. In the next section I reconcile these apparently contradictory findings using a model with fully rational agents and asymmetric information.

2.5. Model of Reputation and Trade Generation

In this section I develop a simple model describing the strategic interaction between an investor and an analyst. The interaction is modelled as a dynamic game of incomplete information. The primary objective of the model is to reconcile the three empirical results found in Section 2.4 within a simple, fully rational agent framework. In addition the model produces three further testable hypotheses that I subsequently empirically test in Section 2.6.

2.5.1 Description of Economy

There are two possible future states of the world (labelled high and low) that are equally likely. A single stock, with initial price P_0, is available with payoffs x_{high} and x_{low} in these two states where $x_L < P_0 < x_H$.

There exists a one-period investor with log utility and initial wealth W. His initial beliefs about the state of the world are given by $\lambda = \Pr(x_{high}) = 0.5$, and he is endowed with δ_0 shares in the single stock where:

$$\delta_0 = \frac{-W(\lambda(x_H - P_0) + (1-\lambda)(x_L - P_0))}{(x_H - P_0)(x_L - P_0)}$$

This seemingly complicated expression is just equal to his optimal shareholding given his initial beliefs (see Appendix 2B for details).

The investor employs a sell-side analyst at the start of the period to forecast the future state. I assume that he only employs a single analyst[20], and is obligated to trade through the brokerage firm employing this analyst if he revises his beliefs on the basis of the analyst's advice[21]. The sell-side analyst may be one of two types; a 'smart' analyst (who receives informative private signals) or a 'dumb' analyst (whose private signals are just noise). Neither the investor nor the analyst initially knows the analyst's actual type, but they hold a common prior that $P(smart)=\theta$ where $0 \leq \theta \leq 1$.

At the start of the period, the sell-side analyst receives a non-contractible binary private signal[22] ($s = s_{high}$ or $s = s_{low}$) about the future state with the following properties:

$$P(s_{high} \mid x_{high}, smart) = p$$
$$P(s_{high} \mid x_{low}, smart) = 1-p$$
$$P(s_{high} \mid x_{high}, dumb) = P(s_{high} \mid x_{low}, dumb) = 0.5$$
$$\text{with } 0.5 < p < 1$$

After receiving the private signal, the analyst chooses a binary message ($m = m_{high}$ or $m = m_{low}$) that she[23] then sends to the investor. This message need not be truthful; ie she may choose to send m_{high} to the investor even after observing s_{low}. The investor receives the message and optimally updates his beliefs about the

---

[20] This allows us to abstract for the moment from well-documented issues including analyst herding and exaggeration. For an examination of these issues see Scharfstein and Stein (1990), Trueman (1994), and Zitzewitz (2001).

[21] For a discussion of the reasonableness of this assumption see Section 2.3.5.

[22] The signal is not verifiable ie a court cannot distinguish whether the analyst received the high or low signal. This prevents a contract being written with payment contingent on the analyst truthfully reporting the private signal.

[23] For clarity of exposition I refer to the analyst as 'she' and to the investor as 'he' throughout the paper.

expected payoff from the stock. His posterior belief is given by $\lambda^* = P(x_{high} \mid m)$ which will be determined in the equilibrium of the model (Section 2.5.2).

Given these revised beliefs, the investor recalculates his optimal holding of the stock and trades to achieve this holding. For simplicity I assume the investor is sufficiently small that his actions do not influence the stock price, which remains constant at P_0[24]. The optimal holding given the revised beliefs is given as:

$$\delta_1 = \frac{-W(\lambda^*(x_H - P_0) + (1 - \lambda^*(x_L - P_0))}{(x_H - P_0)(x_L - P_0)}$$

His desired trade, $\delta_1 - \delta_0$, is directly proportional to his revision in beliefs and is given as:

$$\text{Desired Trade} = \delta_1 - \delta_0 = \alpha(\lambda^* - \lambda)$$

$$\text{where:} \qquad \alpha = \frac{-W(x_H - x_L)}{(x_H - P_0)(x_L - P_0)} > 0$$

To make the model interesting[25] (and realistic) assume that the investor faces binding short-sales constraints with probability q. This implies that the investor may not always be able to implement his desired trade after receiving a low message that

---

[24] Relaxing this assumption to allow the analyst's signal to be partially incorporated in the share price would not alter the economic conclusion of this model but decreases tractability of the calculations. If the signal is fully incorporated in the price, the results will not go through. This assumption also enables us to avoid considerations of how the investor would strategically modify their demand to exploit monopolistic private information as in a typical Kyle (1985) model setup.

[25] Without such an asymmetry in the model, the analyst will always choose to report a truthful message, since her expected posterior reputation is higher than if she sends a dishonest message. There are various ways of introducing asymmetry in the analyst's incentives, for example previous researchers have examined investment banking affiliations and management access as possible incentives for the analyst to lie optimistically in given situations. In this paper I focus on a third channel, the trade generation incentive facing analysts, and introduce the asymmetry via short selling constraints.

induces a downward revision in beliefs[26]. In reality, short-sales constraints are frequently generated by institutional restrictions on short selling. For example Almazan et al (2003 p.11) find that around 2/3 of mutual fund managers are specifically restricted from short selling in their mandates, and of the 1/3 that are allowed, only 10% of them actually *did* short any stocks in 2000 (59 funds out of 1838 sampled). Another reason for limited shorting may be high lending fees or the inability to short some stocks (Lamont (2003)).

If short sales constraints are binding, the investor does not undertake any trade. Otherwise, he is able to implement his full desired trade. The *expected* trade generated given a positive and negative revision in beliefs is given respectively by:

$$\text{E(Trade} \mid \text{Positive Revision)} = \alpha\left(\lambda^{*} - \lambda\right)$$
$$\text{E(Trade} \mid \text{Negative Revision)} = (1-q)\alpha\left(\lambda^{*} - \lambda\right)$$

Thus if $q > 0$, a negative revision to beliefs will generate a *lower* absolute expected trade size than an equivalently sized positive revision to beliefs. This asymmetry is crucial to the results of the model as it provides an incentive for the analyst to be optimistic.

Finally in our model, at the end of the period the true state of the world (x_{high} or x_{low}) is revealed to both the investor and the analyst. The investor compares the message sent by the analyst, m, to the realised state, x, and optimally updates his beliefs about the analyst's ability. The updated posterior belief is given by $\theta^{*} = P(smart \mid m, x)$. The investor truthfully reports this updated belief to a survey,

[26] The use of a random constraint may be justified by assuming that the analyst doesn't know which investors are able to sell after she issues a negative message, but she does know the ex-ante probability that a given investor can sell.

then dies[27]. I will refer to this posterior belief as the 'reputation' of the analyst. Figure 2.1 presents a timeline of events in the model.

The analyst is employed by a brokerage firm that compensates her based on both the level of trading commission she generates in a given period and her survey ranking at the end of the period. The survey ranking is valuable to the brokerage firm since it leads to higher future trading volume in subsequent periods, given that subsequent investors use the ranking as their prior, and that higher reputation analysts generate larger revisions in beliefs (which I prove is true in equilibrium in Section 2.5.2). Let the percentage trading commission rate be given as c.

The analyst's objective function may be specified in reduced form as:

$$\max_{m} c.E(Trade \mid m) + k.E(\theta^{posterior} \mid m)$$

That is, the analyst cares both about how much commission she generates this period, and her reputation for being smart at the end of the period[28]. The parameter $k \in [0,\infty)$ represents how much the analyst cares about her future reputation and is known by the analyst but not the investor. The investor therefore faces incomplete information about the analyst's payoff function, but he does know the probability distribution of k given by the cdf $F(k)$. At the start of the single period game, nature chooses the analyst's k-type according to this probability density function. The investor's uncertainty about k is the primary driving force behind the model's results[29].

---

[27] Investors in subsequent periods use only this reported survey belief in their decisions and no other information from past analyst actions. An identical assumption is made in the seminal paper of Diamond (1989).

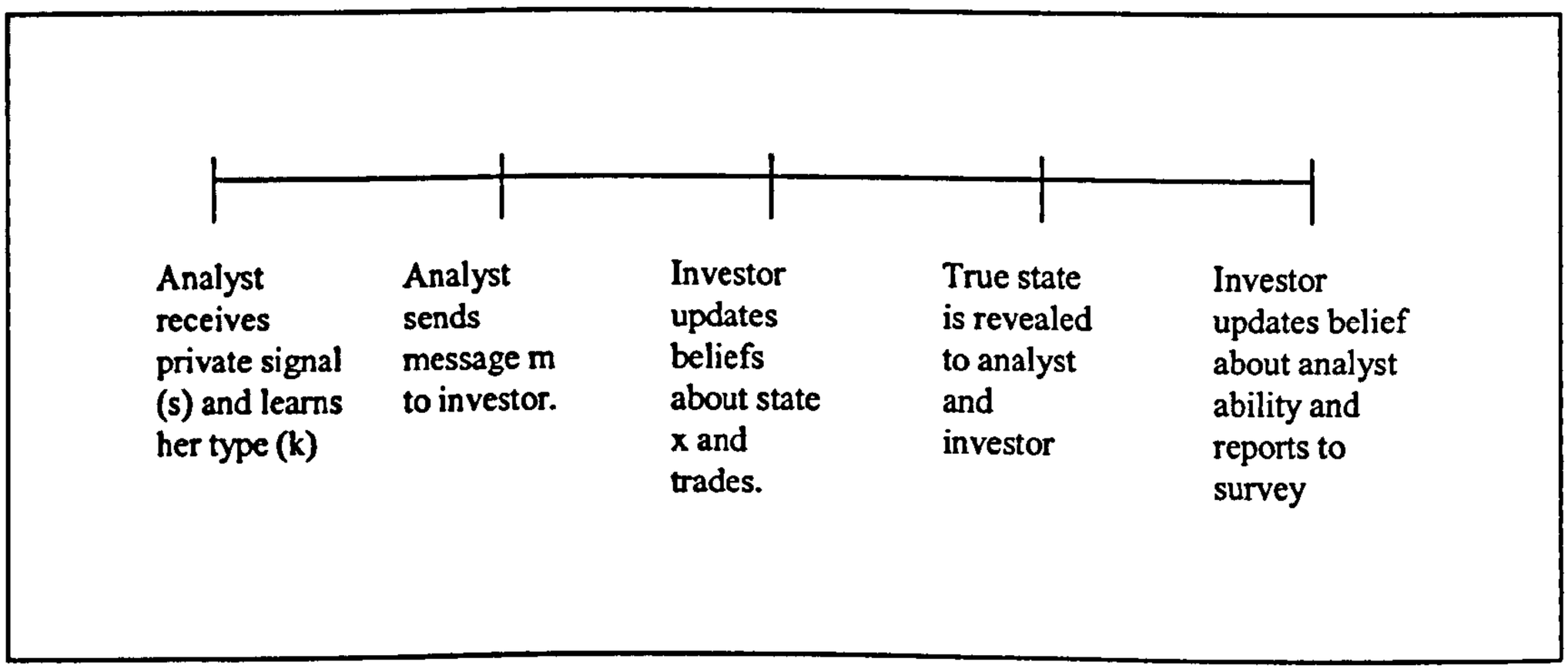
[28] A different way to model this game would be within a repeated game framework, where the analyst's utility function is given as the discounted value of all future trading commissions generated. In this game 'reputation effects' would arise endogenously given certain parameter restrictions. However this approach also greatly complicates the model without adding significantly to the basic point I am trying to make. For examples of this approach see the papers by Diamond (1989) and John and Nachman (1985).

[29] See Morgan and Stocken (2003) for a related model examining how investors uncertainty about an analyst's type leads to equilibrium over-optimism and a significant loss in information transmission.

The parameter k may be interpreted in at least two ways[30]. Firstly k may represent stochastic outside options facing the analyst in that period. A low value of k indicates attractive outside options, so the analyst cares less about her reputation. A high value indicates the analyst values her reputation highly. To prevent the investor learning about k I assume that the outside options facing the analyst are randomly changing over time (eg attractive portfolio management positions arrive randomly at the start of each period).

A second interpretation is that k is persistent across periods for a given analyst. This may reflect the analyst's career concerns, for example a young analyst may have a high k and value her reputation highly (Holmstrom (1982)) while an analyst in her final year may have a low k due to the 'final period' effects of the game. In this interpretation investors must be assumed to be unable to communicate information about the analyst's k-type across periods, or alternatively are unable to tell whether the analyst is near the final round of her game.

Figure 2.1: Timeline of Events in Model



[30] Benabou and Laroque (1992 pp 941) use a similar technique, allowing the agent's type to follow a stochastic process. This makes the learning problem more difficult. In equilibrium, information is transmitted and credibility does not vanish.

We will see in equilibrium that the desire to generate both high current commission and a high end-of-period reputation generates conflict for the analyst only when she receives a *low* private signal. For a *high* private signal, s_{high}, her two objectives are aligned since sending a positive message generates both higher trading volume and a higher expected posterior reputation (since truthful messages are more likely to be correct). Conversely for the *low* private signal, s_{low}, sending the truthful low message will generate lower trading volume than a high message (due to short sales constraints), but will also generate a higher expected end-of-period expected reputation. Thus a conflict exists between trade generation and reputational concerns.

Her ultimate choice is given by the message that maximises her objective function. If the following condition holds then she will truthfully report the low message:

$$c.\left[E(Trade|m_H)-E(trade|m_L)\right] \leq k.\left[E(\theta^{posterior}|m_L)-E(\theta^{posterior}|m_H)\right] \qquad (3)$$

Otherwise she will lie and report m_H. If the analyst's type k is large enough that she reports truthfully I will refer to her as a 'good' analyst, otherwise if she chooses to lie I will refer to her as an 'evil' analyst.

The analyst's choice depends on the investor's beliefs in equilibrium, and on her k-type. The investor's beliefs depend in turn on the strategy he expects the analyst to follow. These issues are examined in the equilibrium discussion to which I now turn.

2.5.2 Specification of Equilibrium

In this section I derive an equilibrium for this simple game. The solution concept I use is perfect Bayesian equilibrium. This solution concept requires that the strategy profile of each agent is sequentially rational given his/her beliefs, and that beliefs are generated using Bayes rule. In addition beliefs off the equilibrium path must be specified. Proposition 1 gives the equilibrium in the investor-analyst game.

Proposition 1: *Suppose the following condition*[31] *holds:*

[C.1]
$$q > 1 - \left(\frac{\pi}{2-\pi}\right)$$

then there exists a perfect Bayesian equilibrium with the following features:

[A.1] *An analyst with type* $k \geq k^*$ *always reports truthfully, sending the low message* m_{low} *when she receives the low private signal* s_{low}*, and the high message* m_{high} *when receiving the high private signal* s_{low}*. An analyst with type* $k < k^*$ *always reports the high message, irrespective of her private signal. The critical level of* k^* *is given implicitly by:*

$$k^* = \frac{c\alpha\left(\frac{\pi}{2-\pi} + q - 1\right)(0.25 - (p-0.5)^2\theta^2\pi^2)}{\pi(1+\pi)(p-0.5)(1-\theta)\theta}$$

where $\pi = \Pr(k \geq k^*)$ *is probability that the analyst's type exceeds* k^**.*

[I.1] *An investor with beliefs* π, θ *will buy* $\delta_2^H - \delta_1$ *shares after receiving a high message from the analyst where:*

$$\delta_2^H - \delta_1 = \alpha(p-0.5)\theta\left(\frac{\pi}{2-\pi}\right)$$

If the investor <u>does not</u> face short sales constraints then he will sell $-\delta_2^H + \delta_1$ *shares after receiving the low message from the analyst where:*

$$-\delta_2^L + \delta_1 = \alpha(p-0.5)\theta$$

---

[31] Since π is determined endogenously in the equilibrium, it is not generally possible to express it as an explicit function of the exogenous parameters. For specific numerical values of the exogenous parameters and a particular distribution of k, it is however possible to solve for the minimum level of q to ensure that this condition holds.

If the investor does face short sales constraints then he will sell no shares after receiving the low message.

[I.2] *After the true state is revealed the investor reports his updated posterior belief* $\theta^* = \Pr(smart \mid x, m)$ *to a survey. There are four possible survey responses given as:*

$$P(smart \mid m_H \,\&\, x_H) = \frac{(0.5+(p-0.5)\pi)\theta}{0.5+(p-0.5)\theta\pi}$$

$$P(smart \mid m_H \,\&\, x_L) = \frac{(0.5-(p-0.5)\pi)\theta}{0.5-(p-0.5)\pi\theta}$$

$$P(smart \mid m_L \,\&\, x_H) = \frac{(1-p)\theta}{0.5-(p-0.5)\theta}$$

$$P(smart \mid m_L \,\&\, x_L) = \frac{p\theta}{0.5+(p-0.5)\theta}$$

This equilibrium has the following characteristics.

[Outcome.1] *The investor's trade size is increasing in their initial belief about the analyst's ability (ie higher reputation analysts generate more trade).*

[Outcome.2] *Analysts with accurate messages achieve increases in reputation.*

[Outcome.3] *High messages generate more trade on average than low messages.*

[Outcome.4] *The average message sent in equilibrium will be optimistic. If all analysts were honest, the average message sent would be 0.5. However in this equilibrium the average message sent will be:*

$$E(m) = 1 - 0.5\pi > 0.5$$

$$when\ \pi < 1$$

Proof of Proposition 1:

See Appendix 2. ■

We can see from Proposition 1 that the three empirical results from Section 2.4 arise in the equilibrium. Higher reputation analysts generate higher future trading volume (Outcome 1), accurate analysts are rewarded with increases in their reputation (Outcome 2), and optimistic analysts generate more trade for their firms (Outcome 3).

We can also see that investors are acting in a rational way in this equilibrium, updating their beliefs via Bayes rule and having correct prior beliefs about the parameters of the model. Investors face asymmetric information about the analyst's concerns for her future reputation, and adjust their actions and beliefs accordingly. The analyst correctly anticipates how investors will react to her messages and in turn optimises her strategy.

We can illustrate the implications of the equilibrium via a simple numerical example. If we assume a uniform distribution of k between 0 and 1, with parameters of $c = 0.05, \theta = 0.5, p = 0.60, q = 0.5$ then the critical level is $k^* = 0.0942$. This implies the analyst has a 90.58% probability of being a 'good' type, and the average message sent in equilibrium is 0.547 (compared to the unbiased true signal of 0.50). Additionally condition [C.1] is satisfied since $q = 0.5 > 0.172 = 1 - \pi/(2-\pi)$.

We can also generate several comparative statics results from this model as outlined in Proposition 2.

Proposition 2: *Given the equilibrium in Proposition 1, the following comparative statics results hold for the equilibrium level of optimism* $E(m)$*:*

[2.1] The expected message is increasing in the level of short sales constraints, q.

[2.2] The expected message is decreasing in the precision of analyst's private signals, p.

Proof of Proposition 2:

See Appendix 3. ■

Proposition 2.1 predicts that consensus optimism will be higher for stocks that have higher probabilities of short-sales constraints. The intuition behind this result is straightforward. As the level of short sales constraints becomes higher, the difference in trading volume between a high and low message becomes larger, so the incentive for the analyst to lie optimistically becomes stronger. Thus the critical level of k increases and the probability that the analyst reports truthfully drops. As long as condition [C.1] holds, optimistic messages will still generate more trade.

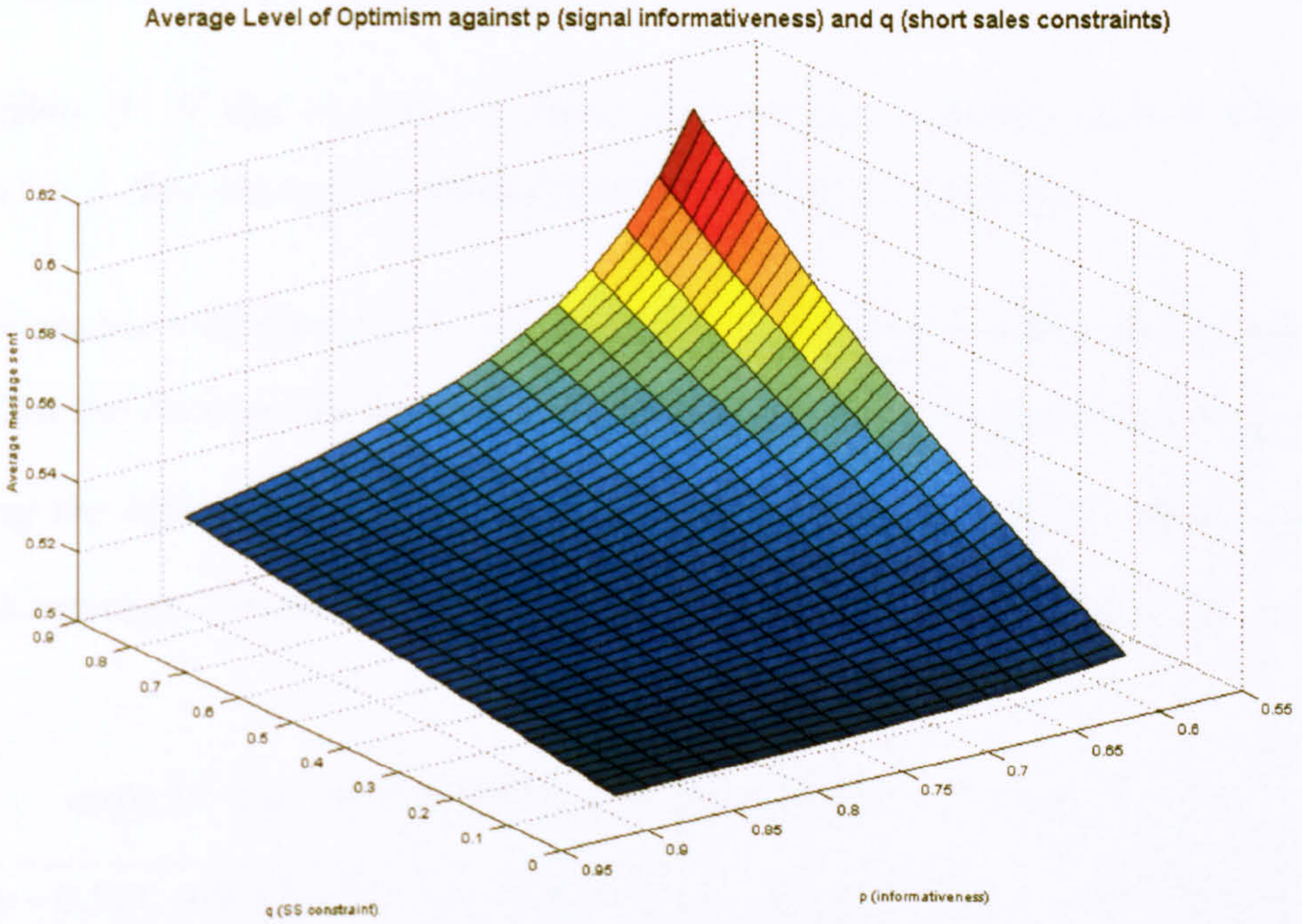
Proposition 2.2 predicts that when the informativeness of the private signal decreases, the equilibrium level of optimism rises. In this case, the analyst is less likely to be 'caught out' by sending an incorrect signal, and therefore the expected loss in reputation from reporting an incorrect message is lower.

Again, using our simple numerical example with k uniform between 0 and 1, and parameters of $c=0.05, \theta=0.5$, Figure 2.2 plots the average message sent in equilibrium for different values of p and q. We can see from Figure 2.2 that the average message sent is always biased optimistically, and this optimism becomes larger as short sales constraints q increase and as signal informativeness p decreases.

So far, we have maintained the assumption that investors act in a fully rational manner and have not needed to appeal to any behavioural explanations to generate the results. However, it is interesting to compare the predictions of this model to a behavioural explanation of the phenomena.

To do this (and subsequently test the Spitzer view against the Neoclassical view), we can adjust our model to impose a specific behavioural error on investors' beliefs. One argument commonly used by regulators (and behaviouralists) is that investors (particularly small investors) often naively believe that analysts always report truthfully, allowing them to be consistently fooled by optimistic messages.

Figure 2.2: Average Optimism Comparative Statics



This figure shows the average message sent in equilibrium for various combinations of p (signal informativeness) and q (short sales constraints). The unbiased private signal has a value of 0.5, so this figure shows that optimism will exist in equilibrium, and this optimism will increase as p decreases and as q increases.

For example the Wall Street Journal (2002) reported that:

> *"Investigators had spent months looking into allegations that analysts had hyped the stocks of companies that generated fat investment-banking fees for their firms, misleading small investors when the companies -- including money-losing Internet start-ups -- went bust."*

We can incorporate this version of the Spitzer view into our model by adding a simple modification[32]. Assume with probability ξ the investor is naïve, in which case he believes that the analyst is always truthful, ie he believes incorrectly that

[32] See Ottaviani and Squintani (2002) for a related model examining equilibria in a cheap talk game where some players act in a non-strategic manner.

$\pi = 1$. With probability $1-\xi$ assume the investor is rational. We can then solve for the new equilibrium in this situation.

Proposition 3: *If the investor is naïve (incorrectly believing that $\pi = 1$) with probability ξ then the optimal strategy for the analyst is as follows:*

[3.1] An analyst with type $k \geq k^N$ always reports truthfully, sending the low message m_{low} when she receives the low private signal s_{low} and the high message m_{high} when receiving the high private signal s_{low}. An analyst with type $k < k^N$ always reports the high message, irrespective of her private signal. The critical level of k^N is given by:

$$k^N = \frac{c\alpha(0.25-(p-0.5)^2\theta^2)(0.25-(p-0.5)^2\theta^2\pi^2)\left(q+(1-\xi)\left(\frac{\pi}{2-\pi}-1\right)\right)}{(p-0.5)(1-\theta)\theta\left(2(0.25-(p-0.5)^2\theta^2\pi^2)\xi+\pi(1+\pi)(0.25-(p-0.5)^2\theta^2)(1-\xi)\right)}$$

where $\pi = \Pr(k \geq k^N)$ is probability that the analyst's type exceeds k^N.

[3.2] The equilibrium level of optimism is increasing in ξ.

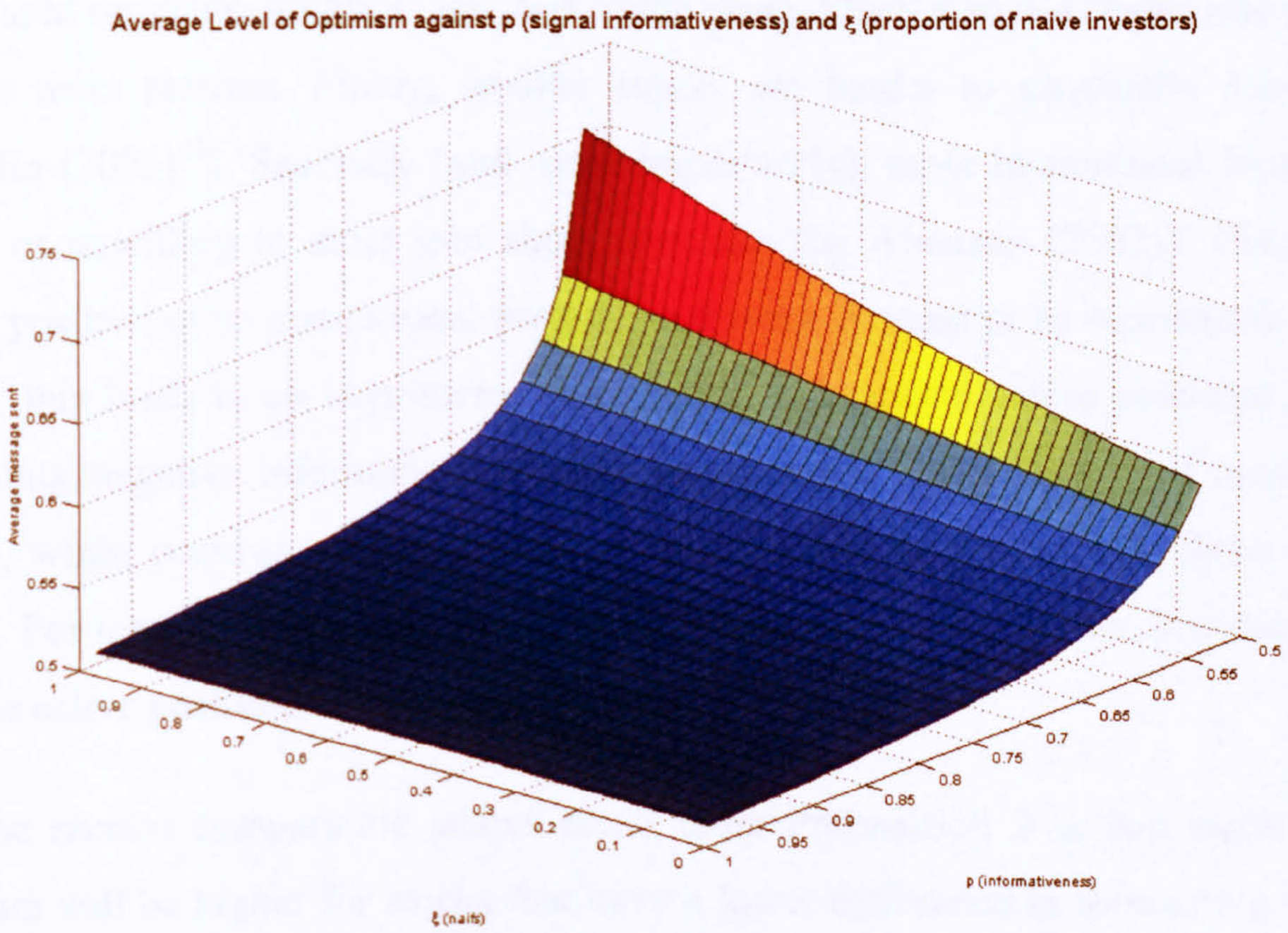
Proof of Proposition 3:

See Appendix 4. ■

Under the Spitzer model of investor irrationality, the equilibrium level of optimism is higher than it is with sophisticated investors. The intuition behind this result is that investors do not correctly adjust for the decreased credibility of a positive message relative to a negative message, so the asymmetry in trading volume between high and low messages facing the analyst is increased. Proposition 3.2 implies that the equilibrium level of optimism will be higher when the proportion of naïve investors, ξ, is higher. Since the Spitzer view typically assumes small investors are more prone to naivety, a natural empirical implication is that equilibrium optimism should be higher for stocks with a larger proportion of retail investors.

Continuing our simple numerical example, with k uniform between 0 and 1, and parameters of $c = 0.05, \theta = 0.5, q = 0.5, p = 0.6$ then the proportion of honest analysts is 90.58% when all investors are rational, 89.17% when 50% of investors are naifs, and 87.59% when all investors are naifs. The average message sent is 0.547, 0.554, and 0.562 respectively. Figure 2.3 plots the average message sent as a function of the proportion of naïve investors ξ. We can see that the message sent is increasing in ξ.

Figure 2.3: Average Optimism Comparative Statics



This figure shows the average message sent in equilibrium for various combinations of p (signal informativeness) and z (proportion of naive investors). The unbiased private signal has a value of 0.5, so this figure shows that optimism will exist in equilibrium, and this optimism will increase as p decreases and as z increases.

In summary the model serves two purposes. Firstly it demonstrates that my three initial empirical findings can be generated by a simple model with fully rational agents in the presence of asymmetric information and short-sales constraints. Secondly the model generates three further empirical predictions about the

equilibrium level of optimism. The final section of the chapter empirically tests these predictions.

2.6. Extension: Tests of Model Predictions

Proposition 2 predicts that the equilibrium level of optimism in analysts' forecasts should be increasing in the level of short sales constraints q facing investors. As a proxy for q, I use stock size measured by the stock's percentile size rank (based on index-weight at the start of the year). Stock size is a reasonable proxy for two main reasons. Firstly, smaller stocks are harder to physically short (eg D'Avolio (2003)[33]). Secondly (and more importantly), most institutional funds are unable or unwilling to enter into short positions (eg Almazan (2002)). Given the neutral position of an institutional investor in a stock is equal to its benchmark index weight, this leads to an asymmetric distribution of possible active positions. With constraints, negative information can only incorporated downwards until zero stock is held, while positive positions may be incorporated up to multiple times index weight. For large index weight stocks, there is less asymmetry in this distribution of possible active positions than for smaller stocks.

The second comparative statics result from Proposition 2 is that equilibrium optimism will be higher for stocks that have a lower difference in forecasting ability between high and low ability analysts, ie low p stocks. The intuition behind this result is that analysts are less likely to be 'caught out' if they cheat on a stock with difficult to forecast earnings. As a proxy for p, I use the dispersion of analyst forecasts (normalised by the stock-price). Dispersion can be interpreted as a measure

[33] D'Avolio (2003) finds that smaller stocks have higher average loan fees and are significantly more likely to be 'on special' ie periodically very expensive to borrow.

of the noise in the private signals received by analysts (eg Johnson (2003)), which fits naturally with our definition of p.

Proposition 3 predicts that equilibrium optimism will be higher when the proportion of naïve investors increases. To proxy for naïve investors, in a manner consistent with the Spitzer view, I use the percentage of trading volume accounted for by retail investors in each stock. To test Proposition 3 we are therefore testing jointly our proxy for naïve investors plus the conjecture that optimism increases as the percentage of naïve investors increases.

The dependent variable required to test these three hypotheses is the average level of optimism in earnings forecasts for each stock. To measure this, each year for each stock I calculate the consensus earnings forecast six months before the earnings announcement that falls in that year. The difference between the consensus and actual forecast is divided by the price at the time consensus is measured, to give an optimism measure. A positive optimism measure implies that the consensus forecast was higher than the actual earnings announcement. To ensure the regression results are robust, I convert this measure to a percentile rank and do the same to each of the independent variables.

To test the three hypotheses I run Fama-Macbeth regressions with consensus optimism as the dependent variable and size, dispersion, and retail ownership as independent variables. In addition to control for any effects of investment banking on post-deal optimism I include control variables indicating whether the stock had undertaken an IPO or SEO in the last three years and a second control variable, analyst coverage.

Table 2.7 presents the regression results. In the univariate regressions, all three hypotheses appear to be supported, with significant t-statistics on size, dispersion, and retail share. However in the multivariate regression, the retail-share coefficient becomes insignificant, while size and dispersion remain strong. Thus, the first two supplementary hypotheses are supported by the data. The third hypothesis is not

strongly supported since we are unable to reject the null that the coefficient on retail share is zero. One reason for this result may be that small investors are not a good proxy for naivety (as suggested by Jackson (2003)).

To ensure these results are robust to my choice of proxy specification, I repeat the regressions using optimism measured three rather than six months prior to the earnings announcements. This generates the same conclusions. I also repeat the multivariate regression separately for each yearly cross-section. In 8 of the 9 cross-sections, the size and dispersion coefficients have the predicted sign, while in 6 of the 9 the retail share variable has the opposite to predicted sign. Next, I repeat the tests controlling for write-offs by firstly excluding stocks experiencing negative earnings in a given year, and secondly excluding stocks where earnings surprises exceed 5% of the stock price. Neither of these variations affects the results. As a final robustness check, to account for the forecast error distributional issues addressed by Abarbanell and Lehavy (2003) I calculate non-parametric estimates of consensus optimism differences. For small stocks (Ex-100) consensus forecasts exceed actual earnings in 63% of cases while for large stocks (Top 50) this statistic is 49.7%. Similarly for high dispersion stocks (top quartile) consensus exceeds actual in 65% of cases, while for low dispersion stocks (bottom quartile) this statistic is 48%. The differences are highly significant, and confirm my earlier parametric results.

In summary, the model predicted higher levels of optimism for smaller and harder to forecast stocks and these predictions are verified in the data. An extension of the model (Proposition 3) allowed us to distinguish somewhat between the Spitzer and Neoclassical explanations of the results. The findings tend to support the Neoclassical view, which is consistent with the objective of the model, ie to demonstrate that the empirical findings can be generated within a rational framework.

Table 2.7: Equilibrium Optimism Tests

This table analyses the factors influencing the consensus level of optimism for a stock. 'Optimism' is defined as the consensus forecast six months prior minus the actual earnings announcement normalised by stock price. 'SizeRank' is the size of the firm at the start of the year, expressed as a percentile rank. 'Dispersion' is the dispersion of earnings forecasts six months prior to the actual earnings announcement. 'Retail Share' is the percentage of turnover accounted for by retail investors. 'IPO' and 'SEO' is a dummy variable indicating whether the firm has undertaken an IPO or SEO in the last 3 years. 'Coverage' measures the number of analysts covering the stock six months prior to the earnings announcement. Fama-Macbeth panel regressions are run annually for 9 yearly cross-sections from 1994-2002. Average cross-sectional R2's are reported as are the average number of observations in each cross-section.

	Dependent Variable = Optimism (t)									
	(1)		(2)		(3)		(4)		(5)	
	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Constant	0.639	31.54	0.379	14.63	0.420	24.43	0.503	8.04	0.546	11.54
Size Rank	-0.235	-6.79					-0.181	-3.78	-0.238	-3.49
Dispersion			0.242	4.88			0.208	5.68	0.148	4.02
Retail Share					0.227	4.11	0.011	0.21	-0.045	-1.13
SEO									0.005	0.41
IPO									-0.003	-0.14
Coverage									0.006	2.48
Av # obs in cross-section	204.6		205.3		205.3		204.6		204.6	
Av adj R2	5.8%		7.6%		4.2%		11.8%		11.0%	

2.7. Conclusion

This chapter presents three main empirical findings. Firstly, high reputation analysts generate higher future trading volume. Secondly, accurate analysts are rewarded with higher end-of-period reputations as measured by survey rankings. Thirdly, optimistic analysts generate more trading volume for their brokerage firms.

To reconcile these results I build a simple model to demonstrate that the incentive to generate trade may lead analysts to lie about their private signals. The analyst trades off an expected loss in reputation against the short-term benefit of higher trading commissions. In equilibrium, optimistic analysts will indeed generate more short-term trading volume, even when investors are fully rational. However, as expected, an analyst's reputation suffers when she misleads investors. An analyst who cares highly about her future reputation will refrain from opportunistic behaviour.

I also confirm two auxiliary cross-sectional predictions from the model, namely that consensus optimism is higher when short sales constraints are more likely to bind (proxied by firm size) and when process noise is higher (proxied by analyst dispersion). A third test to distinguish between behavioural and rational explanations supports the latter view, given that consensus optimism is not increasing in the proportion of retail investors in a stock.

These findings suggest that recent plans to split brokerage firms may be unsuccessful in eliminating analysts' optimism bias. Investment banking incentives are likely to be replaced by trade generation incentives, which ultimately generate a similar effect. However, the results do suggest that an analyst's concerns for her reputation may be a successful way to address excessive opportunism. Recent moves to make previous analyst forecasting track records more transparent increase the ability of investors (especially small investors) to form an opinion on the ability of the analyst. This increased reputation effect reduces the incentive of analysts to report untruthfully.

Extending the theoretical analysis to situations involving multiple analysts would be a fruitful area for future research as would examining models where the analyst is aware of whether the investor is short sales constrained at any given point in time. Such models would generate more sophisticated patterns in analyst behaviour, for example engaging in flicking of forecasts in order to move clients in and out of a given stock over time.

2.8. Appendix 1: Analyst Variable Calculation

This appendix details the construction of the accuracy, optimism, and timeliness measures used in the empirical analysis.

2.8.1 Measuring analyst accuracy

Analyst accuracy has traditionally been measured as the absolute forecast error, scaled by current or lagged price (eg Brown et al (1987 p167)). The forecast used to evaluate accuracy for analyst i for firm j in year t is generally taken to be the most recent forecast made by the analyst prior to some fixed date a month or several months prior to the end of the fiscal year.

$$AE_{i,j,t} = \left| F_{i,j,t} - A_{j,t} \right| / P_{j,t-k}$$

where : $F_{i,j,t}$ = most recent forecast prior to cutoff date.

$A_{j,t}$ = actual earnings per share for firm

$P_{j,t-k}$ = stock price k periods prior to announcement

When comparing analyst forecast errors across different firms, several authors (eg Clement (1999), Hong and Kubik (2003)) note that we should adjust our measure to account for the fact that the information environment may differ across firms. For example a regulated utility company may have earnings that are easier to forecast than an airline. Alternatively some firms may engage in analyst guidance or earnings management while others may not.

To correct for differences in information environments, two alternative measures have been proposed. The first, from Clement (1999), subtracts the average absolute error for all analysts covering the stock in that year.

$$MAE_{i,j,t} = AE_{i,j,t} - Avg(AE_{i,j,t})$$

where $Avg(AE_{i,j,t}) =$ average absolute error for all analysts covering the firm.

The second method, from Hong and Kubik (2003), is to rank analysts based on their absolute forecast error in a given year, then scale analyst rankings between one (most accurate) and zero (least accurate) individually for each firm.

$$AE_Rank_{i,j,t} = 100 - \left[\frac{Rank - 1}{\#Analysts_{j,t} - 1}\right] * 100$$

One shortcoming of all three measures is that analyst accuracy is typically only measured at a single point in the year (eg one or six months prior to the announcement date). In reality however analysts make forecasts throughout the year, and continually revise these forecasts as new information appears. A such it is possible that an analyst may be very inaccurate throughout the year, but have one lucky forecast close to the measurement date and appear to be the most accurate analyst. A more reliable way to measure relative accuracy is to examine accuracy at multiple dates throughout the year, and construct an average accuracy measure over the year[34].

On each day of the 12 months prior to the day of the earnings announcement, I take the most recent forecast for each analyst and calculate the daily absolute forecast error relative to the future announced earnings. The forecast errors across analysts are compared, and each analyst is ranked daily based on their relative forecast error[35]. The average daily percentile rank is calculated for each analyst over the full year as follows.

---

[34] Using information from all forecasts on a high frequency (daily) allows us to match flow based measures of trading volume to a flow based accuracy measure. Since accuracy is an average deviation over time, this approach is quite analogous to the calculation of integrated volatility measures as described by Anderson et al (2003).

[35] To eliminate the impact of stale estimates, I delete forecasts where the age of the forecast exceeds 150 days.

$$Av_Accuracy_Rank = \frac{1}{n}\sum Acc_Rank_t$$

Finally each analyst receives a summary percentile accuracy rank based on this average. Since brokerage commission is a flow variable, measured over the full year, this flow-based measure of accuracy better captures the consistency of an analyst's relative accuracy throughout the year.

2.8.2 Measuring analyst optimism

Optimism is calculated using FY1 (next fiscal year) and FY2 (fiscal year two) earnings estimates. These measures simply remove the absolute value signs from the previous formulae and then follow exactly the same methodology.

$$Opt_{i,j,t} = (F_{i,j,t} - A_{j,t}) / P_{j,t-k}$$

I also calculate the relative optimism of analyst recommendations (buy, hold, and sell)[36] in the same way. As a robustness check, I also use a second measure of optimism equal to the percentage of time the analyst is above consensus. This does not affect my results.

2.8.3 Measuring Forecast Age

Since analysts issue earnings estimates at different points in time, it is important to correct for forecast recency (Brown (1991)). I calculate a relative forecast age measure that ranks analysts daily based on the age of their most recent outstanding estimate. Analysts with the oldest estimates receive a percentile ranking of 1, those

[36] Although IBES provides recommendations on a 1 to 5 scale, I find that this is inconsistently applied across brokers, with some brokers having all recommendations clustered on values of 1, 3 and 5, while other brokers use all 5 numbers. As such I reclassify recommendations 4-5 to buy, 3 to hold, and 1-2 to sell. Using the raw recommendations does not significantly change the empirical results.

with the most recent a rank of 0. Again I take the average of this measure over the year, and rank analysts based on this summary measure.

2.9. Appendix 2: Proof of Proposition 1

In this appendix I firstly derive the optimal beliefs and strategy of the investor, given the conjectured strategy of the analyst. Secondly I verify that given the investor's beliefs and strategy, the strategy of the analyst is indeed optimal. Finally I discuss off-equilibrium beliefs.

2.9.1 Investor's Optimal Beliefs

2.9.1.1 Updating Rules about the State x

The analyst may be 'smart', S, or 'dumb', D. The analyst may also be 'good', G, which means that her $k \geq k^*$, or 'evil', E, if $k < k^*$. I denote the four possible analyst types as SG, SE, DG, and DE respectively. In our posited equilibrium, good analysts report their private signal truthfully via either the high or low message, while evil analysts *always* report the high message. In addition smart analysts receive informative private signals, while dumb analysts receive noise.

If the investor receives the low message m_L then he updates his beliefs about the state x using Bayes rule as follows:

$$\lambda^*(m_L) = \Pr(x = x_H \mid m_L) = \frac{P(m_L \mid x_H)P(x_H)}{P(m_L \mid x_H)P(x_H) + P(m_L \mid x_L)P(x_L)}$$

where:

$$P(m_L \mid x_H) = P(m_L \mid x_H, GH)P(GH) + P(m_L \mid x_H, GD)P(GD)$$
$$+P(m_L \mid x_H, BH)P(BH) + P(m_L \mid x_H, BD)P(BD)$$
$$= (1-p)\theta\pi + 0 + 0.5(1-\theta)\pi + 0$$
$$= (0.5-(p-0.5)\theta)\pi$$

$$P(m_L \mid x_L) = P(m_L \mid x_L, GH)P(GH) + P(m_L \mid x_L, GD)P(GD)$$
$$+P(m_L \mid x_L, BH)P(BH) + P(m_L \mid x_L, BD)P(BD)$$
$$= p\theta\pi + 0 + 0.5(1-\theta)\pi + 0$$
$$= (0.5+(p-0.5)\theta)\pi$$

So the updated belief about the state x after a low message is given by:

$$\lambda^*(m_L) = \Pr(x = x_H \mid m_L) = 0.5-(p-0.5)\theta$$

If the investor receives the high message, m_H then he updates his beliefs about the state x using Bayes rule as follows:

$$\lambda^*(m_H) = \Pr(x = x_H \mid m_H) = \frac{P(m_H \mid x_H)P(x_H)}{P(m_H \mid x_H)P(x_H) + P(m_H \mid x_L)P(x_L)}$$

where:

$$P(m_H \mid x_H) = P(m_H \mid x_H, GH)P(GH) + P(m_H \mid x_H, GD)P(GD)$$
$$+P(m_H \mid x_H, BH)P(BH) + P(m_H \mid x_H, BD)P(BD)$$
$$= p\theta\pi + 1.\theta(1-\pi) + 0.5(1-\theta)\pi + 1.(1-\theta)(1-\pi)$$
$$= 1-0.5\pi + (p-0.5)\theta\pi$$

$$P(m_H \mid x_L) = P(m_H \mid x_L, GH)P(GH) + P(m_H \mid x_L, GD)P(GD)$$
$$+P(m_H \mid x_L, BH)P(BH) + P(m_H \mid x_L, BD)P(BD)$$
$$= (1-p)\theta\pi + 1.\theta(1-\pi) + 0.5(1-\theta)\pi + 1.(1-\theta)(1-\pi)$$
$$= 1-0.5\pi - (p-0.5)\theta\pi$$

So the updated belief about the state x after a high message is given by:

$$\lambda^*(m_H) = P(x = x_H \mid m_H) = 0.5+(p-0.5)\theta\left(\frac{\pi}{2-\pi}\right)$$

We can see that $\left|\lambda^*(m_H)-\lambda\right|<\left|\lambda^*(m_L)-\lambda\right|$, namely the low message induces a larger absolute revision in belief since the investor knows this message only ever comes from a good (ie truthful) analyst.

2.9.1.2 Updating Rules about the Analyst's Type θ

At the end of the period, the investor compares the message sent by the analyst (either m_H or m_L) to the realised state of the world (x_H or x_L). The investor then updates his beliefs about the analyst's ability using Bayes rule. The four possible outcomes are outlined below. If the analyst sends a low message, the investor knows that this is truthful. If the analyst sends a high message the investor must adjust his inference for the probability that the analyst is lying. A straightforward application of Bayes rule yields the following results.

Case 1: Low message and low realised state

$$P(smart \mid m_L \,\&\, x_L)=\frac{P(m_L \,\&\, x_L \mid smart)P(smart)}{P(m_L \,\&\, x_L \mid smart)P(smart)+P(m_L \,\&\, x_L \mid dumb)P(dumb)}$$

$$=\frac{p\theta}{0.5+(p-0.5)\theta}$$

Case 2: Low message and high realised state

$$P(smart \mid m_L \,\&\, x_H)=\frac{(1-p)\theta}{0.5-(p-0.5)\theta}$$

Case 3: High message and high realised state

$$P(smart \mid m_H \,\&\, x_H)=\frac{(0.5+(p-0.5)\pi)\theta}{0.5+(p-0.5)\pi\theta}$$

Case 4: High message and low realised state

$$P(smart \mid m_H \,\&\, x_L) = \frac{(0.5-(p-0.5)\pi)\theta}{0.5-(p-0.5)\pi\theta}$$

2.9.2 Investor's Desired Trade Size

At any given time, the investor chooses his shareholding $\delta = \delta(\lambda)$ to maximise expected utility as a function of his beliefs about the final state λ.

$$\max_{\delta} \lambda \ln(W + \delta(x_H - P_0)) + (1-\lambda)\ln(W + \delta(x_L - P_0))$$

The first order condition gives the optimal holding as:

$$\delta(\lambda) = \frac{-W(\lambda(x_H - P_0) + (1-\lambda)(x_L - P_0))}{(x_H - P_0)(x_L - P_0)}$$

Now the desired trade is the difference between $\delta(\lambda)$ and $\delta(\lambda^*)$. This is given as:

$$\delta(\lambda^*) - \delta(\lambda) = \alpha\left(\lambda^* - \lambda\right)$$

$$\text{where } \alpha = \frac{-W(x_H - x_L)}{(x_H - P_0)(x_L - P_0)} > 0$$

$$\text{since } \; x_L - P_0 < 0 < x_H - P_0 .$$

So the trade is directly proportional to the revision in the investor's beliefs $\lambda^* - \lambda$.

Given the presence of short sales constraints, and the explicit results for the revision in the investor's beliefs, we can work out the expected trade sizes generated by a positive and a negative message.

For a positive message the actual (and expected) trade size is given as:

$$Trade(m_H) = \alpha\left(\lambda^*(m_H) - \lambda\right) = \alpha\left(p - 0.5\right)\theta\left(\frac{\pi}{2-\pi}\right)$$

For a negative message and non-binding short-sales constraints trade size is given as:

$$Trade(m_L \mid notbinding) = \alpha\left(\lambda^*(m_L) - \lambda\right) = -\alpha\left(p - 0.5\right)\theta$$

If the short-sales constraints are binding (with probability q) then the trade is zero.

Thus the expected trade size from a negative message is given as:

$$E\left(Trade(m_L)\right) = -(1-q)\alpha\left(p - 0.5\right)\theta$$

Finally, the expected difference in the absolute trade size between a positive and negative message is given as:

$$E\left(\left|Trade(m_{high})\right| - \left|Trade(m_{low})\right|\right) = \alpha\left(p - 0.5\right)\theta\left(q + \left(\frac{\pi}{2-\pi}\right) - 1\right) > 0 \qquad (5)$$

Given the restriction that $q > 1 - \left(\frac{\pi}{2-\pi}\right)$ then the trade generated by a high message will be larger than the trade generated by a low message. This restriction implies short-sales constraints are sufficiently large to offset the additional credibility of a low message.

2.9.3 Analyst's Optimal Strategy

Given the beliefs and strategy of the investor, we now need to show that the analyst's strategy is indeed optimal. There are two possible cases. In the first case the analyst receives a high private signal. In the second case she receives a low private signal. We analyse the two situations as follows.

Case 1: When the analyst receives the high private signal, she reports the high private message irrespective of her type k. The intuition behind this result is straightforward. A high message generates higher trading volume and additionally generates a higher expected end-of-period reputation θ^* since:

$$E(\theta^* \mid m_H, s_H) = \theta$$

$$E(\theta^* \mid m_L, s_H) = \frac{(0.5-(p-0.5)\theta)p\theta}{0.5+(p-0.5)\theta} + \frac{(0.5+(p-0.5)\theta)(1-p)\theta}{0.5-(p-0.5)\theta} < \theta$$

Case 2: When the analyst receives the low private signal the message she sends depends on her type. A high message generates more trading volume than a low signal, however a low messages generates a higher expected end-of-period reputation than the high message. So the analyst must trade off these two factors. The expected end-of-period reputation from the two messages is given by:

$$E(\theta^* \mid m_L, s_L) = \theta$$

$$E(\theta^* \mid m_H, s_L) = \frac{(0.5-(p-0.5)\theta)(0.5+(p-0.5)\pi)\theta}{0.5+(p-0.5)\pi\theta} + \frac{(0.5+(p-0.5)\theta)(0.5-(p-0.5)\pi)\theta}{0.5-(p-0.5)\pi\theta} < \theta$$

The expected loss in reputation from lying is given by:

$$E(\theta^* \mid s_L, m_L) - E(\theta^* \mid s_L, m_H) = \frac{\theta^2(1-\theta)(p-0.5)^2\pi(1+\pi)}{0.25-(p-0.5)^2\theta^2\pi^2} > 0$$

From equation (3) (in section 2.5.1) we can see that the analyst will tell the truth if k is sufficiently large. There exists a critical level of k, given by k^* at which the analyst will be indifferent between reporting the high or low message. This critical value of k is given by solving equation (3) and equals:

$$k^* = \frac{c\alpha\left(\frac{\pi}{2-\pi}+q-1\right)(0.25-(p-0.5)^2\theta^2\pi^2)}{\pi(1+\pi)(p-0.5)(1-\theta)\theta} \qquad (7)$$

So if the analyst is of type $k \geq k^*$ she will truthfully report the *low* message. If her type is $k < k^*$ then she will lie and report the *high* message. The probability that the analyst is honest is given by $\pi = \Pr(k \geq k^*)$. The analyst type k is distributed on $[0,\infty)$ according to the distribution function $F(k)$. Given a specific distribution function we can express π in terms of the underlying model parameters. However for our purposes it is not necessary to specify a specific $F(k)$ so long as it has some mass to either side of the critical value of k^*.

2.9.4 Other Considerations

We have shown that the investor's beliefs and strategy is optimal given the analyst's strategy. Also the analyst's strategy is optimal given the investor's beliefs and strategy. In this model, we do not need to check the investor's off-equilibrium beliefs since all the information sets of the investor are reached with positive probability. As such, it is sufficient to show that the investor uses Bayes law to update beliefs at each of his information sets, which he indeed does. Thus the equilibrium is a perfect Bayesian equilibrium[37].

2.9.5 Derivation of Equilibrium Outcomes

Outcome 1: It is straightforward to show that $\frac{\partial(\delta_2^H - \delta_1)}{\partial\theta} > 0$ and $\frac{\partial(-\delta_2^L + \delta_1)}{\partial\theta} > 0$.

Thus higher reputation analysts generate higher trading volume.

Outcome 2: Since:

$$P(smart \mid m_L, x_L) > P(smart \mid m_H, x_H) > P(smart \mid m_H, x_L) > P(smart \mid m_L, x_H)$$

We can see that if the analyst's message is ex-post correct ie $m = x$, her posterior reputation will be higher than if her message is incorrect.

Outcome 3: See equation (5).

Outcome 4: The average message sent in equilibrium is given as:

$$E(m) = P(good)(0.5(0) + 0.5(1)) + P(evil)(0(0) + 1(1)) = 1 - 0.5\pi$$

---

[37] Given all the information sets of the investor are reached with positive probability the perfect Bayesian equilibrium is equivalent to the weak perfect Bayesian equilibrium for this game.

2.10. Appendix 3: Proof of Proposition 2

Given the critical value of k^* in equation (7) we want to prove the two partial derivative results in Proposition 2. To enable an analytic solution[38], assume k is uniformly distributed. Then we have $\pi = 1-k$. Define two new variables, $s = \theta(2p-1)$ and $\gamma = \frac{c\alpha}{2(1-\theta)}$ then equation (7) simplifies to:

$$k = \frac{\gamma(q+kq-2k)(1-s^2(1-k)^2)}{s(1+k)(1-k)(2-k)}$$

Define the new function $F(k,q,s)$ as:

$$F(k,q,s) = sk(1+k)(1-k)(2-k) - \gamma(q+kq-2k)(1-s^2(1-k)^2) \qquad (8)$$

Then equation (7) is equivalent to $F(k,q,s)=0$. To prove our result, we first need to prove the following lemma.

Lemma 1: If $F(k,q,s)=0$ then $\frac{\partial F}{\partial k}(k,q,s) > 0$.

Proof: Differentiating (8) we get:

$$\frac{\partial F}{\partial k}(k,q,s) = 2s(1-k-3k^2+2k^3) - \gamma(q-2+s^2(1-k)(2+3kq-6k+q)) \qquad (9)$$

Also we can solve equation (8) for γ using $F(k,q,s)=0$ to get:

$$\gamma = \frac{sk(1+k)(1-k)(2-k)}{(q+kq-2k)(1-s^2(1-k))}$$

Substituting this into (9) gives:

---

[38] I thank Michael Ulm for significant assistance in analytically proving this result and Harjoat Bharmra for valuable suggestions.

$$\frac{\partial F}{\partial k}(k,q,s)=2s(1-k-3k^2+2k^3)-\frac{sk(1+k)(1-k)(2-k)}{(q+kq-2k)(1-s^2(1-k)^2)}(q-2+s^2(1-k)(2+3kq-6k+q))$$

This is positive if and only if:

$$\begin{aligned}&2(1-k-3k^2+2k^3)(q+kq-2k)(1-s^2(1-k)^2)\\&>k(1+k)(1-k)(2-k)(q-2+s^2(1-k)(2+3kq-6k+q))\end{aligned}\qquad(10)$$

Since q enters this expression only linearly, the inequality is true for all values of q if it is true for the values $q=0$ and $q=1$. To see this, the inequality can be written as $qS>T$, so if $0>T$ and $S>T$ then $qS>T$ for $0<q<1$. The variable s enters the equation only in the form s^2 so we can use the same argument. If the inequality holds for $s=0$ and $s=1$ then it holds for all s. So we need to consider four cases.

Case 1: $q=0,s=0$. Inequality (10) simplifies to:

$$2(1-k-3k^2+2k^3)(-2k)>k(1+k)(1-k)(2-k)(-2)$$

or $2k^2(1+4k-3k^2)>0$ which is clearly true since $0\le k<1$.

Case 2: $q=0,s=1$. Inequality (10) simplifies to:

$$2k^2(k-2)^2(1+k^2)>0 \qquad \text{which is true.}$$

Case 3: $q=1,s=0$. Inequality (10) simplifies to:

$$(1-k)^2(2+2k-3k^2)>0 \qquad \text{which is true.}$$

Case 4: $q=1,s=0$. Inequality (10) simplifies to:

$$k^2(1-k)^2(2-k)^2>0 \qquad \text{which is true.}$$

So the inequality is true for all relevant values of q, s . Thus the lemma holds. ■

Using this lemma it is straightforward to prove Proposition 2. By the implicit function theorem we have:

$$\frac{\partial k}{\partial q} = -\frac{\partial F / \partial q}{\partial F / \partial k}$$

and since: $\frac{\partial F}{\partial q} = -\gamma(1-k)(1-s^2(1-k)^2) < 0$

and by the Lemma $\frac{\partial F}{\partial k} > 0$

Combining these results we get $\frac{\partial k}{\partial q} > 0$ as required for Proposition 2.1.

Similarly,

$$\frac{\partial k}{\partial s} = -\frac{\partial F / \partial s}{\partial F / \partial k}$$

Now since: $\frac{\partial F}{\partial s} = k(1+k)(1-k)(2-k) + 2s\gamma(q+kq-2k)(1-k)^2 > 0$

and by the Lemma $\frac{\partial F}{\partial k} > 0$.

Combining these results, we get $\frac{\partial k}{\partial s} < 0$

Finally, since $s = \theta(2p-1)$ we get $\frac{\partial k}{\partial p} < 0$ as required for Proposition 2.2. ■

2.11. Appendix 4: Proof of Proposition 3

To prove this proposition we need to modify the results from Proposition 1 to account for the fact that with some probability the investor is naïve. The proof is identical to Proposition 1, however now the analyst faces a naïve investor with probability ξ. With probability $1-\xi$ she faces a rational investor (who is aware that the naïve investor also exists).

The rational investor has a belief π that the analyst is a good analyst (ie honest). The analyst faces the following trade-off between short-term gains in commission from a high message versus an expected loss in reputation. The short-term gain in commission now becomes:

$$E(Trade|m_H)-E(Trade|m_L)=\alpha(p-0.5)\theta\left(q+(1-\xi)\left(\frac{\pi}{2-\pi}-1\right)\right)$$

The expected loss in reputation becomes:

$$E(\theta^*|m_L)-E(\theta^*|m_H)=$$
$$\frac{\theta^2(1-\theta)(p-0.5)^2\left(2(0.25-(p-0.5)^2\theta^2\pi^2)\xi+\pi(1+\pi)(0.25-(p-0.5)^2\theta^2)(1-\xi)\right)}{(0.25-(p-0.5)^2\theta^2)(0.25-(p-0.5)^2\theta^2\pi^2)}$$

The critical level k^N at which the analyst is indifferent between the two messages is:

$$k^N=\frac{c\alpha(0.25-(p-0.5)^2\theta^2)(0.25-(p-0.5)^2\theta^2\pi^2)\left(q+(1-\xi)\left(\frac{\pi}{2-\pi}-1\right)\right)}{\theta(1+\theta)(p-0.5)\left(2(0.25-(p-0.5)^2\theta^2\pi^2)\xi+\pi(1+\pi)(0.25-(p-0.5)^2\theta^2)(1-\xi)\right)} \quad (9)$$

This replaces the previous critical level k^* and the remainder of the equilibrium is otherwise identical. The naïve investor behaves as the investor in Proposition 1 with $\pi=1$. The rational investor here behaves as the investor in Proposition 1, but replaces k^* with k^N when forming his beliefs.

We now move to proving the second part of Proposition 3, namely that $\frac{\partial k^N}{\partial \xi} > 0$.

Let $s = \theta(2p-1) > 0$ and $\gamma = \frac{c\alpha(0.25-(p-0.5)^2\theta^2)}{\theta(1-\theta)(p-0.5)} > 0$:

We can simplify (9) to the following expression:

$$F(s,k,\xi) = k(1+k)2(1-s^2(1-k)^2)\xi + k(1+k)(1-k)(2-k)(1-s^2)(1-\xi)$$
$$-\gamma(1-s^2(1-k)^2)((1+k)q-(1-\xi)2k) = 0$$

or alternatively:

$$F(s,k,\xi) = q(s^2-1)\gamma - k\left[-2+s^2(2+\gamma(2+q-2\xi))+\gamma(-2+q+2\xi)\right]$$
$$+k^2\left[-1+3\xi+s^2(1+\xi-\gamma(q-4+4\xi))\right]+k^3\left[2(\xi-1)+s^2(2+\gamma(-2+q+2\xi))\right]$$
$$-k^4\left[-1+\xi+s^2(1+\xi)\right] = 0$$

Differentiation yields:

$$\frac{\partial F}{\partial k} = 2-s^2(2+\gamma(2+q-2\xi))-\gamma(-2+q+2\xi)+2k\left[-1+3\xi+s^2(1+\xi-\gamma(-4+q+4\xi))\right]$$
$$+3k^2\left[2(\xi-1)+s^2(2+\gamma(-2+q+2\xi))\right]-4k^3\left[-1+\xi+s^2(1+\xi)\right] > 0$$

and:

$$\frac{\partial F}{\partial \xi} = -k\left[3+s^2(1-4\gamma)\right]+k^2\left[3+s^2(1-4\gamma)\right]+k^3\left[2+2s^2\gamma\right]-k^4\left[1+s^2\right] < 0$$

Given the earlier results from Lemma 1, implicit differentiation yields the following result:

$$\frac{\partial k}{\partial \xi} = -\frac{\partial F/\partial k}{\partial F/\partial \xi} > 0$$

as required. ■

2.12. Appendix 5: Cluster Adjusted t-statistics[39]

In the pooled regressions, the same analyst produces earnings estimates for several stocks in a given year. This implies observations within a single analyst-year cluster may be correlated if common shocks affect all stocks covered by the analyst in a similar way. More generally in the presence of clustering, the observations within a given cluster may not be treated as independent, even though the clusters themselves may be considered independent. The detailed calculation of the covariance matrix allowing for clustering is given as follows, comparing it to the OLS estimator, and the White estimator.

OLS Variance Estimator:

$$V_{OLS} = s^2 (X'X)^{-1}$$

$$\text{where: } s^2 = \left(\frac{1}{N-k}\right)\sum_{i=1}^{N} e_i^2$$

Robust (White) Unclustered Variance Estimator:

$$V_{rob} = (X'X)^{-1}\left[\sum_{j=1}^{n}(e_i x_i)'(e_i x_i)\right](X'X)^{-1}$$

Robust Cluster Variance Estimator:

$$V_{cluster} = (X'X)^{-1}\left[\sum_{j=1}^{n_C} u_j' u_j\right](X'X)^{-1}$$

$$\text{where: } u_j = \sum_{j_{cluster}} e_i x_i$$

In the calculations, e_i is the i^{th} residual and x_i is a row vector of predictors.

[39] Source: http://www.stata.com/support/faqs/stat and Williams (2000).

The variance of the clustered estimator will be higher than the robust estimator when there is positive correlation between elements of the cluster. Positive correlation means the cluster sums have more variability than the individual elements. Negative correlation within a cluster will have the reverse effect.

3. The Aggregate Behaviour of Individual Investors

3.1. Introduction

The ability of behavioural models in finance to explain asset returns rests on three key assumptions. Firstly, a group of investors must deviate from rationality in their trading activity. Secondly, these errors must be sufficiently systematic across the group to avoid them being diversified away upon aggregation. Thirdly, there must be some limits to arbitrage that prevent rational investors immediately correcting any resulting mispricing. While much literature has addressed various deviations from rationality and limits to arbitrage[40], to date relatively little empirical research has focussed on the second assumption. That is, whether individual investors' deviations from rationality (ie errors) are sufficiently systematic to aggregate in a meaningful way.

The aggregation assumption plays a key role in models where the existence of a group of commonly acting irrational investors (or equivalently a representative irrational agent) is evoked[41]. Shleifer (2000) argues that errors *should* aggregate based on the systematic nature of the behavioural biases demonstrated by Kahneman and Tversky (1982) and others. However, since much of this work is experimentally

---

[40] For a survey of deviations from rationality see Barberis and Thaler (2001), for limits to arbitrage see Shleifer (2000).

[41] See for example DeLong et al (1990) where correlated noise traders cause prices to deviate from their true values, Hong and Stein (1999) where a group of boundedly rational momentum traders exist who trade based on past returns, and Baker and Stein (2002) where a group of irrational investors exist who underreact to the information contained in order flow.

based, it is important to examine this assertion in real-life financial settings where agents typically face strong economic incentives to avoid such biases.

This chapter examines the aggregation assumption using a unique database of individual investor trades. The database is somewhat larger than that used in previous studies of small investors, containing data from 47 full-service and 9 internet brokerage firms in Australia over an 11-year period from 1991 to 2002[42]. Additional information from 16 institutional and 15 mixed brokerage firms (dealing with both retail and institutional clients) is also used.

I firstly examine the trading behaviour of a large group of individual investors to assess whether there are any systematic patterns in their trading that remain after aggregation. Secondly I examine whether the actions of unrelated subgroups of individual investors from a large number of independent brokerage firms are positively correlated. Finally I examine the relationship between aggregated trades and future returns. The analysis is performed at two levels, at the market level (examining flows into and out of the equity market as a whole) and at the cross-sectional level (examining flows into and out of individual stocks). Additionally the actions of individuals using both internet and traditional full-service brokers are compared.

Why individuals? Of all investor groups, individual investors seem a priori a likely candidate for a group that might be subject to systematic behaviourally induced errors. For example previous research shows that individuals appear to trade excessively, exercise options in a clearly irrational manner, sell stocks that outperform those they purchase, and hold undiversified stock portfolios with excessive allocations to their employers stock[43]. Additionally as a group, individuals

---

[42] This paper uses information from over 41 million small investor trades. By comparison Barber and Odean (2000) use around 3 million trades and Grinblatt and Keloharju (2001) use 2 million trades. Note however that both these papers have account level information, while this paper uses information aggregated at the individual broker level.

[43] See Barber and Odean (2000), Poteshman and Serbin (2003), Odean (1998), and Benartzi (2001) respectively.

are large, holding 42% of all outstanding equity in the US, 16% in the UK, 20% in Germany, and 24% in Australia. Individuals are also likely to have access to fewer resources relative to institutional investors when making their investment decisions. Some individual investors reduce this problem via trading through full-service brokerage firms where they receive investment advice. Others trade via internet or discount brokers where they receive none.

The first section of the chapter finds a number of strong systematic patterns in behaviour that survive after aggregation across investors from all 56 retail brokerage firms. To reduce the possibility of unintentional data-mining I limit my analysis to four groups of explanatory variables identified by previous small investor research; namely past returns and volatility, calendar effects, and lagged investor flows. In addition I conduct numerous robustness checks across sub-periods and across subsets of stocks. The strongest pattern I find is that individual investors appear to engage in significant levels of negative feedback trading (buying after falls, selling after rises) at *both* the market and cross-sectional levels. This result is consistent with previous studies of small investors at the cross-sectional level (Odean (1998), Grinblatt et al (2000)) however at the aggregate market level it differs from the results in the mutual fund flow literature (Edelen and Warner (2001)). Negative feedback trading is most pronounced for internet brokerage clients, but is also significant for full-service retail brokerage clients. I also find that the net flows of individual investors are extremely persistent over time especially at the cross-sectional level. Furthermore, individuals are net buyers after high individual stock volatility, and individual trading activity is high in the first half of January and in the second half of June (before tax-year end).

In the second section I examine small investor trades disaggregated to the individual broker level to assess the similarity of actions of small investors from independent brokerage firms. The results show that individual investors' actions are highly correlated between different brokerage firms. For example the average pairwise cross-sectional correlation for weekly net flows into the top 50 stocks between unrelated pairs of internet brokerage firms is 0.439. For full-service firms

the average pairwise correlation is 0.235. In contrast, for unrelated pairs of institutional brokers the correlation is zero. The correlation result for retail brokers is remarkably robust with significant positive correlation coefficients found for *every* single unique pair of retail brokerage firms over the sample period. This result is striking given there are 1081 unique retail broker pairs where overlapping data is available. Concurrent related research by Barber, Odean, and Zhu (2003) shows that investors *within* a *single* broker have strongly correlated trades. While correlation within a single broker may be due to common investment advice, or common social networks, it is highly unlikely that such explanations would be valid for the current chapter's result that positive correlations exist *across* 56 unrelated brokerage firms with millions of retail clients. As such the results here provide strong evidence that the aggregation assumption is likely to hold.

The final section of the chapter examines the relationship between individual investor flows and future stock returns at the market and cross-sectional levels. Vector autoregression analysis is used to allow for joint endogeneity between returns and flows. At the aggregate market level I find small investor flows positively predict future 1-week market returns with evidence of subsequent reversals, consistent with short-term price pressure. At the cross-sectional level, I find small investors predict future 2-3 week excess returns with no evidence of reversals, consistent with information motivated trading or liquidity provision. Further examination reveals that this result is entirely driven by full-service brokerage clients. In a related finding, Pan and Poteshman (2003) show that full-service brokerage clients' equity option trades positively forecast future stock returns. My results do not support the conjecture that behavioural errors are causing individual investors to act in an irrational manner at either the aggregate market or cross-sectional level.

In summary, the main contribution of this chapter is to provide strong evidence that the aggregation assumption *does* appear to hold for small investors. Aggregated flows from this group of investors *do* contain strong systematic patterns and the

actions of individual investors from a large number of unrelated retail brokerage firms are highly correlated. The final section of the chapter reveals however that behavioural models assuming small investors are in aggregate irrational or uninformed may not be justified, given the net flows of full-service investors *positively* predict future short-term cross-sectional and aggregate market returns.

The chapter is structured as follows. Section 3.2 describes the data used in the study. Section 3.3 examines aggregated small investor flows for evidence of systematic patterns. Section 3.4 examines correlations in net flows across a large number of brokerage firms. Section 3.5 examines the relationship between net flows and future returns. Section 3.6 concludes.

3.2. Data Description

This chapter uses a unique dataset of small investor trading on the Australian Stock Exchange. The dataset contains information on the trades of 56 retail brokers (9 internet brokers and 47 traditional full-service brokers) over the period September 1991 to December 2002[44,45]. For each of the 56 brokers, data is available on the number of trades, the value traded, and the average price of all buy and sell trades executed during each week, for each stock. The dataset is proprietary and was supplied by a leading Australian broker. Trading on the Australian Stock Exchange during this period was conducted using the automated SEATS system that operates via a fully transparent electronic open limit order book with no designated market makers[46]. In this system all trades are reported and tagged with a unique broker

---

[44] Not all brokers exist for the full period. For example most internet brokers start after 1995, while several brokers leave the sample over the 11-year period.

[45] Additional information on 16 institutional brokers and 15 mixed brokers (that deal with both retail and institutional clients) is also used.

[46] For more details on the institutional structure of the ASX market see Aitken et al (1995). Note that although the trading system does not include a designated market maker, there exists a well established pseudo-market making

identification code that enables us to track the specific broker engaged on each side of each transaction. The broker identification data used in this study was observable in real-time on a trade-by-trade basis by some market participants (brokers) but was not available to investors in general.

Summary statistics describing the dataset appear in Table 3.1. In aggregate around 21.6 million buy and 20.2 million sell transactions are represented with an aggregate value of $A191 billion and $A188 billion respectively. Additionally the sample is split into two groups[47], internet brokerage investors and traditional full-service brokerage investors. Table 3.1 shows that the clients of these retail brokers trade small amounts, with an average transaction size of $7,570 for internet brokerage investors and $10,080 for full-service brokerage investors[48]. As found in Barber and Odean (2000), the average value of stocks sold is slightly higher than the average value of stocks purchased ($9,292 vs $8,824) and the number of buy transactions is higher than the number of sell transactions.

system as evidenced in Fong et al (2001) who show that for the highest decile volume stocks, 34% of trading is done in the upstairs market (via crossings).

[47] Internet brokers were identified using the ASX 2001 Participants directory and confirmed via examining broker web sites.

[48] Occasionally institutional trades appear in the records for these brokers. To remove the influence of these trades, stock weeks with an average trade size of >$A50000 for either buy or sell transactions were removed and the data for the week was treated as missing. The number of records removed was small, 0.15% of Internet broker stock-weeks were removed, and 1.07% of full-service broker trades were removed.

Table 3.1: Summary Description of Aggregate Data (in $A)

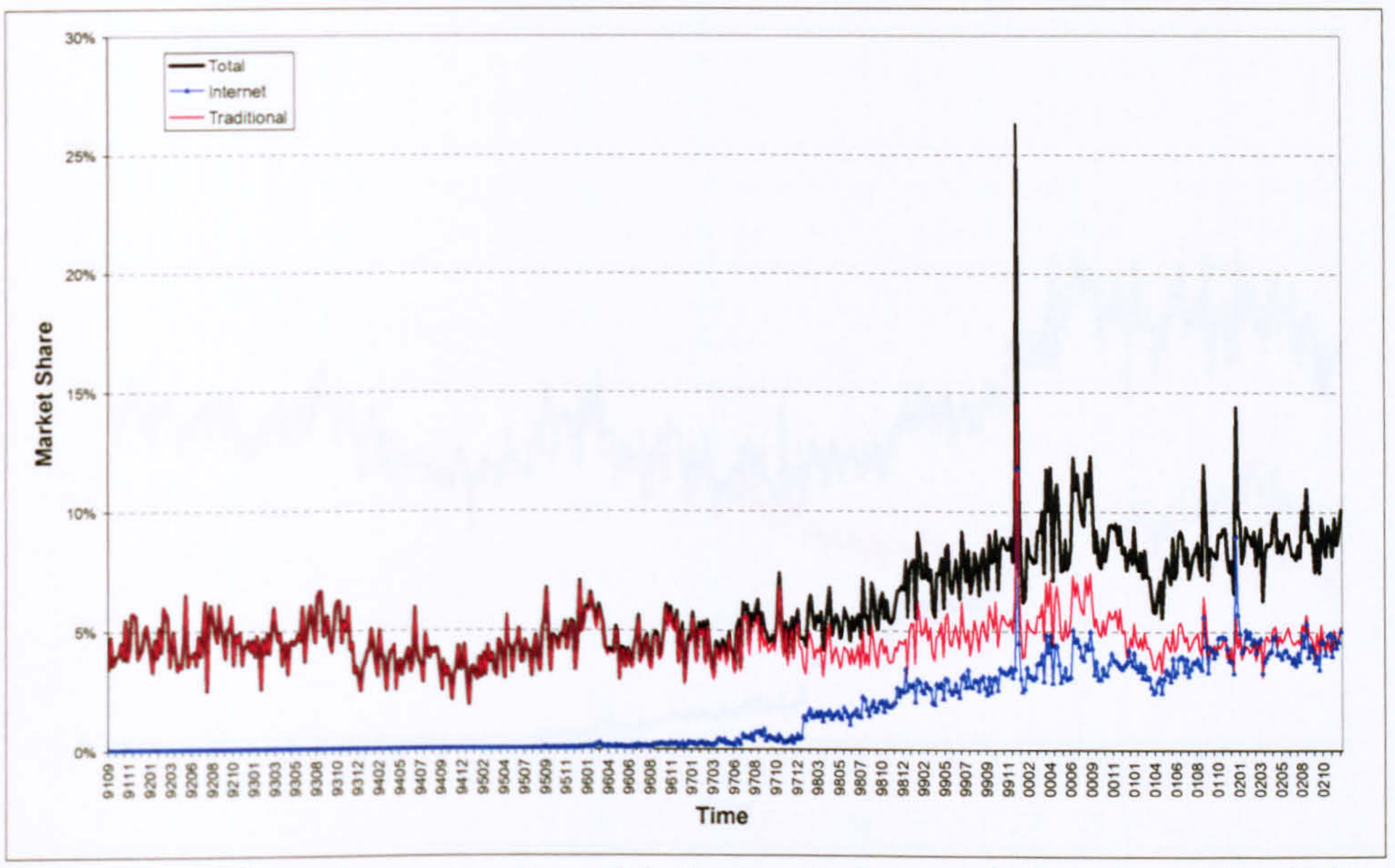
This table presents summary statistics on the data sample. Results are split into internet brokerage investors and traditional full-service brokerage investors. The sample period is September 1991 to December 2002.

	Broker Type		
Variable	**Internet**	**Traditional**	**All**
Total # Buys	9,305,566	12,351,483	21,657,049
Total # Sells	7,723,715	12,509,032	20,232,747
Total Buy Value ($M)	67,271	123,839	191,110
Total Sell Value ($M)	61,152	126,861	188,012
Average Buy Trade	7,229	10,026	8,824
Average Sell Trade	7,917	10,142	9,292
# Brokers	9	47	56
Stock-Weeks	597,985	2,664,365	3,262,350

The 56 brokers in the sample account for a significant proportion of all trades executed by small investors over this period. It is difficult to quantify this figure exactly for the full-service portion of the sample, as some small investors trade through full-service mixed brokerage houses that deal with both institutional and small investors. Based on the number of trades from these mixed firms I estimate that the percentage of all full-service retail brokerage trades captured by my 47 full-service brokers is over 50%. In contrast the 9 internet brokers in the sample capture virtually all internet based trades.

The economic significance of the sample can be demonstrated by calculating the total market share (by value traded) of these brokers relative to the total value traded in the entire market. Market share results are shown in Figure 3.1. The market share of the 47 full-service brokers remains at around 5% throughout the entire period. The market share of the 9 internet brokers grows rapidly from early 1998 to reach around 5% by the end of the sample period[49]. Figure 3.2 shows that the market share of these brokers calculated by the *number* of trades is significantly higher reaching around one-third of all trades executed in the final three years of the sample.

Figure 3.1: Aggregate market share of sampled brokers over time (by $ value)

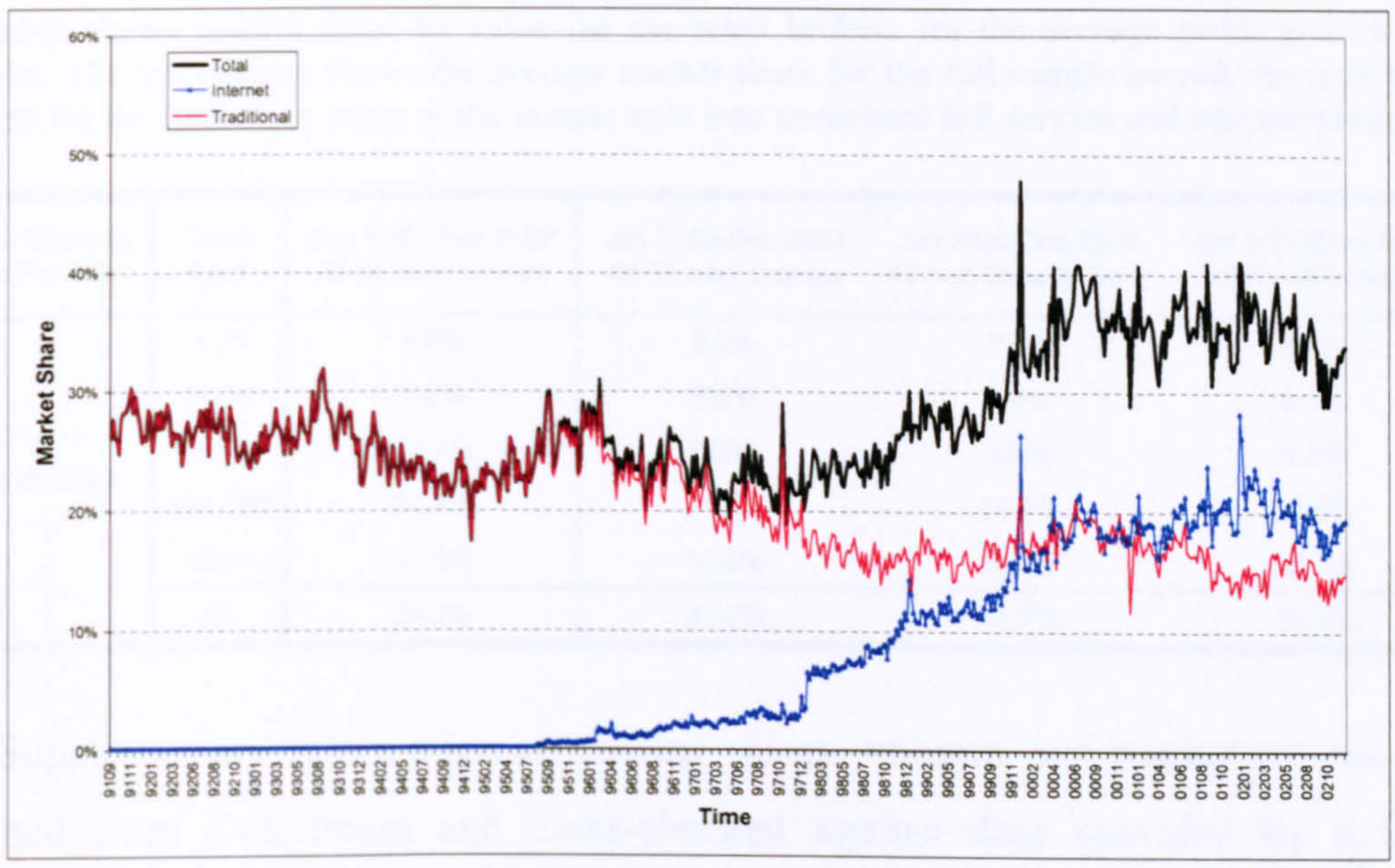


This figure shows the market share of the brokers in the sample from September 1991 to December 2002. Market share is calculated as the value of trades done through these brokers relative to the value of total market turnover.

While the combined aggregate market share of less than 10% may lead us to conclude that this group of investors is likely to have little impact on market dynamics, the market share of these investors in the *typical* index stock is significantly higher. As found in Lee, Shleifer, and Thaler (1991), market share of small investors increases as firm size decreases. Table 3.2 shows the average market share (by value) for stocks in five different size groupings. For large stocks the market share of these investors is small, however for the average index stock in the sample the market share of these investors is around 25% for the full sample, and around 33% in the most recent three years. This indicates that small investors are of sufficient size to potentially influence market dynamics for the average index stock.

[49] A large spike in market share can be seen in late December 1999 driven by the Y2K effect, with extremely low institutional volume in the week prior to the new millennium.

Figure 3.2: Aggregate market share of sampled brokers over time (by # trades)



This figure shows the market share by number of trades of the brokers in the sample from September 1991 to December 2002. Market share is calculated as the number of trades done through these brokers relative to the total number of trades.

This chapter restricts the universe of stocks to those included in the All Ordinaries Index up to April 2000, and stocks in the ASX300 index for the remainder of the period[50]. These stocks represent over 90% of total outstanding market capitalisation over the period, are held by both institutional and individual investors, and are of reasonable liquidity. The number of stocks in the index varies between 250 and 350 for each week of the sample with the total market cap of the index varying between $A160bn and $A680bn. Stocks outside the All Ordinaries suffer from extreme illiquidity and resulting non-synchronous trading problems. As a group these micro-cap stocks are not particularly economically important relative to the market as a whole and are omitted from this study.

---

[50] On the 3rd April 2000 the construction of the All Ordinaries index was changed significantly, with the number of stocks in the index going from around 266 to 500. The ASX300 post April 2000 is the closest market index to the All Ordinaries prior to the changes and became the equivalent institutional benchmark. Source: ASX Factbook 2001.

Table 3.2: Market Share of Retail Brokers By Size Group

This table shows market share by value for the retail brokers for the average stock in each size portfolio. The left column shows the average market share for the full sample period, the right three columns for the final three years of the sample split into traditional full-service and internet brokerage firms.

Market Share by Size Portfolio	Stock Rank	Sep 1991-Dec 2002 All 56 retail brokers	Jan 2000-Dec 2002 All 56 retail brokers	Jan 2000-Dec 2002 47 traditional brokers	Jan 2000-Dec 2002 9 internet brokers
Size Portfolio	1-20	4.9%	6.2%	2.8%	3.5%
	21-50	7.2%	9.2%	3.9%	5.3%
	51-100	12.7%	13.9%	5.6%	8.3%
	101-200	25.4%	34.5%	14.7%	19.8%
	200+	41.9%	57.4%	26.7%	30.7%
	All	24.2%	33.2%	14.7%	18.5%

Supplementary information on closing prices, volume, and market returns was obtained from Datastream and cross-checked against data provided by a large Australian investment management firm[51]. Information on daily index composition was obtained from a brokerage firm and verified against the ASX Monthly Index Almanacs.

Finally, it is interesting to know how potentially generalizable the results in this chapter are to other equity markets. Table 3.3 presents some summary statistics on the ownership composition and size of the Australian market relative to other international markets. Individual investor participation is high in Australia (41%) relative to other countries. However the actual percentage of the market owned by individuals (20%) is similar to levels in Japan, UK, and France but is lower than US individual ownership (42%). Foreign ownership of Australia equities (24%) is similar to that of Japan and UK, but significantly higher than foreign ownership in the US (11%).

[51] Where data did not match I performed a third check against independent data from the ASX monthly almanacs and data from a brokerage firm. This process identified several errors in the Datastream series, a list of which are available from the author on request.

Table 3.3: Descriptive Market Statistics for 2001 (or most recent available)

Country	# Individual Shareholders (m)	Participation Rate	% Market Owned by Retail Investors	Ownership by Foreign Investors	Turnover Velocity	# Listed Companies	Market Capitalisation (US$bn)
Australia	5.7	41%	24%	24%	67%	1410	375
Finland	0.8	18%	12%	51%	99%	155	190
France	5.6	13%	17%	14%	138%	1132	1844
Germany	na	na	17%	14%	118%	983	1072
Hong Kong	1.0	17%	na	na	44%	867	506
Ireland	0.4	13%	na	na	24%	87	75
Japan	32.1	30%	18%	19%	60%	2141	2294
UK	11.5	24%	16%	33%	84%	2332	2165
US	33.8	26%	42%	11%	141%	7069	13827

3.3. Systematic Patterns in Aggregate Individual Trading Behaviour

This section examines whether there are systematic patterns that exist in the aggregated net flows of the 56 retail brokers. Since all completed trades have both a buyer and a seller, then across the entire market net flows must sum to zero for any given stock each week. As such, if we take a randomly chosen group of traders and aggregate their trades, then in a regression of aggregate net trades on any factor we choose we should observe a zero loading. For example since net trades sum to zero, not all investors can be momentum investors. If however we aggregate the net flows from a group of investors that we believe a priori may behave in a similar way (here small investors), then we may get significant coefficients if we regress these flows on factors that capture the commonality in behaviour.

To limit the possibility of unintentional data mining I limit my analysis to four groups of factors identified by previous small investor research. Specifically, the next two sub-sections examine the relationship between net flows of small investors and past net flows, current and past returns, calendar effects, and current and past volatility. This is done firstly at the market level (flows into and out of the equity market as a whole) and secondly at the cross-sectional level (flows into and out of

individual stocks). In addition buy, sell, and total flows are examined individually to further understand the nature of the patterns found.

From a behavioural perspective, if no patterns were found that survived after aggregation, then the crucial second assumption of many behavioural models in finance would be violated for this particular group of investors and choice of regressors. Conversely if patterns in aggregate trades are found, this would support the aggregation assumption.

3.3.1 Market Level Flows

Market level flows represent the total dollar amount injected or removed from the equity market as a whole by small investors each week. Market level buy (sell) flows are calculated by summing the total value of buy (sell) orders across all stocks each week for all retail brokers.

Market level *net flows* are defined as:

$$NetFlowMkt_t = \frac{buyflow_t - sellflow_t}{buyflow_t + sellflow_t}$$

This gives a value between +1 (100% buys) and –1 (100% sells) each week. When analysing buy, sell, and total market flows separately, I detrend these series to ensure stationarity (see Appendix 1 for a discussion of the methodology).

Net flows, together with buy, sell, and total flows are regressed on the following factors:

- Contemporaneous and lagged weekly market returns
- Lagged flow information
- Calendar dummies
- Contemporaneous and lagged intra-week market return volatility

For each flow variable, two regressions are run. The first regression includes contemporaneous return and volatility variables together with four weekly lags for returns, flows, and volatility. Two longer-term lagged return and flow measures are also included. The second regression uses a reduced number of regressors, with the specification chosen by a recursive procedure designed to minimise the Akaike Information Criterion (AIC). Table 3.4 presents the results of the analysis.

The first result from Table 3.4 is that aggregate market level net flows are significantly *negatively* related to contemporaneous and lagged aggregate market returns. In economic terms the relationship is significant. A 10% positive return leads to a –0.273 contemporaneous change in the net flow variable (equivalent to a –2.46 standard deviation shock). The result contrasts with *mutual fund* flows where previous research finds a *positive* relationship between mutual fund inflows and contemporaneous aggregate market returns (Warther (1995)), and a *positive* relationship between fund flows and lagged short-term returns (Edelen and Warner (2001))[52]. The difference between flows into equity mutual funds and direct equities is somewhat puzzling. However the negative feedback trading result is consistent with the *cross-sectional* results of Grinblatt and Keloharju (2001) and Odean (1998) that find individual investor net flows are negatively related to recent returns[53].

The second result involves lagged flows. Warther (1995) finds that aggregate mutual fund flows are highly persistent. Consistent with this result, we can see from Table 3.4 that market level net flows are also strongly persistent for the retail brokers in this sample.

---

[52] Cohen (1999) also finds a positive relationship between individual's allocation to equities and past market returns over longer horizons. He attributes this result to relative risk aversion of households decreasing in wealth.

[53] These latter papers explain this result via the disposition effect of Shefrin and Statman (1985). Alternative rational explanations may include portfolio rebalancing or individuals acting as liquidity providers to the market. For the objectives of this chapter we are not concerned whether the patterns are rationally or behaviourally induced, we are simply interested whether systematic patterns exist that survive in a large group of small investors.

Table 3.4: Aggregate Weekly Flow Regressions: Net Flows, Buy Flows, Sell Flows and Total Flows

This table analyses aggregated weekly net flows, buy flows, sell flows, and total flows. Aggregated weekly flows are calculated by summing buys and sells for all individual stocks across all brokers each week. Buy, sell, and total flows are calculated as de-trended raw flows. Net flows are calculated as (buyflow-sellflow)/(buyflow+sellflow). 'MktRet' is the weekly market return and 'Volatility' is the weekly return volatility calculated using the daily market returns. T-statistics are calculated using the Newey-West method with four lags. The period of the regression is September 1991 to December 2002 and the frequency is weekly.

Variable	Net Flows				Buy Flows				Sell Flows				Total Flows				Variable
	Full Model		Reduced model		Full Model		Reduced model		Full Model		Reduced model		Full Model		Reduced model		
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		p-value
adjR2	57.4%		58.0%		56.6%		56.9%		63.2%		63.7%		60.9%		61.3%		adjR2
AIC																	AIC
Constant	0.00	0.31	0.01	1.75	-0.13	-0.90	-0.07	-0.78	-0.06	-0.42	-0.11	-1.42	-0.09	-0.57	-0.13	-1.12	Constant
MktRet (t)	-2.73	-12.01	-2.71	-11.36	-2.85	-1.75	-2.92	-1.80	14.82	8.42	14.71	8.33	6.74	4.10	6.82	4.16	MktRet (t)
MktRet (t-1)	-0.77	-3.22	-0.58	-2.79	1.28	0.76			6.41	3.30	6.16	3.30	4.19	2.42	4.49	2.56	MktRet (t-1)
MktRet (t-2)	-0.30	-1.28			4.40	2.45	4.09	2.35	5.91	3.51	5.47	3.21	5.35	3.22	5.69	3.44	MktRet (t-2)
MktRet (t-3)	0.08	0.38			3.37	2.00	3.70	2.24	3.33	2.01	3.37	2.04	3.54	2.26	3.85	2.59	MktRet (t-3)
MktRet (t-4)	-0.04	-0.20			3.09	1.91	3.82	2.39	1.49	1.01			2.42	1.68	2.91	2.09	MktRet (t-4)
MktRet (t-5:t-8)	-0.15	-1.41	-0.15	-1.62	1.11	1.25	1.72	2.13	2.25	2.45	2.11	2.73	1.70	1.88	2.01	2.36	MktRet (t-5:t-8)
MktRet (t-9:t-12)	0.12	1.20			1.82	2.08	1.94	2.34	0.91	1.27			1.52	1.94	1.38	1.82	MktRet (t-9:t-12)
Flow (t-1)	0.41	9.25	0.45	12.14	0.39	11.18	0.42	13.36	0.39	8.71	0.41	10.88	0.38	9.91	0.40	11.40	Flow (t-1)
Flow (t-2)	0.01	0.24			0.03	0.63			0.00	0.04			0.02	0.51			Flow (t-2)
Flow (t-3)	0.12	2.85	0.11	3.35	0.10	3.31	0.10	3.30	0.09	2.66	0.08	2.42	0.09	3.08	0.10	3.38	Flow (t-3)
Flow (t-4)	0.01	0.29			0.10	2.38	0.07	2.14	0.14	3.61	0.16	4.84	0.14	3.63	0.13	3.74	Flow (t-4)
Flow (t-5:t-8)	0.04	2.63	0.04	3.18	0.01	0.70			0.00	0.06			0.00	-0.10			Flow (t-5:t-8)
Flow (t-9:t-12)	0.03	2.37	0.02	1.93	0.03	2.96	0.05	5.16	0.04	4.10	0.04	5.67	0.04	3.95	0.04	6.01	Flow (t-9:t-12)
Dec_h1_dummy	0.05	2.68	0.03	1.98	0.20	1.68			-0.10	-0.65			0.05	0.35			Dec_h1_dummy
Dec_h2_dummy	0.02	1.22			-1.88	-8.34	-2.01	-9.87	-1.91	-9.56	-1.87	-11.89	-2.03	-9.32	-2.06	-11.23	Dec_h2_dummy
Jan_h1_dummy	0.02	1.78			0.52	2.55	0.35	2.34	0.23	1.12	0.30	2.20	0.40	1.88	0.34	2.40	Jan_h1_dummy
Jan_h2_dummy	0.02	1.46			0.40	1.65			0.22	1.11	0.29	2.16	0.36	1.57	0.28	1.61	Jan_h2_dummy
Feb_dummy	0.04	3.08	0.03	2.66	0.49	3.36	0.26	4.16	0.12	0.80	0.16	2.33	0.28	1.88	0.25	4.17	Feb_dummy
Apr_dummy	0.00	0.29			-0.22	-1.40	-0.39	-3.54	-0.28	-1.70	-0.24	-2.35	-0.27	-1.57	-0.31	-2.88	Apr_dummy
May_dummy	0.02	1.51			0.17	1.28			-0.01	-0.07			0.07	0.48			May_dummy
Jun_h1_dummy	0.01	0.44			-0.08	-0.42	-0.24	-1.42	-0.16	-0.78			-0.14	-0.69			Jun_h1_dummy
Jun_h2_dummy	0.02	1.53			0.75	4.79	0.62	5.07	0.50	3.41	0.56	6.24	0.67	4.35	0.65	5.73	Jun_h2_dummy
Jul_h1_dummy	0.04	2.18			0.16	0.95			-0.14	-0.87			0.02	0.10			Jul_h1_dummy
Jul_h2_dummy	0.01	1.13			0.09	0.57			-0.07	-0.48			-0.02	-0.15			Jul_h2_dummy
Aug_dummy	0.02	1.63			0.20	1.62			0.03	0.19			0.12	0.94			Aug_dummy
Sep_dummy	-0.01	-0.82	-0.03	-2.03	-0.06	-0.43	-0.24	-2.55	-0.03	-0.16			-0.05	-0.34			Sep_dummy
Oct_dummy	0.01	0.43			0.10	0.73			0.00	0.03			0.05	0.41			Oct_dummy
Nov_dummy	0.03	2.08	0.02	1.14	0.31	2.16			0.02	0.14			0.17	1.18			Nov_dummy
Volatility(t)	-0.15	-0.29			29.21	7.18	28.07	6.84	27.21	6.87	26.50	6.71	30.22	7.65	30.25	7.61	Volatility(t)
Volatility(t-1)	-0.41	-1.07			-20.11	-5.13	-21.58	-6.37	-15.55	-4.08	-16.49	-4.71	-18.90	-4.94	-19.67	-5.42	Volatility(t-1)
Volatility(t-2)	0.64	1.61	0.60	1.64	-2.21	-0.70			-5.07	-1.78	-5.67	-2.13	-4.39	-1.49			Volatility(t-2)
Volatility(t-3)	0.06	0.17			-3.69	-0.56			-4.38	-0.65			-4.33	-0.62	-4.93	-0.71	Volatility(t-3)
Volatility(t-4)	-0.84	-2.35	-0.97	-2.98	-3.23	-0.97			0.67	0.23			-1.78	-0.57			Volatility(t-4)

Table 3.5: Broker Type Regressions of Aggregate Net Flows, Buy Flows, Sell Flows and Total Flows

This table analyses normalised weekly net flows, buy flows, sell flows, and total flows, allowing for different behaviour for internet investors and traditional full-service broker investors. Unnecessary regressors are eliminated using an iterative process that maximises the AIC information criterion. Variable definitions are defined as in Table 3.4. For each flow type, internet results are on the left and full-service on the right. The regression is from Dec 1996 to Dec 2002.

Variable	Net Flows				Buy Flows				Sell Flows				Total Flows				Variable
	Internet		Traditional		Internet		Traditional		Internet		Traditional		Internet		Traditional		
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	
p-value	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000		p-value
adjR2	15.7%		39.8%		43.3%		57.8%		38.5%		49.6%		50.9%		55.9%		adjR2
AIC																	AIC
Constant	0.01	0.80	0.00	-0.15	-0.04	-0.45	-0.05	-0.48	0.07	0.58	-0.06	-0.53	0.03	0.31	-0.08	-0.75	Constant
MktRet (t)	-2.22	-4.34	-1.91	-6.45	-6.06	-2.37	-8.47	-4.53	7.28	2.68	4.94	2.04					MktRet (t)
MktRet (t-1)																	MktRet (t-1)
MktRet (t-2)							4.82	2.07	4.04	1.61	5.02	2.08	3.41	1.72	4.35	1.85	MktRet (t-2)
MktRet (t-3)			0.65	3.32													MktRet (t-3)
MktRet (t-4)																	MktRet (t-4)
MktRet (t-5:t-8)					3.95	3.27			2.94	2.30			1.92	1.62			MktRet (t-5:t-8)
MktRet (t-9:t-12)																	MktRet (t-9:t-12)
Flow (t-1)	0.16	2.33	0.43	10.36	0.33	6.69	0.43	9.83	0.35	7.07	0.37	5.66	0.37	7.72	0.42	7.68	Flow (t-1)
Flow (t-2)									0.10	2.13			0.12	1.87			Flow (t-2)
Flow (t-3)			0.13	2.48			0.11	2.36					0.09	1.65	0.11	2.47	Flow (t-3)
Flow (t-4)	-0.14	-1.32					0.12	2.90			0.10	2.48			0.10	2.64	Flow (t-4)
Flow (t-5:t-8)	0.07	2.18			0.03	2.05											Flow (t-5:t-8)
Flow (t-9:t-12)																	Flow (t-9:t-12)
Dec_h1_dummy									-0.32	-1.46							Dec_h1_dummy
Dec_h2_dummy			-0.03	-2.68	-1.75	-7.60	-2.04	-8.96	-1.71	-6.02	-2.24	-9.28	-1.81	-6.55	-2.24	-9.28	Dec_h2_dummy
Jan_h1_dummy							0.77	3.16					0.67	2.20	0.62	2.33	Jan_h1_dummy
Jan_h2_dummy			0.02	1.35	0.60	2.02	0.68	3.08	0.37	1.12			0.67	1.88	0.46	2.25	Jan_h2_dummy
Feb_dummy	0.04	2.76			0.78	4.86	0.50	5.08			0.32	2.36	0.41	3.21	0.42	4.00	Feb_dummy
Apr_dummy	-0.06	-2.25	-0.02	-1.64													Apr_dummy
May_dummy							0.25	2.52									May_dummy
Jun_h1_dummy			-0.03	-1.92													Jun_h1_dummy
Jun_h2_dummy					0.63	4.11	0.88	6.01	0.62	4.15	0.85	7.60	0.71	5.96	0.89	8.21	Jun_h2_dummy
Jul_h1_dummy	-0.05	-1.59															Jul_h1_dummy
Jul_h2_dummy					0.35	1.61											Jul_h2_dummy
Aug_dummy			0.02	1.53			0.20	2.04									Aug_dummy
Sep_dummy			-0.05	-2.97													Sep_dummy
Oct_dummy	-0.06	-2.35															Oct_dummy
Nov_dummy					0.38	2.85	0.39	3.19					0.27	2.79	0.25	2.27	Nov_dummy
Volatility(t)					18.01	4.45	20.36	3.25	17.16	3.74	26.67	4.86	21.16	4.90	24.83	4.11	Volatility(t)
Volatility(t-1)			0.67	1.84	-21.49	-4.12	-18.69	-4.72	-16.88	-3.39	-22.85	-6.28	-18.90	-3.84	-23.25	-6.43	Volatility(t-1)
Volatility(t-2)	1.94	2.93							-14.80	-3.02			-9.44	-1.60			Volatility(t-2)
Volatility(t-3)																	Volatility(t-3)
Volatility(t-4)			-0.80	-2.94			-6.11	-2.10	9.65	2.54							Volatility(t-4)

The third group of regressors examines calendar effects on flows. Ritter (1988) finds using US data a low buy/sell ratio for individuals in late December and a high ratio in early January due to tax year-end driven trading. The results in Table 2.4 show no economically significant effects for net flows at the turn of the year. There is weak evidence that net flows are positive in early July, corresponding to the start of the new tax year in Australia. A more robust result is that the *level* of trading activity is strongly influenced by calendar effects. Buy, sell, and total flows are significantly lower in the second half of December due the Christmas / New-Year holiday effect. Total trading activity is also significantly higher than usual in the second half of June, coinciding with the end of the tax year (30th June) and in the first two weeks of January.

The final group of regressors examines short-term market volatility. Table 3.4 shows that net flows are unaffected by market volatility while buy, sell, and total flows are all strongly increasing in contemporaneous volatility. This matches the well-known relationship between volume and volatility shown by Karpoff (1987).

Next I examine whether aggregate flows generated by internet broking clients are significantly different from aggregate flows generated by full-service broking clients. As noted by Odean (1998) the actions of discount and full-service brokerage clients may differ. He uses discount clients in his analysis to avoid the need to disentangle the decisions and motivations of individual investors from those of their retail brokers. In Table 3.5 both sets of clients are separated to see if there are any significant differences in their aggregated flows.

Net flows, buy, sell, and total flows are constructed separately for both internet and full-service brokers. The same de-trending and normalisation process is applied as for the full sample results. For brevity only the results for the AIC optimised regressions are presented, rather than the full regression results. Since the internet sample began only part way through the sample period, the period for the two regressions is from December 1996 to December 2002.

The results of this analysis in Table 3.5 show that there is little economic difference in the trading behaviour between clients of internet brokers and full-service brokers at the aggregate market level. Both exhibit a negative relationship between returns and net flows, and both show significant positive persistence in one-week aggregate flows. The two groups have similar seasonalities in buy, sell, and total flows and also show common behaviour relative to intra-week market volatility[54].

In summary, based on the results for *market* level flows, it appears that systematic patterns *do* exist in the aggregated flows of individual investors. Moreover, these patterns are consistent across investors from both full-service and internet brokerage firms. The next section examines cross-sectional flows to see if a similar result holds at the individual stock level.

3.3.2 Cross Sectional Flows

Cross sectional flows are defined as the aggregated flows from individual investors in and out of individual stocks on a weekly basis. Net flows are calculated using the same definition as for market flows, giving a value between –1 and +1 for each stock in each week of the sample. In addition I separately model buy, sell, and total flows, which are logged and divided by market capitalisation to generate standardised measures for each stock. As discussed earlier, only index stocks are analysed.

Weekly aggregate flows are regressed on the following factors:

- contemporaneous and lagged weekly stock returns
- lagged weekly stock flows
- intra-week stock return volatility

[54] An F-test comparing the two groups of coefficients doesn't reject the null that they are equal at the 5% level.

Weekly Fama-Macbeth cross-sectional regressions are used to calculate regression coefficients and Newey-West standard errors are used to adjust for serial dependence in the coefficients caused by the use of overlapping lagged data. In addition contemporaneous and one-week lagged returns are split into positive and negative components to examine any asymmetry in the reaction to positive and negative stock returns.

The results of the regressions using all retail brokers appear in Table 3.6. The first set of regressors show that weekly net flows are strongly *negatively* related to contemporaneous weekly stock returns for both positive and negative returns. Additionally weekly net flows have a significant negative relationship with lagged stock returns out to a lag of two months. The results support the existence of very strong negative feedback (contrarian) trading at the individual stock level by small investors. This confirms the cross-sectional findings of Odean (1998), Grinblatt and Keloharju (2000), and Shapira and Venezia (2000) that individuals tend to be contrarian investors[55,56].

Examining the results for buy and sell flows splitting out positive and negative returns reveal some interesting results. Notably, positive returns lead to an increase in both buying and selling by individuals, however sell flows dominate leading to a net negative flow in response to positive returns. Negative returns by contrast have little impact on selling behaviour, but have a strong effect on buying behaviour, leading to contrarian behaviour. Why is this interesting? Previous research has explained the negative relationship between net flows and returns as a manifestation

---

[55] Dhar & Kumar (2001) find using account level detail that there is some heterogeneity of small investors in their response to past price trends. As a group though, individual investors appear to behave in a contrarian manner.

[56] Barber, Odean, and Zhou (2003) find that over longer horizons (1-3 years) individual investors become positive feedback traders while over shorter horizons (1-6 months) they are contrarian. Repeating my regressions using quarterly net flows rather than weekly net flows provides some support for this finding. In unreported analysis I also find evidence of positive feedback trading over a 12-18 month horizon. However, while the result is statistically significant, the economic magnitude of the long term positive flow-return reaction is significantly weaker than the contrarian trading patterns demonstrated in this section.

of the disposition effect of Shefrin and Statman (1985). The disposition effect implies that investors are reluctant to sell their losers and eager to sell their winners, implying net flows will be negatively related to recent weekly returns. However, the disposition effect operates entirely through *selling* behaviour by individuals. The results in Table 3.6 show that *both* buying and selling behaviour is driving the negative feedback relationship. The relationship between buys and past returns is not predicted by the disposition explanation[57,58].

The second set of regressors in Table 3.6 show that net flows are extremely persistent at the cross-sectional level with significant coefficients extending out to two months. High persistence may be indicative of individuals using correlated information sources that are acted on with a lag by some investors.

The third set of regressors in Table 3.6 shows that there is a *positive* relationship between contemporaneous stock volatility and net, buy, sell, and total flows. The total-flow/volatility relationship is consistent with the results of Gallant et al (1992) and Epps & Epps (1976) who find that volatility and trading volume are positively related. However, more intriguingly, high volatility generates more buys than sells, resulting in a positive *net* flow on average from individual investors. Barber and Odean (2002) also show that small investors are more likely to trade in stocks that have had recent extreme performance, possibly due to attention effects.

---

[57] An alternative behavioural explanation of the negative flow/return relationship may be anchoring (Kahneman and Tversky (1982)). Shiller (1998) suggests that the past stock price may be an important anchor point for individual investors, and suggests that the more ambiguous the value of the commodity, the more important anchoring is likely to be. If investors were using the past price as an anchor, this would immediately explain the negative return-flow relationship, as after a stock price fall investors would perceive the stock as cheap and vice versa.

[58] Again, the primary objective of this chapter is to assess whether there exist systematic patterns that survive aggregatation, not to determine whether they are behaviourally or rationally driven. Negative feedback trading may also be generated by portfolio rebalancing to restore diversification after price movements or by small investors rationally following mean reversion strategies. The important point here is that there are very strong systematic patterns in individual investor flows related to past returns.

Table 3.7 separates individuals into full-service and internet brokerage clients to compare their behaviour over the period December 1996 to December 2002. Table 3.7 reveals that both groups have a negative relationship between contemporaneous stock returns and net flows, however the relationship is significantly *stronger* for internet investors[59]. The economic pattern is still the same though; both groups are strongly contrarian, buy flows strongly increase with contemporaneous negative returns, while sell flows are affected little. Conversely sell flows increase strongly with positive returns while buy flows actually increase slightly.

Both groups exhibit very strong persistence in cross-sectional order flow, with significant coefficients extending out to a two-month lag. Additionally both groups increase trading activity when volatility is high, with increased volatility leading to net buying, especially for the internet investors.

We can conclude from this analysis that there *are* strong systematic patterns that exist in the net flows of individual investors at the cross-sectional level. These patterns survive aggregation over a large number of individuals, and are robust to the type of brokerage firm that the individuals trade through.

---

[59] An F-test rejects, at the 5% level, the null hypothesis that coefficients for the two groups of clients are equal.

Table 3.6: Cross Sectional Flow Regressions (All Brokers): Net, Buy, Sell and Total Flows

This table analyses weekly net flows, buy, sell and total flows. Individual weekly flows are calculated by summing buys and sells for each individual stock across all brokers in the sample each week. Net flows are defined as (buyflow – sellflow)/(buyflow + sellflow). Other flows are divided by weekly market capitalisation and are standardised cross sectionally so may be interpreted as z-scores. Buy, sell and total flows are in log terms. The period of the regression is September 1991 to December 2002. Contemporaneous and one period lagged stock returns are split into two separate variables, for positive and negative weeks eg StockRetPos = 0 if return is negative during that week and equals the return if the return is positive.

Variable	Net Flows		Buy Flows		Sell Flows		Total Flows		Variable
	All Brokers		All Brokers		All Brokers		All Brokers		
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	
Constant	0.00	0.35	-0.09	-18.05	-0.11	-20.23	-0.10	-21.25	Constant
PosStockRet (t)	**-1.39**	**-17.70**	1.43	12.61	**4.03**	**29.32**	3.01	28.20	**PosStockRet (t)**
NegStockRet (t)	**-1.86**	**-18.30**	**-3.23**	**-17.63**	-0.14	-0.99	-2.14	-14.33	**NegStockRet (t)**
PosStockRet (t-1)	-0.63	-9.91	0.16	1.75	1.36	12.51	0.65	7.56	**PosStockRet (t-1)**
NegStockRet (t-1)	-1.11	-12.40	-1.09	-8.76	0.58	4.56	-0.36	-3.29	**NegStockRet (t-1)**
StockRet (t-2)	-0.39	-9.72	-0.30	-4.76	0.27	4.24	-0.08	-1.31	StockRet (t-2)
StockRet (t-3)	-0.25	-7.58	-0.30	-7.73	0.03	0.47	-0.18	-4.42	StockRet (t-3)
StockRet (t-4)	-0.14	-4.72	-0.23	-5.30	-0.11	-2.29	-0.22	-5.17	StockRet (t-4)
StockRet (t-5:t-8)	-0.11	-5.85	-0.21	-7.75	-0.18	-5.06	-0.24	-8.11	StockRet (t-5:t-8)
StockRet (t-9:t-12)	0.00	-0.26	-0.09	-3.64	-0.27	-8.28	-0.19	-6.47	StockRet (t-9:t-12)
Flow (t-1)	0.26	33.73	0.32	29.82	0.26	26.64	0.31	25.68	Flow (t-1)
Flow (t-2)	0.10	33.15	0.14	36.22	0.12	34.21	0.13	39.94	Flow (t-2)
Flow (t-3)	0.08	22.65	0.09	26.34	0.10	28.88	0.10	27.53	Flow (t-3)
Flow (t-4)	0.06	16.48	0.07	20.85	0.08	19.17	0.08	20.24	Flow (t-4)
Flow (t-5:t-8)	0.03	21.08	0.04	20.72	0.05	25.72	0.04	20.94	Flow (t-5:t-8)
Flow (t-9:t-12)	0.03	22.98	0.03	23.00	0.04	25.89	0.03	20.72	Flow (t-9:t-12)
Stock_Volatility(t)	0.71	6.16	6.38	24.70	5.84	25.51	6.55	28.92	Stock_Volatility(t)
Stock_Volatility(t-1)	-0.16	-1.32	-2.15	-11.40	-1.65	-8.06	-2.23	-11.34	Stock_Volatility(t-1)
Stock_Volatility(t-2)	-0.21	-3.26	-1.58	-10.98	-1.06	-7.63	-1.56	-11.31	Stock_Volatility(t-2)
Stock_Volatility(t-3)	-0.42	-3.84	-1.18	-7.43	-0.51	-3.07	-0.99	-5.96	Stock_Volatility(t-3)
Stock_Volatility(t-4)	-0.47	-4.93	-1.58	-10.76	-0.94	-5.95	-1.51	-10.76	Stock_Volatility(t-4)

Table 3.7: Cross Sectional Flow Regressions (Internet vs Traditional): Net, Buy, Sell and Total Flows

This table analyses weekly net flows, buy flows and sell flows separately for internet and traditional full-service brokerage clients. Individual weekly flows are calculated by summing buys and sells for each individual stock across all brokers in the sample each week. Net flows are defined using raw flows as (buyflow – sellflow)/(buyflow + sellflow). Buy, sell and total flows are divided by weekly market capitalisation and are standardised cross sectionally so may be interpreted as z-scores. The period of the regression is December 1996 to December 2002.

Variable	Net Flows				Buy Flows				Sell Flows				Total Flows			
	Internet		Traditional		Internet		Traditional		Internet		Traditional		Internet		Traditional	
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*
Constant	0.01	1.66	0.00	-0.02	-0.11	-17.29	-0.10	-15.30	-0.12	-14.41	-0.11	-14.72	-0.11	-16.92	-0.10	-16.23
PosStockRet (t)	**-2.61**	**-18.61**	**-0.93**	**-9.18**	0.73	4.63	1.89	10.84	**4.79**	**23.50**	**3.80**	**21.24**	2.95	18.63	3.12	19.98
NegStockRet (t)	**-3.27**	**-14.84**	**-1.37**	**-9.10**	**-4.69**	**-19.12**	**-3.21**	**-21.00**	0.49	2.23	-0.81	-3.52	-3.01	-17.17	-2.42	-15.61
PosStockRet (t-1)	-0.87	-5.57	-0.72	-9.56	-0.02	-0.14	-0.06	-0.68	1.39	8.69	1.38	9.83	0.37	3.65	0.57	6.88
NegStockRet (t-1)	-1.55	-7.39	-0.98	-10.21	-1.57	-8.10	-1.17	-8.00	0.24	1.51	0.28	2.10	-0.89	-6.51	-0.50	-3.80
StockRet (t-2)	-0.44	-6.14	-0.41	-9.25	-0.42	-5.30	-0.48	-6.94	-0.05	-0.76	0.18	2.57	-0.40	-6.30	-0.22	-3.53
StockRet (t-3)	-0.13	-2.25	-0.27	-5.83	-0.20	-3.25	-0.41	-7.10	-0.31	-3.80	-0.03	-0.63	-0.27	-4.31	-0.28	-5.99
StockRet (t-4)	-0.01	-0.18	-0.15	-4.38	-0.04	-0.82	-0.22	-3.57	-0.35	-5.62	-0.07	-1.26	-0.14	-2.89	-0.17	-2.63
StockRet (t-5:t-8)	-0.04	-1.09	-0.09	-4.40	-0.06	-2.22	-0.18	-5.12	-0.31	-6.75	-0.15	-3.53	-0.16	-5.44	-0.19	-4.87
StockRet (t-9:t-12)	0.08	2.33	0.01	0.43	0.04	1.91	-0.05	-1.75	-0.28	-5.78	-0.21	-4.87	-0.07	-2.82	-0.13	-3.84
Flow (t-1)	0.23	21.25	0.26	49.95	0.34	23.39	0.33	53.27	0.24	19.92	0.28	32.71	0.33	21.75	0.34	39.57
Flow (t-2)	0.08	16.86	0.11	33.83	0.13	25.77	0.15	30.40	0.12	21.54	0.12	22.46	0.14	34.03	0.14	34.54
Flow (t-3)	0.08	13.90	0.08	18.08	0.10	18.86	0.09	20.99	0.10	20.31	0.10	20.98	0.11	20.89	0.10	23.72
Flow (t-4)	0.05	9.32	0.06	11.39	0.07	11.70	0.07	14.48	0.08	20.19	0.08	14.57	0.07	16.00	0.08	17.73
Flow (t-5:t-8)	0.03	18.62	0.03	16.44	0.04	18.60	0.04	15.56	0.05	23.02	0.04	17.87	0.04	17.30	0.04	19.10
Flow (t-9:t-12)	0.02	13.00	0.03	17.49	0.03	13.18	0.03	16.46	0.04	17.12	0.04	21.16	0.03	12.90	0.03	17.63
Stock_Volatility(t)	1.28	5.62	0.42	3.18	6.52	17.25	6.20	19.18	5.38	12.28	5.74	15.98	6.27	15.52	6.38	18.22
Stock_Volatility(t-1)	-0.14	-0.66	0.03	0.20	-2.18	-9.68	-2.00	-10.55	-1.55	-7.54	-1.85	-8.45	-2.30	-9.58	-2.40	-11.92
Stock_Volatility(t-2)	-0.31	-2.78	-0.20	-1.82	-1.70	-11.09	-1.65	-7.28	-0.79	-3.94	-1.08	-5.01	-1.53	-9.08	-1.49	-6.81
Stock_Volatility(t-3)	-0.36	-2.51	-0.45	-4.18	-1.42	-8.22	-1.32	-7.74	-0.55	-2.74	-0.93	-4.42	-1.40	-8.33	-1.35	-8.03
Stock_Volatility(t-4)	-0.58	-3.83	-0.32	-2.54	-1.65	-9.96	-1.56	-6.36	-0.84	-3.96	-1.21	-5.15	-1.43	-9.22	-1.47	-7.28

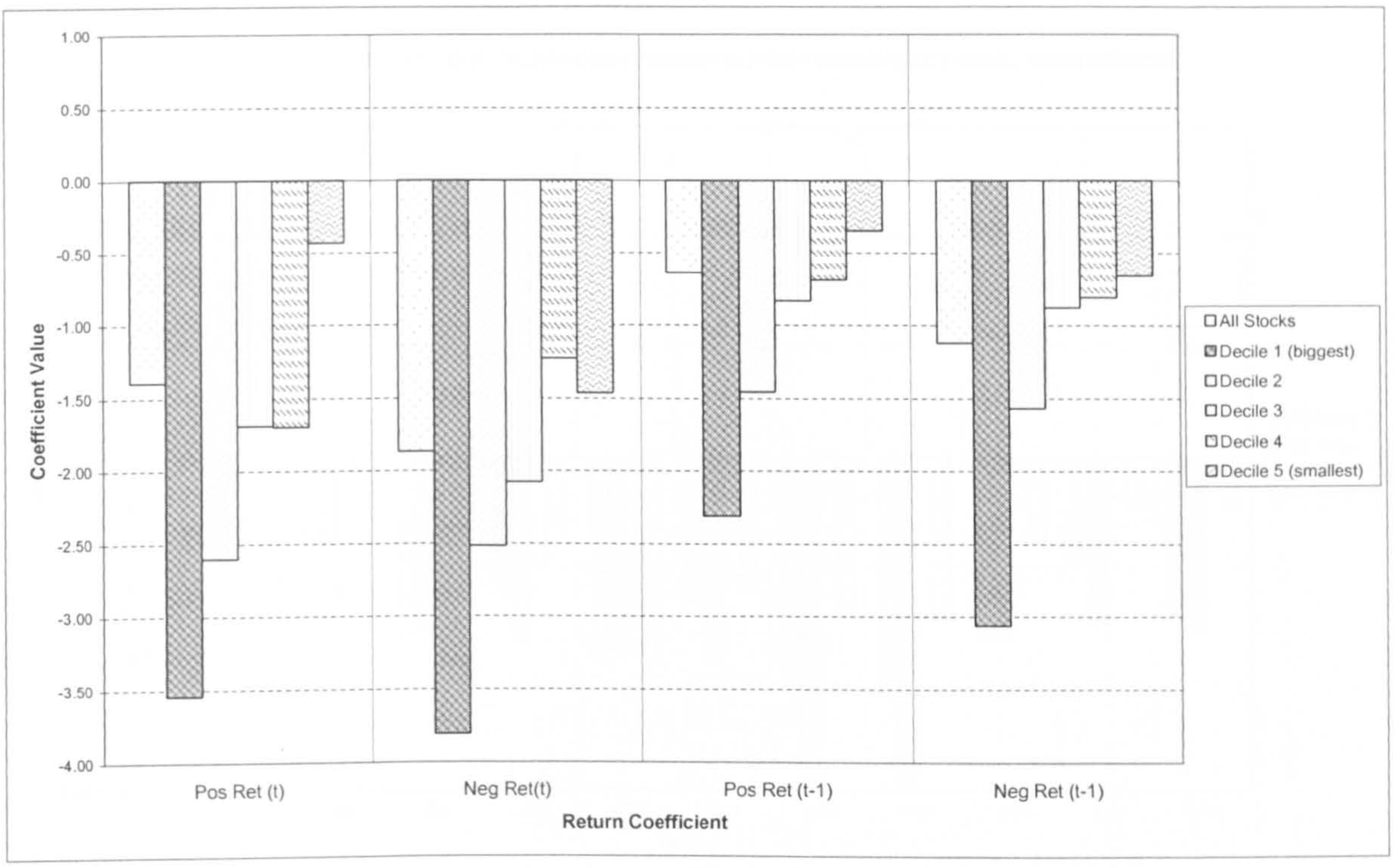
To ensure these patterns are robust under alternative empirical specifications I re-estimate Table 3.6 and Table 3.7 using the *number* of buys and sells in place of the value of buys and sells. The results are extremely similar, revealing strong contrarian patterns, strong persistence in the number of buys and sells, and net buying in response to increased volatility.

Next I divide the stocks into five size quintiles and estimate the regressions for Table 3.6 separately for each size group. Flows remain highly persistent across the five size groups. The positive relationship between volatility and net flows is strongest for the largest stocks, and becomes progressively weaker for smaller stocks. The negative feedback trading result is robust across each size quintile. Figure 3.3 plots the coefficients for contemporaneous and one week lagged positive and negative flows for each size quintile. We can see that the negative feedback results are actually stronger for *larger* stocks[60].

As a third robustness check I re-estimate Table 3.6 individually for 49 of the 56 individual retail brokerage firms (those with sufficiently long panels). Of these 49 there are 40 full-service brokers and 9 internet brokers. This analysis finds significant negative coefficients on PosReturn(t) and NegReturn(t) for 47 of the 49 brokers. Significant positive persistence in flows is also present for 47 of the 49 brokers. The volatility result appears to be slightly less robust, with a significant coefficient on contemporaneous volatility found for only 9 of the 49 brokers (5 of these 9 were internet brokers).

[60] In related research Ranguelova (2001) finds in the US that the disposition effect is strongest for larger stocks, while a reverse disposition effect holds for small stocks. My data confirms that negative feedback trading is stronger for larger stocks but I find that feedback effects have the same sign for all size quintiles.

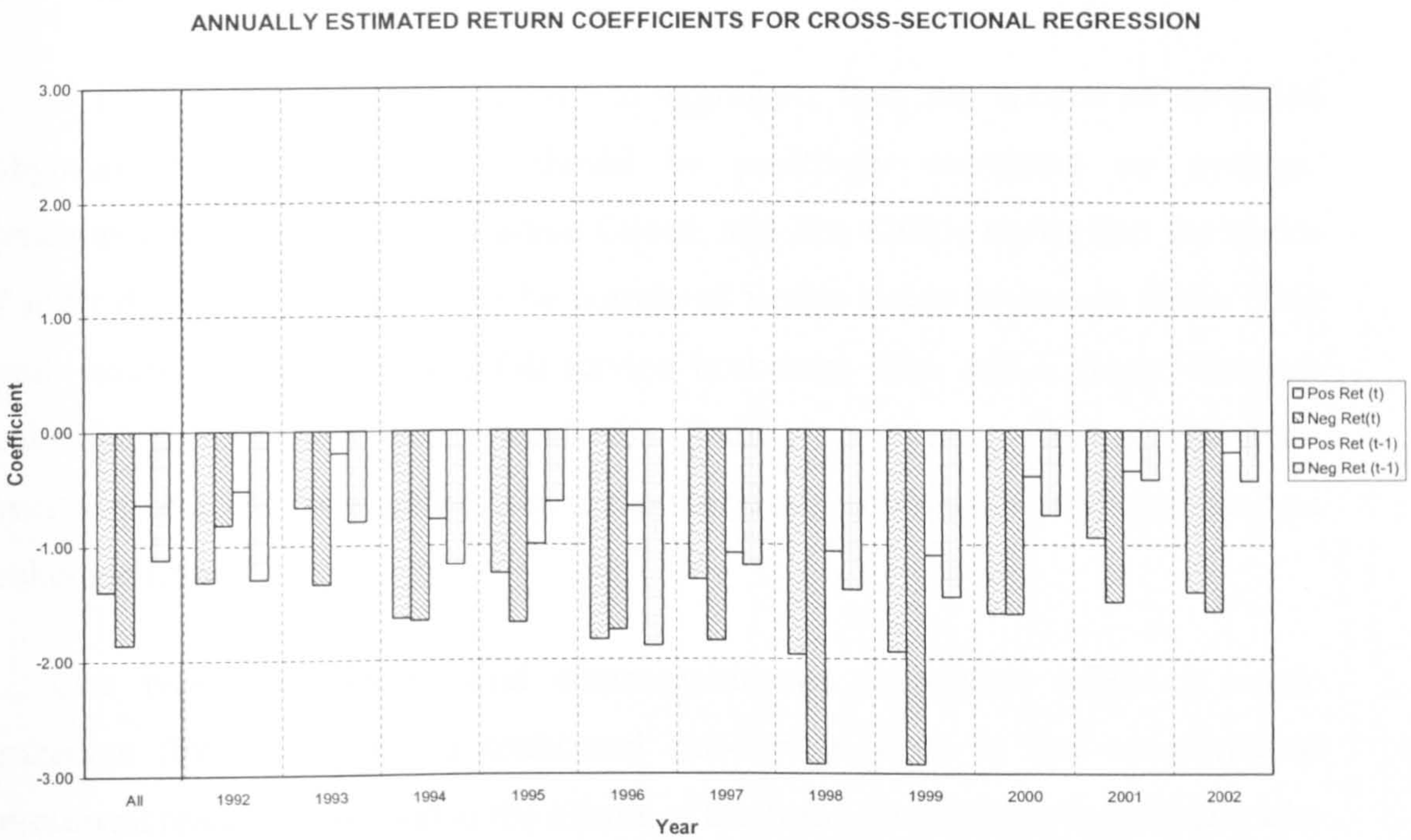
Figure 3.3: Return Coefficients Estimated For Different Size quintiles



This figure shows the coefficients for the four asymmetric return regressors corresponding to the regression in Table 3.6. The coefficients are estimated firstly on all stocks (leftmost bar in each group of six), then on quintiles of stocks starting with the largest stock group second from the left, and moving to the smallest quintile of stocks (on the rightmost of each group of six). This figure demonstrates that the regression coefficients are of consistent sign for both large and small stocks, and are stronger for larger stocks.

As a final robustness check, to show the negative feedback trading pattern is consistent over time, I break the sample into yearly blocks from 1992 to 2002 and re-estimate Table 3.6 separately for each year. Figure 3.4 plots the estimated coefficients for the contemporaneous and one period lagged return variables. This shows that in *every* single year in the sub-sample there is a negative relationship between net flows and returns at the cross-sectional level. In addition, the t-statistics on these coefficients are large ranging from –1.93 (PosReturn(t-1) in 1992) as the least significant to –17.95 as the most significant (NegReturn(t) in 1999). Moreover in 9 of the 11 subperiods, the negative contemporaneous return coefficients are higher than the positive return coefficients. These results confirm our earlier finding that in aggregate individual investors are consistently *contrarian* in their investment decisions, and react more strongly to negative than to positive returns.

Figure 3.4: Return Coefficients for Different Sub-periods



This figure shows the return coefficients for the four asymmetric regressors corresponding to the regression in Table 3.6. The regressions are estimated separately in each year and the coefficients for the PosRet(t), NegRet(t), PosRet(t-1) and PosRet(t-2) are reported above. This shows that in every subperiod these coefficients are significantly negative.

To summarise, the results in Section 3.3 demonstrate that there *are* systematic patterns in individual investor trades that survive after aggregation over a large number of small investors. These patterns are remarkably consistent over an 11-year time period and are somewhat stronger for larger stocks. In the next section we examine the relationship between small investor flows at a more disaggregated level to assess the similarity of small investor behaviour across a wide range of independent brokerage firms.

3.4. Correlations Between Flows from Different Retail Brokerage Firms

If individual investors trades are to aggregate, then the actions of unrelated subgroups of small investors should be positively correlated on average. Contemporaneous research by Barber, Odean, and Zhu (2003) shows that the trades of individual investors appear to be correlated *within* single brokerage firms. This result holds for both a single full-service brokerage firm and a single discount brokerage firm. This section extends this finding to examine whether individual investor trades are systematically contemporaneously correlated *across* independent brokerage firms.

One reason we might find contemporaneous correlation *within* a single brokerage firm (especially a traditional full-service firm) is due to correlated investment research provided to the clients of the firm. Alternatively individuals who live in a certain region may use the same brokerage firm, leading to correlated liquidity shocks due to regional economic factors. However if we can show that individual investor transactions are correlated *across* a large number of independent brokerage firms, this would significantly strengthen the result that individual investor trades contain a strong systematic component.

To test this result an expanded sample of brokers is used including 16 institutional brokers, 15 mixed brokers (institutional and retail), 47 full-service brokers, and 9 internet brokers, giving 87 brokers in total. If individual investors have strong common systematic patterns in their trading activity, we should see positive pairwise correlations between the 56 full-service and internet brokers. We should also see positive but weaker correlations for these 56 brokers with the 15 mixed brokers (who deal with some small investors) and zero (or negative) correlations with the institutional brokers.

To test for contemporaneous correlations, I first calculate the cross-sectional net flows each week for each broker for each index stock. Next, for each of the 3741

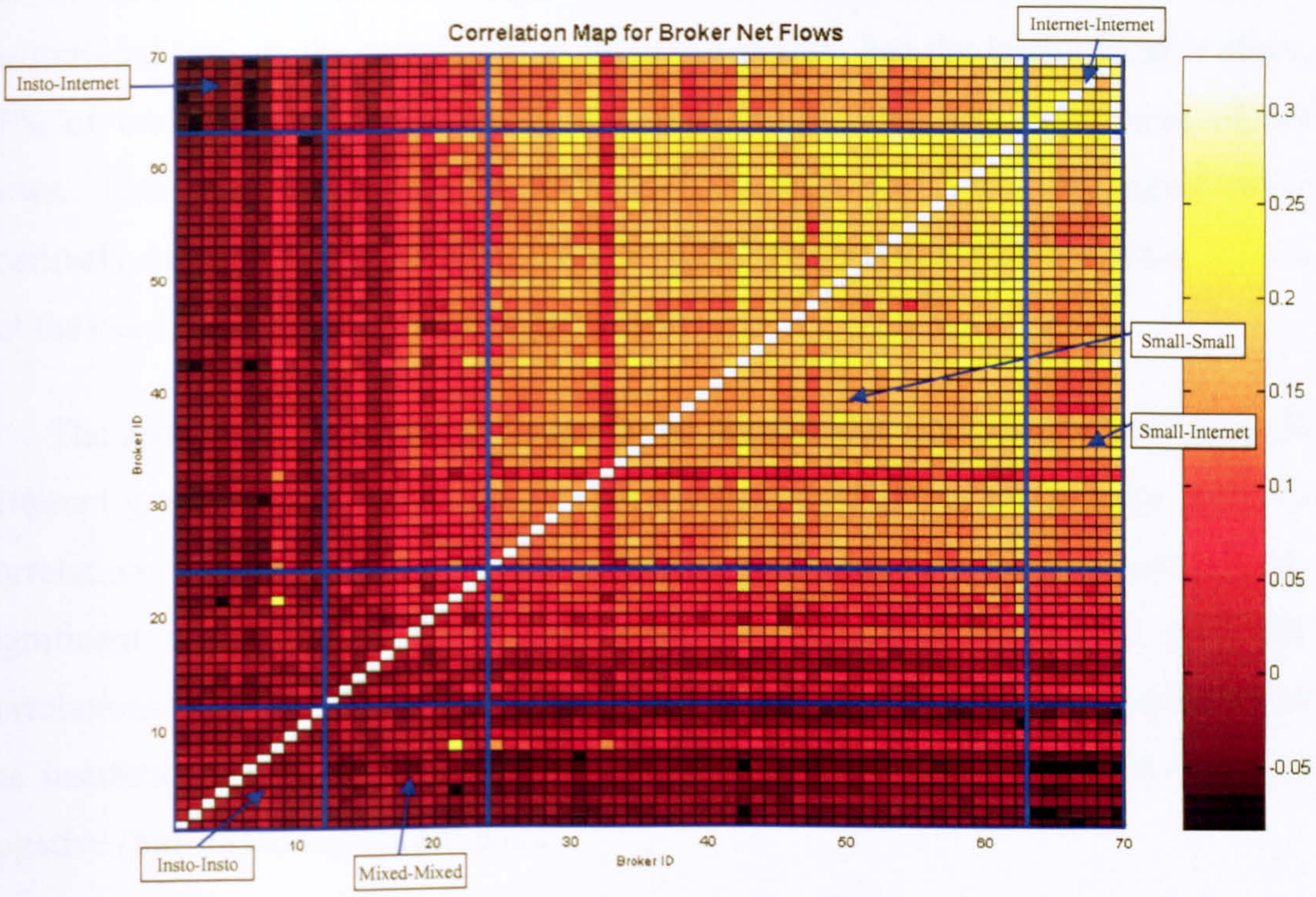
broker pairs (87x86/2) I calculate the contemporaneous correlation between the net flows of the brokers for each weekly cross-section of between 250 and 350 stocks. Then I take the time series average of this correlation across the full sample and estimate the standard deviation of this series (allowing for autocorrelation in this series induced by persistent flows over time using the Newey West approach). This gives me a point estimate for the correlation and a corresponding t-statistic relative to the null of zero correlation for each individual broker pair. Broker pairs where there are less than 12 overlapping weeks are omitted from the analysis (266 broker pairs or 7.1% of all pairs are omitted). Such excluded pairs arise because some brokers (eg internet brokers) enter partway through the sample, while some brokers exit the sample during the period.

Figure 3.5 presents a correlation heat map between the net flows of 70 of the 87 brokers[61]. The brokers are sorted into four blocks; 11 institutional, 12 mixed, 39 full-service small brokers, and 8 internet small brokers. To further highlight patterns in the data, brokers within each block are sorted in descending order by average trade size for that broker. This implies brokers with smaller clients will be on the right hand side of each block.

Figure 3.5 reveals some interesting results. Looking at the top right corner, we see that the correlation between the 8 internet brokers is high, as is the correlation between the 39 full-service brokers. For the full-service brokers, the correlations rise as the average trade size of the brokerage firm falls, indicating that full-service brokers with smaller clients tend to have more highly correlated net flows. Correlations between pairs of institutional brokers are slightly negative, and correlations between institutional and retail brokerage firms are negative.

[61] 17 brokers with a large number of missing pairs are omitted for the figure to improve visualisation of the correlation structure. These brokers are not excluded from the tables.

Figure 3.5: Correlation map for broker net flows – all stocks



This figure shows the average weekly cross-sectional correlation between the net flows from 70 separate brokerage firms for all index stocks. Brokers are split into 11 institutional brokers (ID: 1-11), 12 mixed brokers (ID: 12-23), 39 full-service small brokers (ID: 24-62) and 8 internet brokers (ID: 63-70). Within each block, brokers are sorted in descending order of average trade size, so brokers with smaller clients will appear to the right of each block.

The period of the analysis is September 1991 to December 2002. Dark colours represent areas of low or negative correlation; lighter colours indicate higher correlations as shown in the colourbar on the right of the plot. The labels indicate the type of correlation in each sub-block eg 'Small-Internet' refers to the correlation of the 9 internet brokers with the 39 full-service small brokers.

Average summary correlations for each group are presented in Table 3.8. The first four rows of the table report within-group correlations. We can see that the average correlation of weekly cross-sectional net flows for internet broker pairs is 0.229, which is significantly positive. Additionally and remarkably for the 28 unique internet broker pairs, *every* single pair has a positive correlation between cross-sectional net flows that is statistically significant. For full-service brokers, the average correlation is 0.159, and again *every* single one of the 741 unique broker

pairs has a positive correlation between net flows that is statistically significant. For the 12 mixed brokers, the average correlation is close to zero, and for the 11 institutional brokers the correlation is slightly negative. For the institutional brokers, 87% of broker-pairs have a statistically significantly negative correlation of net flows. These results show that individual investor net flows are significantly positively correlated across a large number of different retail brokerage firms. This is not the case for institutional broker net flows.

The second half of Table 3.8 reports the correlations between pairs of firms in different groups. For the 312 full-service/internet broker pairs the average pairwise correlation is 0.156. Again 100% of these pairs have a positive statistically significant correlation. For the internet/mixed and full-service/mixed pairs the correlations are 0.059 and 0.060 respectively. Finally for broker pairs matched with the institutional brokers, the correlations are negative, with a large percentage of negative (but small) significant pairs.

Table 3.8: Broker-Pair Correlation Analysis

This table shows the average pairwise correlation between brokers for four different broker types. Brokers are divided into 8 Internet brokers, 39 full-service retail brokers, 12 mixed brokers, and 11 institutional brokers. The first row of the table is the average pairwise correlation and other statistics for the 8x7/2 = 28 internet broker pairs, the fifth row shows the statistics for the 39x8 = 312 full-service / internet broker pairs. The second column shows the correlation across the pairs, the third column the percentage of pairwise correlations that are greater than zero. The percentage of t-statistics that are greater than or less than 2 are given in the next two columns, and the number of broker pairs used in the sample are given in the final column. The sample period for the analysis is September 1991 to December 2002. To calculate the correlations for a single broker pair, the correlation between net flows for each week is taken for the cross-section of all index stocks, this is repeated for each week, then the time series of the correlation is taken to get a point estimate and calculate t-statistics.

Broker Type 1	Broker Type 2	Average Pairwise Correlation	Stdev Pairwise Correlation	Percentage of Positive Correlation Pairs	Percentage of pairs with t-stats > 2	Percentage of pairs with t-stats <- 2	Number of BrokerPairs
Internet	Internet	0.229	0.045	100%	100%	0%	28
Full-Service	Full-Service	0.159	0.040	100%	100%	0%	741
Mixed	Mixed	0.021	0.033	67%	32%	68%	66
Insto	Insto	-0.003	0.013	36%	13%	87%	55
Internet	Full-Service	0.156	0.042	100%	100%	0%	312
Internet	Mixed	0.059	0.038	93%	85%	15%	96
Internet	Insto	-0.023	0.044	27%	23%	77%	88
Full-Service	Mixed	0.060	0.043	90%	77%	23%	468
Full-Service	Insto	-0.008	0.032	31%	24%	76%	429
Mixed	Insto	-0.004	0.031	36%	15%	85%	132

Table 3.9: Broker-Pair Correlation Analysis: Top 50 stocks only

This table shows the average pairwise correlation between brokers for four different broker types for the Top 50 stocks only. Brokers are divided into 8 Internet brokers, 39 full-service retail brokers, 12 mixed brokers, and 11 institutional brokers. The first row of the table is the average pairwise correlation and other statistics for the 8x7/2 = 28 internet broker pairs, the fifth row shows the statistics for the 39x8 = 312 full-service / internet broker pairs. The second column shows the correlation across the pairs, the third column the percentage of pairwise correlations that are greater than zero. The percentage of t-statistics that are greater than or less than 2 are given in the next two columns, and the number of broker pairs used in the sample are given in the final column. The sample period for the analysis is September 1991 to December 2002. To calculate the correlations for a single broker pair, the correlation between net flows for each week is taken for the cross-section of all index stocks, this is repeated for each week, then the time series of the correlation is taken to get a point estimate and calculate t-statistics.

Broker Type 1	Broker Type 2	Average Pairwise Correlation	Stdev Pairwise Correlation	Percentage of Positive Correlation Pairs	Percentage of pairs with t-stats > 2	Percentage of pairs with t-stats <- 2	Number of BrokerPairs
Internet	Internet	0.439	0.073	100%	100%	0%	28
Full-Service	Full-Service	0.235	0.056	100%	100%	0%	741
Mixed	Mixed	0.032	0.046	68%	38%	62%	66
Insto	Insto	-0.011	0.016	24%	4%	96%	55
Internet	Full-Service	0.270	0.062	100%	100%	0%	312
Internet	Mixed	0.125	0.080	100%	92%	8%	96
Internet	Insto	-0.024	0.058	30%	19%	81%	88
Full-Service	Mixed	0.091	0.066	98%	82%	18%	468
Full-Service	Insto	-0.011	0.041	25%	19%	81%	429
Mixed	Insto	-0.011	0.030	35%	11%	89%	132

Since net flows for a given stock each week for the entire market must sum to zero by definition, if we find positive correlations between one group of investors' net flows, they must be negatively related to some other group to satisfy this adding up constraint. It appears that the net flows from small investors are significantly positively correlated, while institutional brokerage firms are negatively correlated with small investors' net trades. Interestingly the correlations of net flows for different institutional brokerage firms are also negatively correlated with each other, so systematic common patterns in the trades of their clients are significantly less prevalent (or are internalised within the broker via the matching process).

To test whether the overall correlation results are robust, I repeat the analysis firstly using only the largest 50 index stocks in each week of the sample[62]. The correlation heat map for this test appears in Figure 3.6. We can see that the correlations are significantly *higher* if we only consider the top 50 stocks. This arises because brokers are more likely to trade in these stocks each week, while for small stocks there are quite a few zero net flow weeks that tend to reduce the correlations. The results for the top 50 stocks appear in Table 3.9. The average correlation between internet brokerage pairs rises to 0.439, between full-service pairs to 0.270, and between internet/full-service pairs to 0.235. In a second robustness check I re-estimate the pairwise correlations for all brokers and all stocks excluding a subset of widely held large stocks that were privatised or demutualised over the sample period. These 18 stocks[63] were initially allocated disproportionately to small investors who consistently sold down their large initial holdings over the post-listing period, which

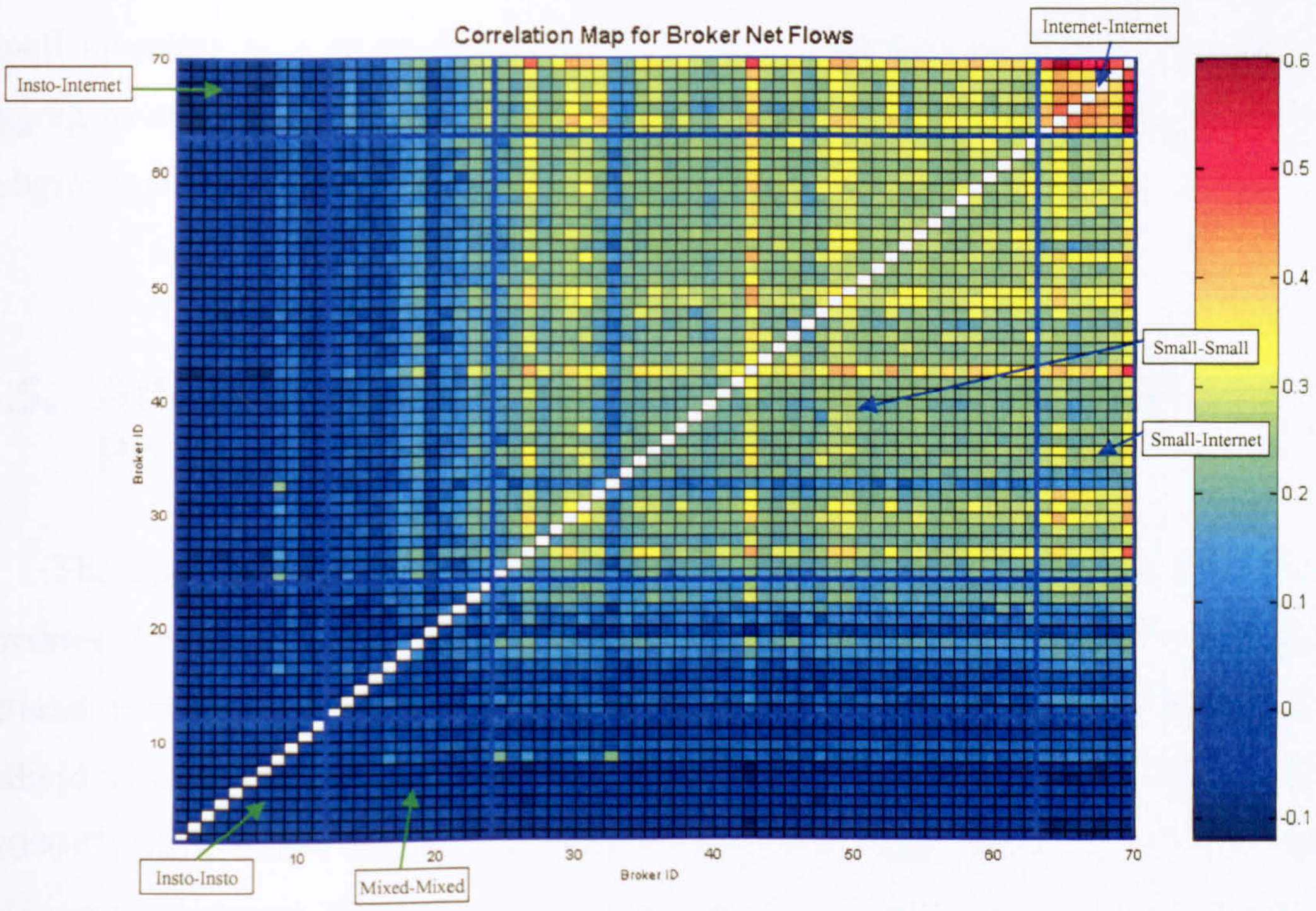
---

[62] In additional analysis I calculate the correlation of *aggregate market* flows between the 87 brokers. I find that the average correlation of aggregate net flows between internet brokers is 0.530, between full-service brokers is 0.153. These are statistically significantly different from zero at the 1% level. This implies that in addition to the correlation of cross-sectional flows, aggregate flows are also highly correlated across different firms.

[63] The specific stocks excluded were Commonwealth Bank (CBA), St George (SGB), Adelaide Bank (ADB), Commonwealth Serum Laboratories (CSL), Tabcorp (TAH), Qantas (QAN), Bankwest (BWA), AMP (AMP), NSW Tab (TAB), Telstra (TLS), Australian Stock Exchange (ASX), Queensland Tab (UTB), AXA (AXA), NRMA (IAG), Colonial (CGH), Suncorp (SUN), SGIO (SGI) and GIO (GIO). These stocks were identified from the RBA report "Privatisation in Australia" (1997) and from industry reports.

might lead to correlated trading. The results of this test reveal however that removing these stocks has little impact on the results. The average correlation between internet pairs is 0.217, between full-service pairs is 0.147, and between internet/full-service pairs is 0.143.

Figure 3.6: Correlation map for broker net flows – top 50 stocks



This figure shows the average weekly cross-sectional correlation between the net flows from 70 separate brokerage firms for the top 50 index stocks only. Brokers are split into 11 institutional brokers (ID: 1-11), 12 mixed brokers (ID: 12-23), 39 full-service small brokers (ID: 24-62) and 8 internet brokers (ID: 63-70). Within each block, brokers are sorted in descending order of average trade size, so brokers with smaller clients will appear to the right of each block.

The period of the analysis is September 1991 to December 2002. Dark colours represent areas of low or negative correlation; lighter colours indicate higher correlations as shown in the colourbar on the right of the plot. The labels indicate the type of correlation in each sub-block eg 'Small-Internet' refers to the correlation of the 9 internet brokers with the 39 full-service small brokers.

These results confirm that the actions of individual investors are strongly systematic *across* a wide range of different brokerage firms. The economic source of

this correlation is unlikely to be individual brokerage firm advice, since the investors are coming from 56 separate brokerage firms. The results from section 3.3 suggest that the correlation in behaviour is more likely driven by a common reaction to recent stock returns and volatility.

The results of these first two sections empirically confirm that the aggregation assumption used in many behavioural finance models may well be reasonable for small investors as a group. There appear to be strong systematic patterns in their aggregate trading behaviour, and trades are strongly correlated across different subgroups of small investors.

3.5. Affect of Aggregate Systematic Patterns on Prices

The final section of this chapter examines the relationship between individual investor flows and *future* stock returns to assess whether flows can predict future returns. Previous research on this issue is mixed with some researchers finding that individual investors underperform on average (eg Odean (1998), Barber and Odean (2000)[64]) while others find they outperform (eg Choe, Kho, and Stulz (2000)). The relationship between flows and future performance is important as it allows us to assess whether individual investors are acting in an irrationally systematic, or conversely a rationally systematic fashion.

To evaluate the relationship, I treat returns and net flows as jointly endogenous and estimate a vector autoregression (VAR) system at both the aggregate market and cross-sectional levels.

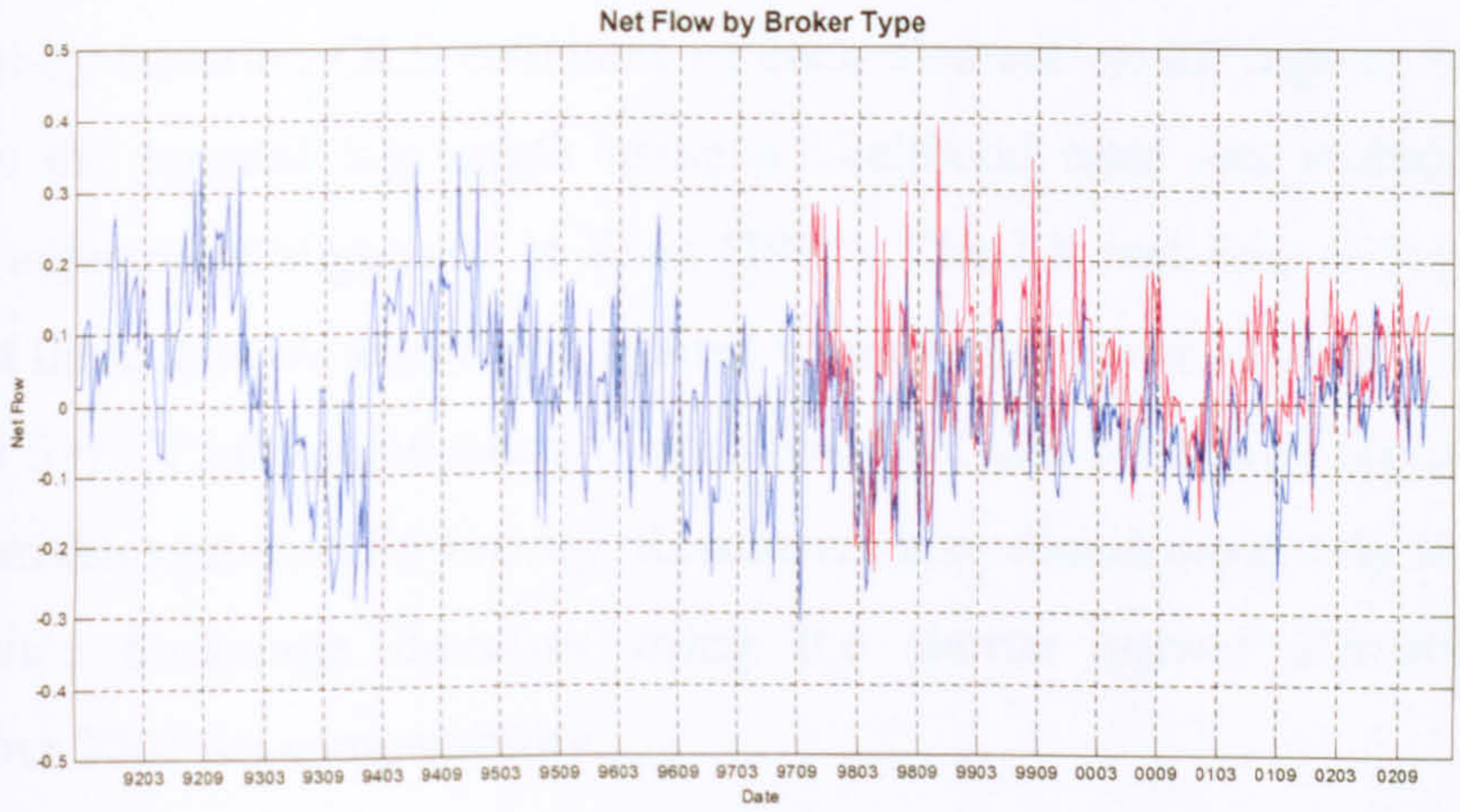
---

[64] Barber and Odean (2000) provide convincing evidence that small investors' *post-cost* investment performance is generally negative relative to buy and hold strategies. The underperformance appears to be increasing in the level of trading intensity. My analysis focuses entirely on pre-cost performance given the limitations of my data sample.

3.5.1 Market Level Flows

In this section I examine the relationship between aggregate market returns and the aggregated market level flows of individual investors. In addition I repeat the analysis separately for full-service and internet brokers. Figure 3.7 plots the market level aggregated net flows for internet and full-service brokers over time. Figure 3.8 plots the market accumulation index over the sample period.

Figure 3.7: Market Level net flows for full-service and internet brokers



This figure shows aggregate market level net flows for internet and full-service brokers over the sample period from September 1991 to December 2002.

Figure 3.8: Market Accumulation Index (Rebased to 100 at start of sample)

This figure shows the market accumulation index over the sample period from September 1991 to December 2002.

To analyse the relationship I estimate the following reduced form VAR system with net flows and aggregate market returns as the dependent variables.

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \Phi_3 z_{t-3} + \Phi_X x_t + \varepsilon_t$$

$$\text{where: } z_t = [f_t, r_t]^T,\ E(\varepsilon_t \varepsilon_t') = \Omega = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

A vector of constants Φ_0 is included, along with a vector of exogenous variables x_t (lagged volatility and calendar dummies). Hamilton (1994) shows that the maximum likelihood estimates of the VAR parameters Φ_i can be found through equation-by-equation OLS estimates of each element on all lags in the system[65]. I estimate the optimal lag length using a likelihood ratio test including the small sample adjustment suggested in Sims (1980). The LR test fails to reject at 5% the null that three lags are significant against various alternative choices[66]. The estimated reduced form VAR specification is presented in Table 3.10, with the coefficients on the x-variables omitted for brevity. Results are also shown separately for internet and full-service brokerage investors using the shorter period December 1996 to December 2002 for comparability.

The results presented in Table 3.10 can be interpreted as a forecasting equation and are quite striking. Aggregate market returns are positively related to the previous week net flows of small investors, and are negatively related to net flows from the week preceding this. The Granger causality test indicates that flows significantly granger cause returns with a p-value of 0.02. The effect is economically significant, with a 2 standard deviation positive shock to small investor net flows (increasing the net flow ratio by 0.22) increasing the expected market return in the following week by 47bp[67]. From the VAR we can also see that weekly net flows are highly persistent.

---

[65] The MLE estimate of the VAR error covariance matrix is also obtained by using the residuals from the OLS estimation.

[66] The AIC is also minimised at a choice of three lags.

[67] Including different numbers of lags in the VAR does not alter this result significantly. For example 4 lags gives an equivalent result of +50bp, while 2 lags gives +52bp.

Table 3.10 – VAR for market returns and flows

This table reports parameter estimates for a VAR containing weekly aggregate market returns (in %) and aggregate weekly raw flows. The process is estimated for all brokers, and separately for internet and full-service brokers. The parameters in the table correspond to the reduced VAR system described in section 3.5.1. The bottom section of the table contains the results of the Granger causality test with the associated F-statistics and p-values.

Variable	All Investors (1991-2002)				Internet Investors (1996-2002)				Full-Service Investors (1996-2002)			
	Flow (t)	*tstat*	*Return (t)*	*tstat*	*Flow (t)*	*tstat*	*Return (t)*	*tstat*	*Flow (t)*	*tstat*	*Return (t)*	*tstat*
Constant	-0.002	-0.18	-0.0031	-1.12	-0.001	-0.02	-0.0028	-0.73	-0.033	-1.74	-0.0021	-0.55
Flow (t-1)	0.451	8.58	**0.0215**	**2.66**	0.221	3.09	**0.0096**	**1.42**	0.380	6.58	**0.0329**	**2.86**
Flow (t-2)	0.129	2.26	**-0.0235**	**-2.23**	-0.045	-0.62	**-0.0054**	**-0.64**	0.156	2.09	**-0.0489**	**-3.13**
Flow (t-3)	0.155	3.11	0.0097	1.17	-0.019	-0.26	0.0073	0.92	0.078	1.09	0.0243	2.01
StockRet (t-1)	-0.584	-1.88	0.0412	0.89	-0.290	-0.65	-0.0228	-0.36	-0.298	-0.81	0.0574	0.80
StockRet (t-2)	-0.023	-0.08	0.0001	0.00	-0.378	-0.57	-0.0068	-0.12	0.351	0.93	-0.0754	-1.36
StockRet (t-3)	0.254	1.04	0.0549	1.19	-0.283	-0.47	0.0305	0.50	0.667	2.49	0.0160	0.26

Variable	All Investors				Internet Investors				Traditional Investors			
	Flow (t)	*p value*	*Return (t)*	*p value*	*Flow (t)*	*tstat*	*Return (t)*	*p value*	*Flow (t)*	*p value*	*Return (t)*	*p value*
Granger F-Stat Flow (t-1)	117.78	0.000	6.26	0.023	5.67	0.001	0.92	0.430	31.97	0.000	5.01	0.002
Granger F-Stat Return (t-1)	1.907	0.127	0.809	0.489	0.712	0.545	0.111	0.954	2.386	0.069	1.046	0.373

Splitting the VAR into full-service and internet brokerage clients, we see from Table 3.10 that the positive relationship between lagged flows and future returns is predominantly driven by full-service brokerage flows. Additionally the persistence of net flows for these brokers is significantly higher than for internet brokers.

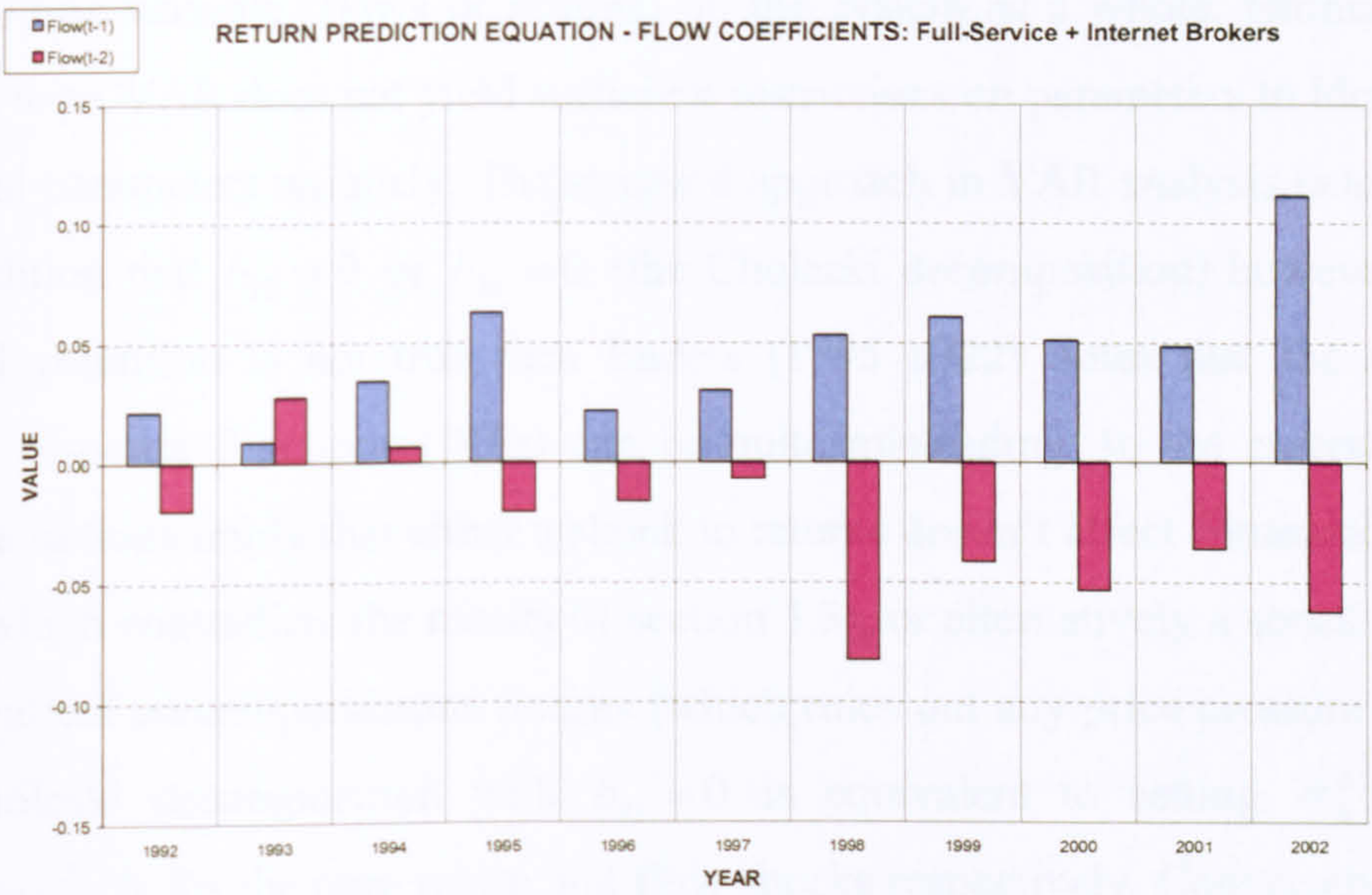
To check whether the flow-returns relationship is robust, I re-estimate the VAR equations separately for each year in the sample to assess the stability of the coefficients over time. The flow coefficients in the return prediction equation are shown in Figure 3.9. We can see that the coefficient for flow(t-1) is positive for every single year in the sample. The coefficient on flow(t-2) is negative on average. Under the null that the coefficient is zero, the probability of observing 11 positive values is highly significant using a sign test. Additionally I estimate year-by-year coefficient estimates for each type of broker separately (internet, full-service, mixed and institutional). These results are presented in Figure 3.10. This shows that in contrast to the full-service broker results, the corresponding coefficients for mixed and institutional brokers are not consistently positive over the sample period. Over the full period, mixed and institutional brokers have a zero (or slightly negative) coefficient for flow(t-1) in predicting future aggregate market returns. Finally I repeat the analysis using excess market returns over the risk-free rate. This results in virtually identical coefficients to those in Table 3.10.

It appears that the flow – return relationship is reasonably robust, at least for full-service brokerage clients. The reduced form VAR can be used to make forecasts of future returns, however to interpret the VAR economically we need to identify the coefficients in the underlying structural VAR.

$$Az_t = B_0 + B_1 z_{t-1} + B_2 z_{t-2} + B_3 z_{t-3} + B_x x_t + e_t$$

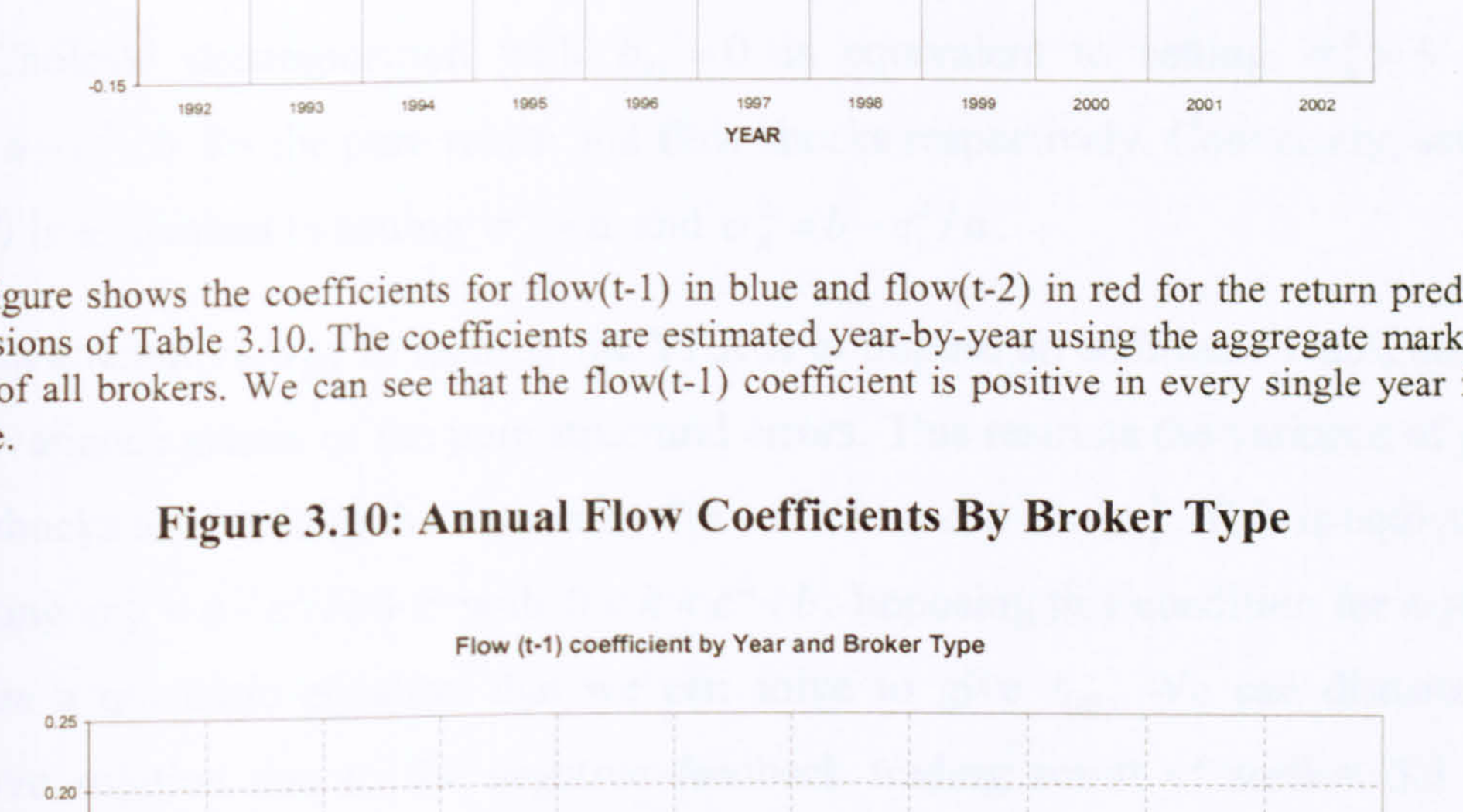
$$\text{where } A = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, E(e_t e_t') = \Omega = \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_R^2 \end{bmatrix}$$

Figure 3.9: Annual Flow Coefficients in Return Forecasting Equation



This figure shows the coefficients for flow(t-1) in blue and flow(t-2) in red for the return prediction regressions of Table 3.10. The coefficients are estimated year-by-year using the aggregate market net flows of all brokers. We can see that the flow(t-1) coefficient is positive in every single year in the sample.

Figure 3.10: Annual Flow Coefficients By Broker Type



This figure shows the coefficients for flow(t-1) in the return prediction regression of Table 3.10 by broker type for each separate year. The coefficients are estimated year-by-year using the aggregate market net flows of all brokers. The flow (t-1) coefficient is positive in every year for full-service and internet brokers, unlike for other broker types.

Identification allows us to make statements about the economic effect of a pure shock to one variable (flows or returns) on the system as a whole. Estimating the reduced form VAR does not yield sufficient restrictions on parameters to identify the structural parameters uniquely. The standard approach in VAR analysis is to impose the condition that $b_{12}=0$ or $b_{21}=0$ (the Choleski decomposition) however if this imposed condition is not true then Enders (1995 p322) notes that the resulting impulse response functions (IRFs) can be quite misleading. In the current setting these restrictions imply that either a shock to returns doesn't affect contemporaneous flows (which contradicts the results of section 3.3), or alternatively a shock to flows doesn't affect contemporaneous returns (which rules out any price pressure effects). The Choleski decomposition with $b_{21}=0$ is equivalent to setting $\sigma_R^2=b$ and $\sigma_f^2=a-c^2/b$ for the pure return and flow shocks respectively. Conversely, setting $b_{12}=0$ is equivalent to setting $\sigma_f^2=a$ and $\sigma_R^2=b-c^2/a$.

An alternative way to identify the VAR is to impose an additional restriction on the covariance matrix of the pure structural errors. This restricts the variance of pure flow shocks to lie within the extremes of $a-c^2/b$ and a for σ_f^2. This is equivalent to setting $\sigma_f^2=a-c^2/b+k$ with $0<k<c^2/b$. Imposing this condition for a given k gives a quadratic equation that we can solve to give b_{12}. We can discard the negative solution due to the negative feedback trading result of section 3.3 that implies $b_{12}>0$. The impulse response functions generated using this assumption are shown in Figure 3.11 and Figure 3.12 for settings of k=0 (equivalent to the standard Choleski approach with $b_{21}=0$), and k= 0.01[68]. The top panel of each figure shows the reaction of flows and returns to a one-standard deviation shock to returns. The bottom panel shows the reaction to a one-standard deviation shock to flows. A

[68] As is common with all VAR analysis, a significant drawback is that the identification choice is necessarily arbitrary (Hamilton (1994 pp335)). I present charts for k = 0.01. Increasing k leads to a larger contemporaneous reaction of returns to flows, but has little impact on the remainder of the IRF (the effect of our identification assumption quickly becomes negligible). Charts for other settings of k are available on request.

bootstrapping procedure is used to generate one standard error bands for the impulse response functions, following the method suggested by Hamilton (1994)[69].

The major difference between the two IRFs generated is the contemporaneous reaction of returns to flows. In the Choleski there is no immediate effect (due to the restriction). In the second IRF, returns react positively to a shock to returns.

We can economically interpret the IRFs as follow. For a pure positive return shock, small investors trade in a negative feedback fashion leading to immediate negative flows. These flows dampen the original positive return shock leading to short-term under-reaction to new information. Flows are persistently negative for several weeks after the return shock delaying the incorporation of new information into prices. Conversely for a positive flow shock, returns increase contemporaneously and in the following week consistent with short-term price pressure. The effects of the short-term price pressure are reversed in the subsequent week, despite the strong persistence of flows.

In summary, this section shows that there is a *positive* relationship between individual investor flows and future short-term returns at the aggregate market level[70]. At the market level, flows predict one-week ahead market returns, but returns exhibit a subsequent reversal in the following week. This result is consistent with short-term price pressure induced by the aggregate actions of small investors moving into or out of the market.

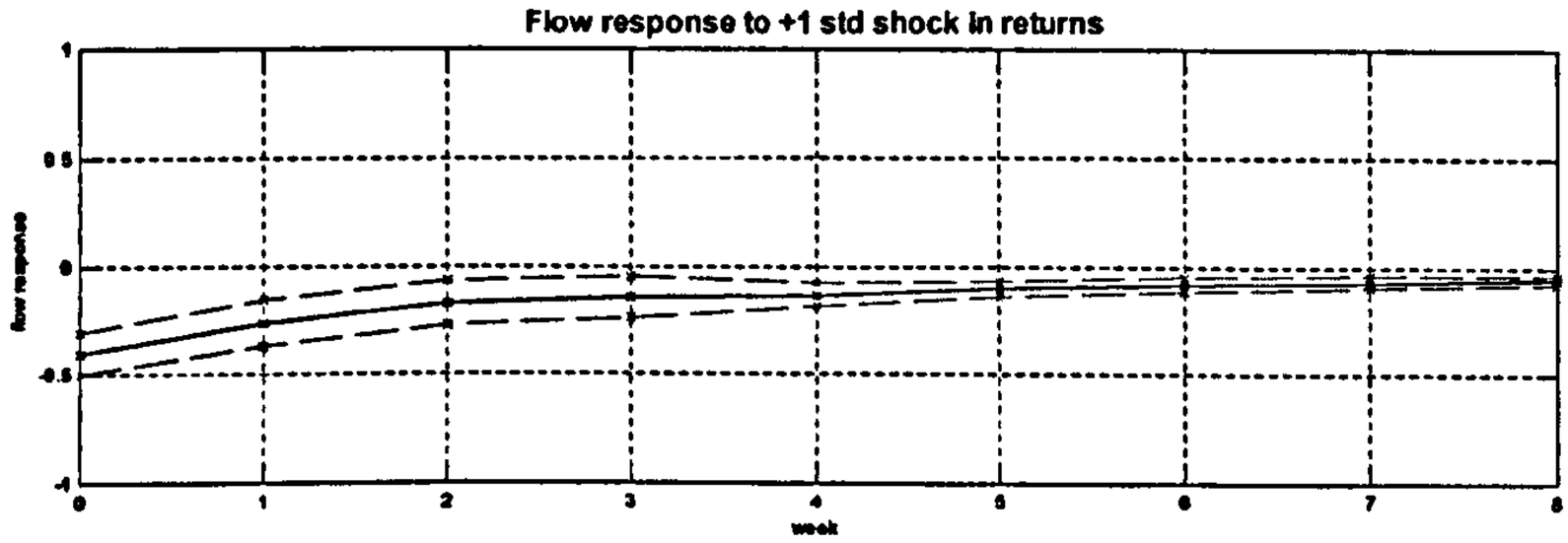
---

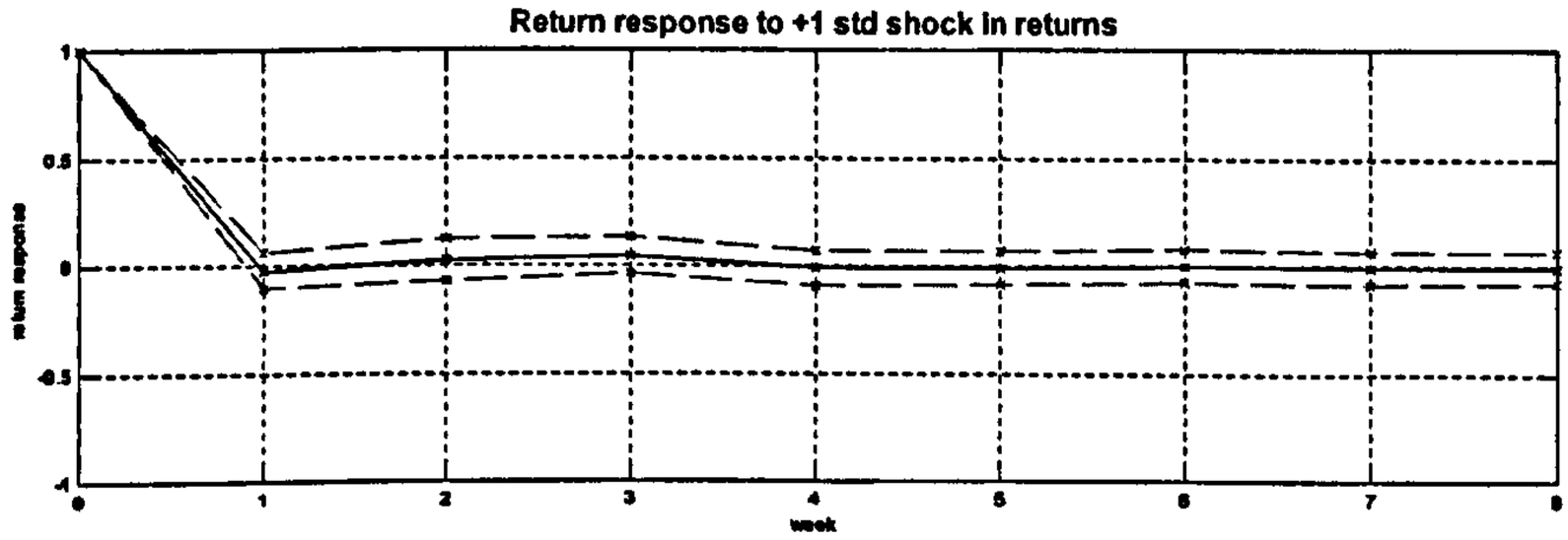
[69] First the VAR is estimated using the original return and flow series, yielding an estimate of the reduced form phi parameter matrices. The residuals from this equation are used as an empirical distribution for the bootstrap. 1000 bootstrap samples are generated drawing from this distribution using these estimated phi parameter matrices. For each sample, a new path for returns and flows is generated and the phi matrices are re-estimated. Using these new phi estimates, the restriction on the variance of flows is imposed and the corresponding structural VAR is identified. This generates a new impulse response function for each bootstrap sample. Finally, the standard error of the impulse response function is estimated from the distribution of the 1000 draws of the impulse response function from the bootstrapping process.

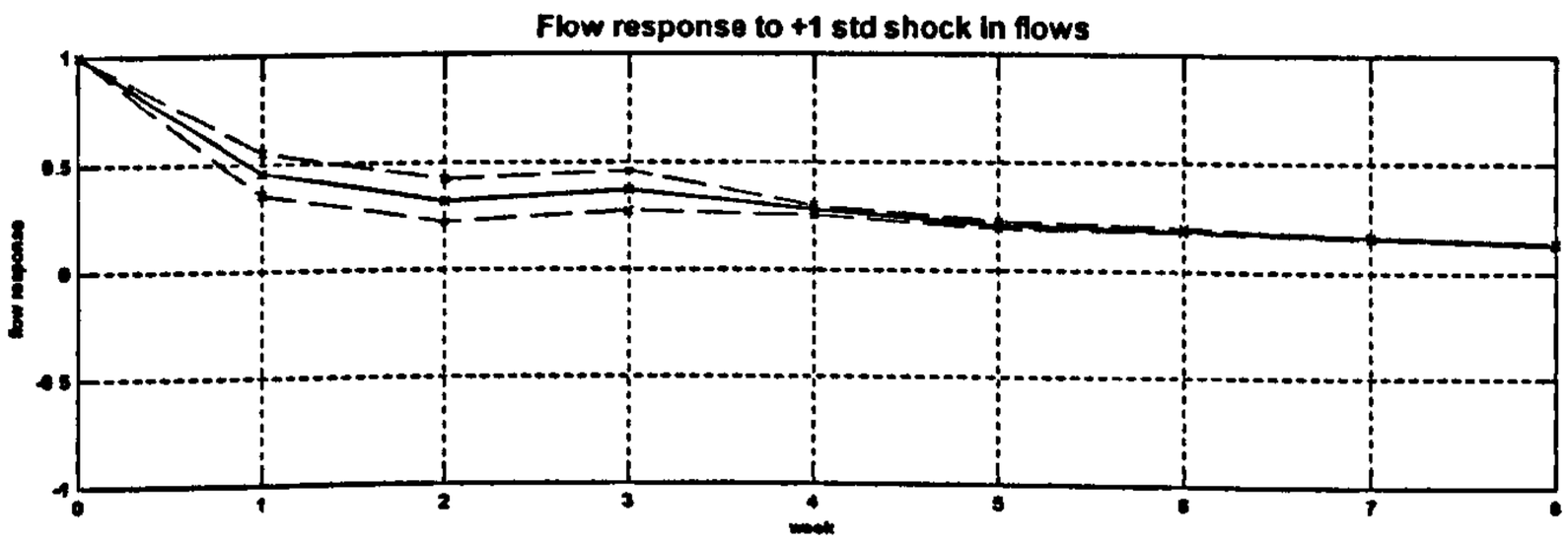
[70] The relationship for longer-term returns is not examined in this paper. Kelly (1997) finds that individual investor participation predicts low aggregate market returns over the following one and three years.

Figure 3.11: Impulse Response Functions for Market Flows and Returns, K=0.0

The upper two panels in this figure show the response of flows and returns to a +1 standard deviation shock in returns. The lower two panels show the response to a +1 standard deviation shock in flows. One standard deviation error bands are calculated using the bootstrap method described in the chapter. For this figure, returns and flows are expressed in standard deviation units. Figure 3.11 shows the results for Choleski decomposition.







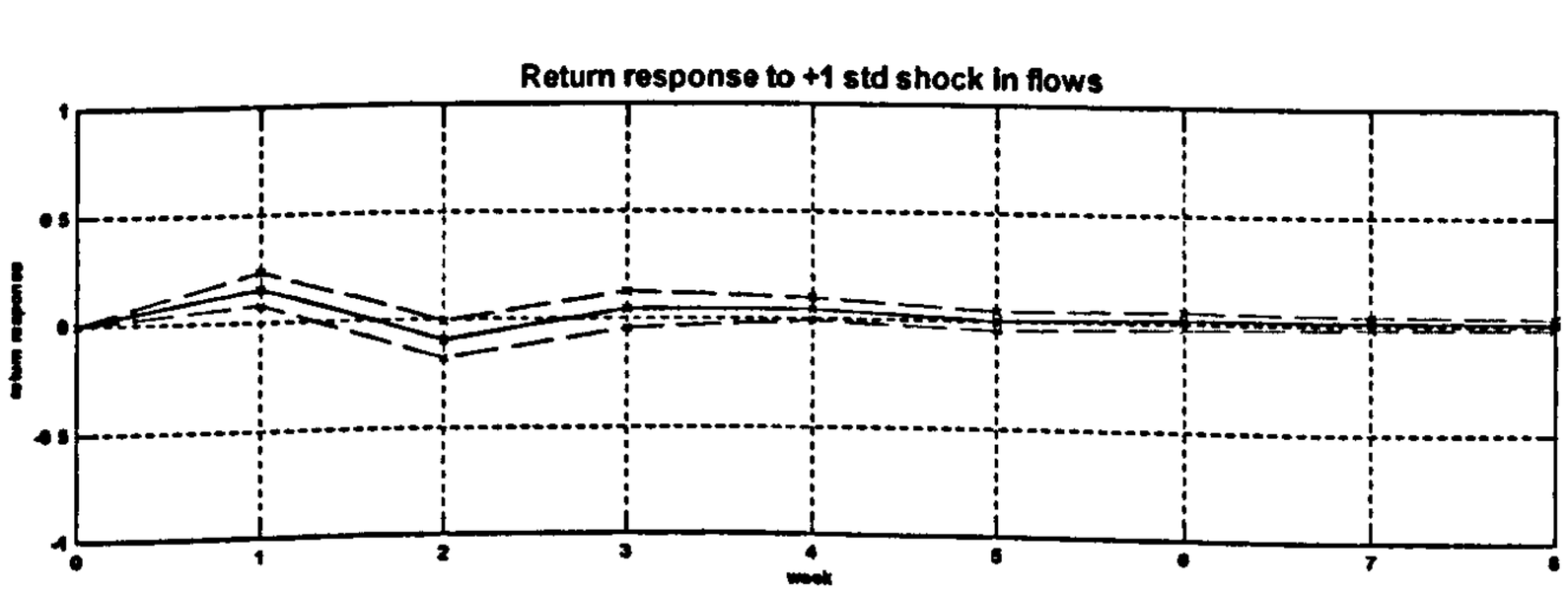
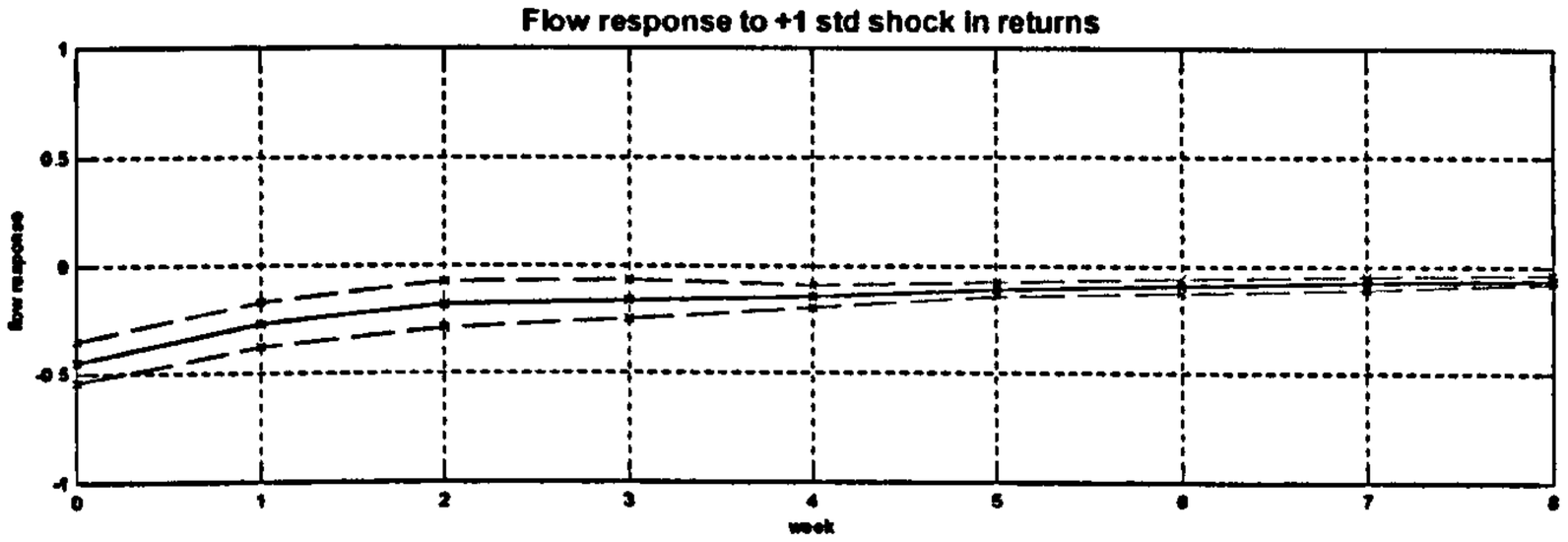


Figure 3.12: Impulse Response Functions for Market Flows and Returns, K=0.01

The upper two panels in this figure show the response of flows and returns to a +1 standard deviation shock in returns. The lower two panels show the response to a +1 standard deviation shock in flows. One standard deviation error bands are calculated using the bootstrap method described in the chapter. For this figure, returns and flows are expressed in standard deviation units. Figure 3.12 shows the results for the covariance restriction identification.



3.5.2 Cross Sectional Flows

In this section the relationship between retail investor flows and returns at the cross-sectional level is examined. Net flows are also examined separately for full-service and internet brokerage clients. To estimate the relationship I estimate a cross-sectional reduced form VAR system with the following specification:

$$\begin{bmatrix} f_{it} \\ r_{it} \end{bmatrix} = \begin{bmatrix} \alpha_f \\ \alpha_r \end{bmatrix} + \begin{bmatrix} \phi_{11}(L) & \phi_{11}(L) \\ \phi_{11}(L) & \phi_{11}(L) \end{bmatrix} \begin{bmatrix} f_{it-1} \\ r_{it-1} \end{bmatrix} + \begin{bmatrix} e_{it}^f \\ e_{it}^r \end{bmatrix} \qquad (6)$$

where :

$f_{i,t}$ = net flow for stock i in week t,

$r_{i,t}$ = return for stock i in week t.

As in Vuolteenaho (2002), the phi parameters are assumed to be the constant across each stock to reduce the number of parameters to estimate. Parameters are estimated using equation-by-equation Fama-Macbeth OLS estimates on the panel of stocks. The VAR system is estimated with 12 lags in order to capture the full dependence of flows and returns on prior values. In the estimation, returns are cross-sectionally demeaned and standardised by the cross-sectional standard deviation to remove time-specific common shocks to mean returns and to reduce the effects of time-varying cross-sectional dispersion. In the robustness checks, various alternative excess return measures are also used.

Table 3.11 shows the parameter estimates for the reduced form VAR on cross-sectional flows. Net flows are strongly persistent over time and negatively related to past returns, as shown in the earlier results. Returns also show the usual autocorrelation patterns identified in the literature, with short-term (1-2 week) mean reversion (Conrad, Gultekin and Kaul (1997)) and medium term (3-12 week+) momentum (Jegadeesh and Titman (1993)). Additionally, the most novel finding is that retail investor flows *positively* and significantly predict returns over the following two weeks. This effect seems to have a permanent price impact, as there is no evidence of subsequent reversals in returns (in contrast to the aggregate market

results). Splitting the net flows up into internet and full-service flows over the reduced 1996-2002 period shows that this positive predictability is again driven predominantly by the net flows of *full-service* brokerage clients.

Is this predictability result robust? To assess this issue, firstly I re-estimate the VAR system replacing raw returns with excess returns calculated by matching each stock to 10 size benchmark portfolios. This doesn't affect the result, with the t-statistics on flow(t-1), flow(t-2), flow(t-3) in the return equation being 3.78 (vs 4.29), 2.83 (vs 3.59), and 2.09 (vs 1.85) respectively. Similarly using 25 (5x5) size and value sorted matching portfolios to calculate excess returns, or 27 (3x3x3) size, value, 6-month momentum matching portfolios does not alter the predictability result[71]. In each case past flows predict future short-term excess returns[72].

As an additional check I re-estimate the VAR parameters for full-service brokers for each individual year 1992-2002. The coefficient estimates for the sum of the flow (t-1) and flow (t-2) coefficients are shown in Figure 3.13. For comparison, the coefficients for the other broker types are also included; institutional, mixed, and internet. The results show that the coefficients for the full-service brokers are positive in *every* single yearly sub-period for the full-service brokers. This is not the case for the institutional or mixed brokers. The coefficients for the internet brokers are smaller than those of the full-service brokers.

---

[71] For the 25 size value portfolios the t-stats are 4.01, 3.81, and 1.14 and for the 27 size, value, momentum portfolio the t-stats are 4.22, 3.84, 0.31 respectively. ie the results are not significantly effected by the benchmark chosen for excess returns.

[72] Using unstandardised returns doesn't alter the result either. The t-statistics on the flow coefficients are 4.88, 2.83, and 2.09 respectively indicating the results are if anything slightly stronger.

Table 3.11 – VAR Estimates for Cross-Sectional returns and flows

This table reports parameter estimates for the cross sectional flow VAR. The process is estimated for all investors and separately for internet and traditional full-service investors. The parameters in the table correspond to the reduced form VAR system with 12 lags. For each parameter I report the estimated coefficient value and the T-statistic of the coefficient estimated using Newey West standard errors. The VAR is estimated from September 1991 to Dec 2002.

Variable	Full-Service + Internet: 1991-2002				Internet: 1996-2002				Full-Service: 1996-2002			
	Flow (t)	*tstat*	*Return (t)*	*tstat*	*Flow (t)*	*tstat*	*Return (t)*	*tstat*	*Flow (t)*	*tstat*	*Return (t)*	*tstat*
Constant	0.001	0.47	0.001	0.52	0.03	4.67	0.008	1.72	0.00	-0.42	0.000	0.07
StockRet (t-1)	-0.035	-15.72	-0.076	-10.36	-0.05	-7.39	-0.052	-7.31	-0.04	-14.59	-0.049	-6.92
StockRet (t-2)	-0.017	-10.51	-0.014	-2.84	-0.02	-6.40	-0.003	-0.46	-0.02	-9.85	0.000	0.00
StockRet (t-3)	-0.013	-8.60	0.013	2.57	-0.01	-3.03	0.017	2.21	-0.02	-7.01	0.024	3.29
StockRet (t-4)	-0.009	-6.43	0.019	3.71	-0.01	-2.51	0.020	2.48	-0.01	-5.69	0.030	4.50
StockRet (t-5)	-0.008	-5.86	0.010	2.08	-0.01	-1.89	0.011	1.37	-0.01	-7.25	0.017	2.95
StockRet (t-6)	-0.008	-5.04	0.010	2.22	0.00	-0.92	0.008	1.11	-0.01	-3.59	0.010	1.83
StockRet (t-7)	-0.005	-3.76	0.013	2.85	0.00	-1.15	0.007	0.99	-0.01	-2.91	0.013	2.15
StockRet (t-8)	-0.004	-3.04	0.015	3.31	0.00	-0.07	0.016	2.32	0.00	-1.70	0.018	2.94
StockRet (t-9)	-0.004	-3.06	0.014	2.71	0.00	-0.22	0.014	1.90	0.00	-3.70	0.013	1.85
StockRet (t-10)	0.000	-0.25	0.013	2.48	0.00	1.53	0.022	3.02	0.00	-1.06	0.022	3.69
StockRet (t-11)	0.001	0.45	0.012	2.81	0.00	-0.10	0.013	1.76	0.00	1.07	0.014	2.15
StockRet (t-12)	0.000	0.27	0.009	2.17	0.01	1.86	0.014	2.13	0.00	1.71	0.014	2.48
Flow (t-1)	0.262	34.18	**0.036**	**4.29**	0.25	23.00	**-0.010**	**-1.06**	0.26	46.26	**0.065**	**7.25**
Flow (t-2)	0.101	28.81	**0.026**	**3.59**	0.08	15.14	**0.018**	**1.99**	0.11	29.36	**0.034**	**3.28**
Flow (t-3)	0.074	20.06	0.013	1.85	0.08	11.53	0.015	1.47	0.08	16.89	0.017	1.62
Flow (t-4)	0.056	15.04	0.006	0.86	0.05	6.85	0.006	0.65	0.06	10.31	0.003	0.35
Flow (t-5)	0.043	12.10	0.000	-0.01	0.04	5.94	-0.010	-1.01	0.04	8.48	-0.008	-0.93
Flow (t-6)	0.033	8.09	-0.006	-0.81	0.03	4.97	0.001	0.12	0.04	7.98	-0.007	-0.68
Flow (t-7)	0.023	7.22	0.003	0.39	0.02	3.44	-0.004	-0.42	0.02	4.34	0.009	0.90
Flow (t-8)	0.029	8.09	-0.012	-1.60	0.03	4.38	-0.007	-0.68	0.02	4.89	-0.024	-2.05
Flow (t-9)	0.032	9.04	0.000	0.02	0.03	6.99	0.004	0.44	0.03	6.97	-0.001	-0.12
Flow (t-10)	0.030	7.89	0.005	0.67	0.03	5.79	-0.009	-0.93	0.03	6.19	0.012	1.39
Flow (t-11)	0.024	7.28	-0.017	-2.33	0.02	4.86	-0.022	-2.49	0.02	4.74	-0.013	-1.29
Flow (t-12)	0.028	8.23	-0.009	-1.24	0.02	5.37	0.004	0.58	0.03	8.27	-0.021	-2.39

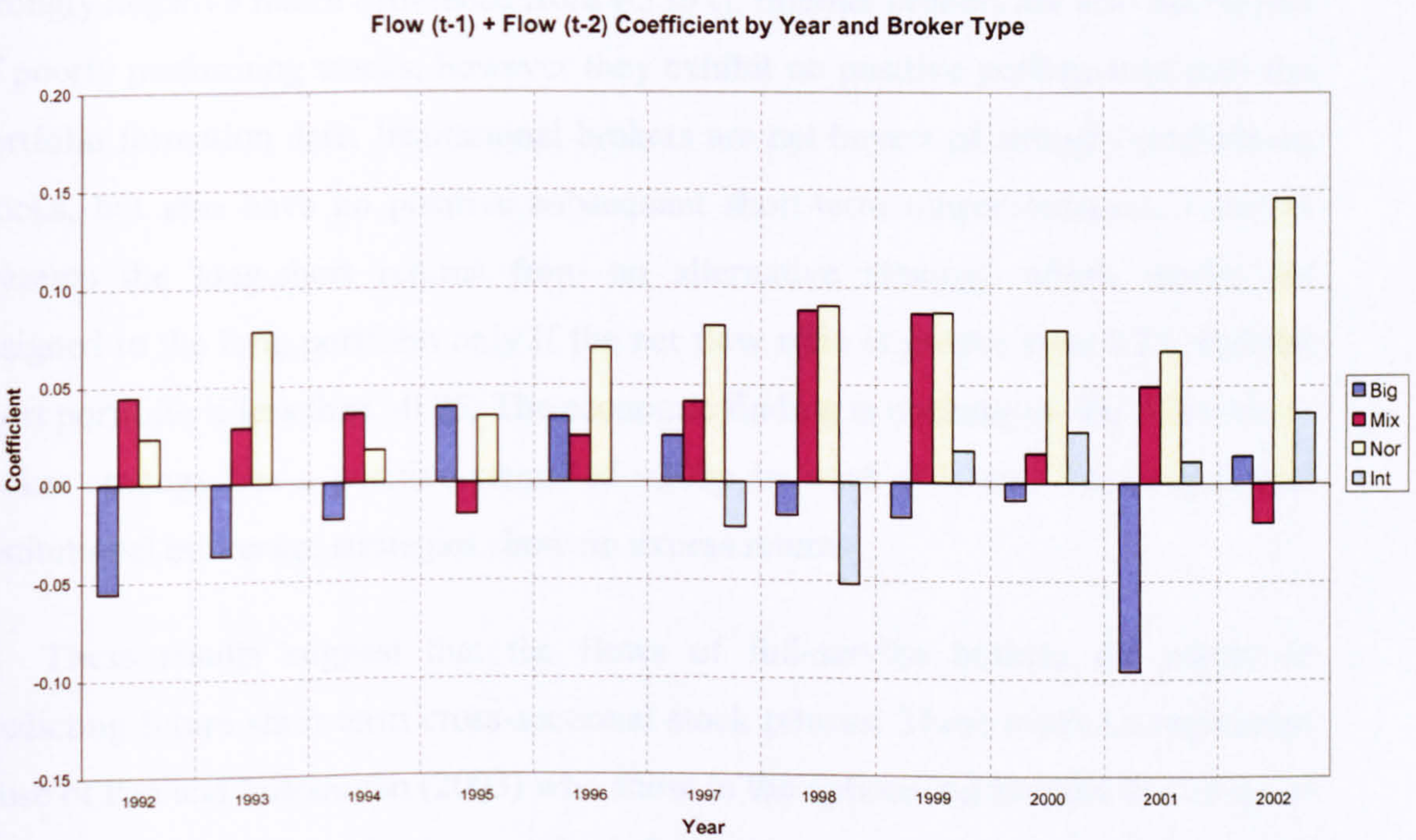
Table 3.12 – VAR Estimates for Cross-Sectional returns and flows: Large and Small Stocks

This table reports parameter estimates for the cross sectional flow VAR estimated separately for large and small stocks. The process is estimated for all investors and separately for internet and full-service investors. The parameters in the table correspond to the reduced form VAR system with 12 lags. For each parameter I report the estimated coefficient value and the T-statistic of the coefficient estimated using Newey West standard errors. The VAR is estimated from September 1991 to December 2002. Only the return equations are reported, with the flow equations omitted to save space.

Variable	Full-Service + Internet: 1991-2002				Internet: 1996-2002				Full-Service: 1996-2002			
	Large (Top 100)		Small (Ex 100)		Large (Top 100)		Small (Ex 100)		Large (Top 100)		Small (Ex 100)	
	Return (t)	tstat	Return (t)	tstat	Return (t)	tstat	Return (t)	tstat	Return (t)	tstat	Return (t)	tstat
Constant	-0.026	-5.28	0.011	3.75	-0.026	-2.74	0.015	2.39	-0.021	-3.15	0.011	3.64
StockRet (t-1)	-0.103	-13.28	-0.072	-8.23	-0.100	-9.20	-0.043	-4.02	-0.098	-12.02	-0.045	-5.66
StockRet (t-2)	-0.017	-2.53	-0.019	-3.04	-0.014	-1.41	-0.006	-0.64	-0.003	-0.34	-0.003	-0.45
StockRet (t-3)	0.008	1.11	0.008	1.29	0.007	0.76	0.020	1.85	0.014	1.54	0.021	2.28
StockRet (t-4)	0.013	1.72	0.019	3.20	0.011	1.07	0.028	3.27	0.016	1.52	0.032	4.65
StockRet (t-5)	0.012	1.49	0.007	1.31	0.014	1.35	0.001	0.09	0.023	2.38	0.011	1.75
StockRet (t-6)	0.008	1.22	0.010	1.73	-0.001	-0.11	0.005	0.61	0.005	0.56	0.012	1.82
StockRet (t-7)	0.004	0.50	0.016	2.81	-0.007	-0.51	0.009	0.94	0.005	0.47	0.015	1.95
StockRet (t-8)	0.011	1.51	0.020	3.54	0.005	0.43	0.025	2.76	0.013	1.24	0.024	3.40
StockRet (t-9)	0.018	2.43	0.017	2.83	0.025	2.70	0.007	0.75	0.011	1.24	0.015	1.96
StockRet (t-10)	0.018	3.07	0.007	1.14	0.029	3.22	0.006	0.62	0.018	1.90	0.018	2.44
StockRet (t-11)	0.008	1.26	0.012	2.32	0.014	1.28	0.009	1.11	0.010	1.05	0.010	1.47
StockRet (t-12)	0.011	2.09	0.008	1.49	0.006	0.62	0.016	1.94	0.015	2.12	0.015	2.16
Flow (t-1)	0.077	4.79	0.019	1.78	0.001	0.05	-0.028	-2.52	0.116	5.50	0.052	4.56
Flow (t-2)	0.001	0.08	0.022	2.60	0.018	1.03	0.013	1.28	0.006	0.23	0.035	3.21
Flow (t-3)	0.048	3.55	0.007	0.76	0.023	1.07	0.006	0.44	0.065	3.60	0.004	0.34
Flow (t-4)	-0.015	-1.05	0.012	1.34	0.011	0.66	0.009	0.77	-0.026	-1.58	0.013	1.36
Flow (t-5)	0.021	1.16	-0.007	-0.65	-0.013	-0.60	0.001	0.07	0.011	0.49	-0.010	-0.92
Flow (t-6)	-0.022	-1.57	-0.001	-0.08	0.005	0.26	-0.009	-0.66	-0.030	-1.46	0.006	0.49
Flow (t-7)	0.033	1.97	-0.003	-0.33	-0.002	-0.07	0.012	0.95	0.052	2.38	-0.002	-0.17
Flow (t-8)	-0.020	-1.32	-0.006	-0.64	0.006	0.28	-0.001	-0.08	-0.055	-2.97	-0.014	-1.09
Flow (t-9)	0.018	1.07	-0.006	-0.58	0.052	2.50	-0.018	-1.20	0.005	0.21	-0.007	-0.55
Flow (t-10)	-0.008	-0.51	0.006	0.57	-0.010	-0.44	-0.019	-1.46	-0.013	-0.70	0.016	1.38
Flow (t-11)	-0.031	-2.24	-0.016	-1.68	-0.012	-0.55	0.001	0.07	-0.012	-0.58	-0.014	-1.04
Flow (t-12)	-0.007	-0.47	-0.002	-0.21	0.002	0.10	0.008	0.62	-0.018	-0.88	-0.024	-2.26

As a third robustness check, I estimate the equations for small and large stocks separately. Large stocks are defined as the largest 100 stocks in each weekly cross-section, while small stocks are all stocks in the index outside the top 100. For brevity, I only report the coefficients in the return prediction equation. We can see from Table 3.12, that the basic return patterns are similar for large and small stocks (short-term mean reversion, medium term momentum). Additionally, small investor flows predict future weekly stock returns for *both* large and small stocks (albeit with a lag for smaller stocks). Again, full-service broker flows appear to be driving this effect.

Figure 3.13: Flow Coefficients in Return Forecasting Equation: Year by Year



This figure shows the sum of the coefficients on flow(t-1) and flow(t-2) for the cross-sectional return(t) prediction regressions of Table 3.11. The coefficients are estimated year-by-year separately for each broker type. The sum of the coefficients is positive in every year for the full-service retail brokerage firms.

An alternative approach to assess the robustness of these results is to construct portfolios using a methodology similar to that in Pan and Poteshman (2003). Each week I form a long and a short portfolio based on the net flows of the four different broker types. Stocks are allocated to the long portfolio if the broker type is a net buyer, and to the short portfolio if a net seller. The returns of the long and short portfolio are assessed over a window of t-5 to t+10 weeks for each of the 4 broker types (full-service, internet, mixed, and institutional). Table 3.13 Panel A presents the returns of a long-short portfolio based on this strategy around the weekly event date. We can see that the full-service broker zero investment strategy has a positive 21bp return in week t+1 (t-stat = 7.9), and a return of +38bp over the first four weeks. Stocks in this portfolio have been poor previous performers (shown by the strongly negative return difference from t-5 to t). Internet brokers are also net buyers of poorly performing stocks, however they exhibit no positive performance post the portfolio formation date. Institutional brokers are net buyers of strongly performing stocks, but also have no positive subsequent short-term outperformance. Panel B presents the long-short returns from an alternative strategy, where stocks are assigned to the long portfolio only if the net flow ratio is greater than 0.25, and the short portfolio if less than –0.25. The economic finding is unchanged, the full-service broker strategy has a positive return of +30bp in week t+1, and the internet and institutional brokerage strategies show no excess returns.

These results suggest that the flows of full-service brokers are useful in predicting future short-term cross-sectional stock returns. These results complement those of Pan and Poteshman (2003) who show in the options market that the trades of full-service brokerage clients contain information about future short-term stock returns. The difference in forecasting ability of the internet and full-service brokerage clients is consistent with the latter paying higher brokerage commissions in return for valuable private information. Choe, Kho, and Stulz (2000) also find that domestic individuals have valuable short-term private information in the Korean

market[73]. Alternatively Griffin et al (2003) do not find evidence that individual order imbalances predict future returns in 9-months of Nasdaq data, however they do not separate internet from full-service investors.

An interesting question is why full-service broker flows predict future returns? Given no evidence of subsequent reversals, temporary price pressure seems unlikely. One possible explanation is that these investors have valuable private information. Alternatively, individual investors may be acting as liquidity providers to institutional brokers, allowing them to unload inventories of stock acquired through principal trades. Institutional brokers cross much of their stock internally via upstairs trading, and might be willing to pay on average a small premium to small investors to take stock that cannot be disposed of in this system. Future research will explore these possible explanations in more detail.

Finally I estimate the impulse response function for net flows and stock returns using the variance restriction outlined in section 3.5.1. The estimated impulse response functions are shown in Figure 3.14. A positive flow shock leads to a persistent flow reaction in following weeks. Additionally short-term returns are increased at week 0, and slightly at weeks 1 and 2 consistent with the reduced form results. A positive shock to returns leads to an immediate and persistent negative shock to flows, consistent with negative feedback trading by small investors. The positive return shock also generates well-documented patterns of one-week mean reversion and medium term price momentum.

---

[73] Froot, O'Connell, and Seasholes (2001) examine international investor flows and find evidence that flows predict future market returns in emerging equity markets. They also employ a bivariate VAR methodology in their analysis.

Table 3.13 – Portfolio Returns Based on Net Flows By Brokerage Type

This table reports the portfolio returns for long and short portfolios constructed based on the net flows of each broker type. The zero cut-off portfolios assign a stock to the long portfolio is the brokerage type is a net buyer each week, the 0.25 cut-off assigns a stock to the long portfolio only if the net flow ratio is >0.25. Results are reported in basis points per week for the various portfolios.

Panel A: Zero Cut-off

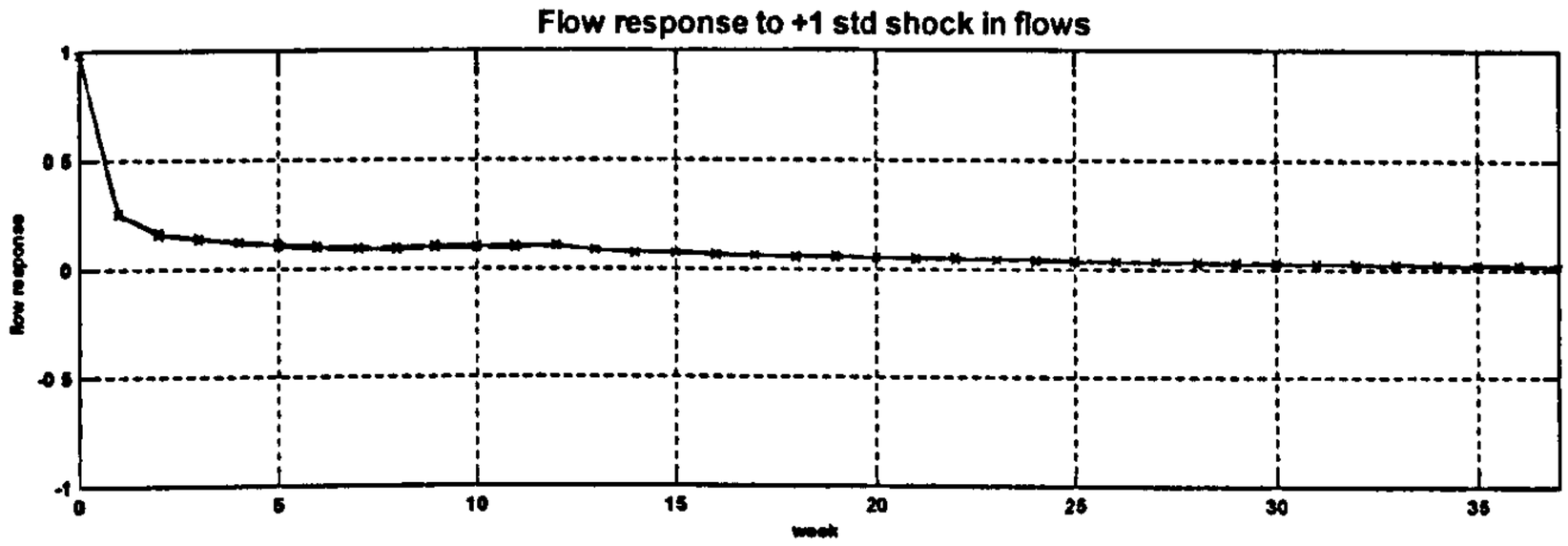
Broker	Portfolio	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10
Big	Long	34.0	34.2	31.5	33.2	45.5	60.7	10.4	13.4	12.8	13.6	14.7	13.3	13.8	14.1	13.8	12.7
(1991-2002)	Short	0.4	-1.3	1.7	-1.3	-12.6	-28.2	15.8	12.3	12.9	10.9	9.5	11.2	12.4	11.6	11.7	13.2
	Difference	**33.7**	**35.5**	**29.8**	**34.4**	**58.1**	**88.9**	**-5.4**	**1.1**	**-0.1**	**2.7**	**5.2**	**2.2**	**1.4**	**2.5**	**2.1**	**-0.5**
	t-stat	10.6	10.9	8.7	9.7	15.2	21.9	-1.8	0.4	0.0	0.9	1.7	0.7	0.5	0.9	0.7	-0.2
Mix	Long	23.5	19.5	22.7	21.4	22.4	27.2	16.7	14.3	16.9	12.4	12.3	16.0	15.5	17.0	13.4	13.5
(1991-2002)	Short	8.4	12.2	9.0	9.3	7.6	2.7	9.8	10.5	8.2	11.3	10.9	8.8	9.8	8.0	12.1	11.4
	Difference	**15.0**	**7.3**	**13.7**	**12.1**	**14.8**	**24.5**	**6.9**	**3.9**	**8.7**	**1.1**	**1.3**	**7.1**	**5.7**	**9.0**	**1.4**	**2.2**
	t-stat	5.2	2.6	4.9	4.2	5.1	8.2	2.4	1.4	3.1	0.4	0.5	2.5	2.1	3.2	0.5	0.8
Full-Service	Long	-4.6	-3.6	-8.1	-10.3	-23.4	-25.0	23.7	18.6	13.2	14.6	14.0	11.5	13.7	10.1	11.6	12.7
(1991-2002)	Short	36.0	33.9	37.8	38.7	50.4	51.2	2.5	7.0	12.0	10.1	10.2	13.0	11.8	16.0	13.0	11.7
	Difference	**-40.6**	**-37.5**	**-45.9**	**-49.1**	**-73.8**	**-76.2**	**21.2**	**11.6**	**1.2**	**4.4**	**3.8**	**-1.5**	**1.9**	**-6.0**	**-1.4**	**1.0**
	t-stat	-13.8	-12.9	-15.2	-17.1	-22.2	-21.2	7.0	3.9	0.4	1.6	1.3	-0.5	0.7	-2.1	-0.5	0.4
Int	Long	-13.6	-14.5	-17.7	-29.3	-45.7	-85.5	5.8	4.1	4.5	0.1	3.4	3.0	0.7	-0.2	-0.3	-1.3
(1996-2002)	Short	42.5	44.7	48.2	60.9	85.7	139.2	8.9	14.5	11.3	13.9	14.2	15.3	15.3	15.3	14.3	15.8
	Difference	**-55.8**	**-59.2**	**-65.7**	**-90.8**	**-132.1**	**-224.5**	**-3.0**	**-12.0**	**-6.9**	**-13.8**	**-11.7**	**-13.2**	**-14.5**	**-15.7**	**-14.6**	**-16.9**
	t-stat	-12.3	-12.7	-14.9	-19.2	-26.8	-33.8	-0.6	-2.6	-1.6	-3.1	-2.4	-2.9	-3.1	-3.4	-3.4	-3.9

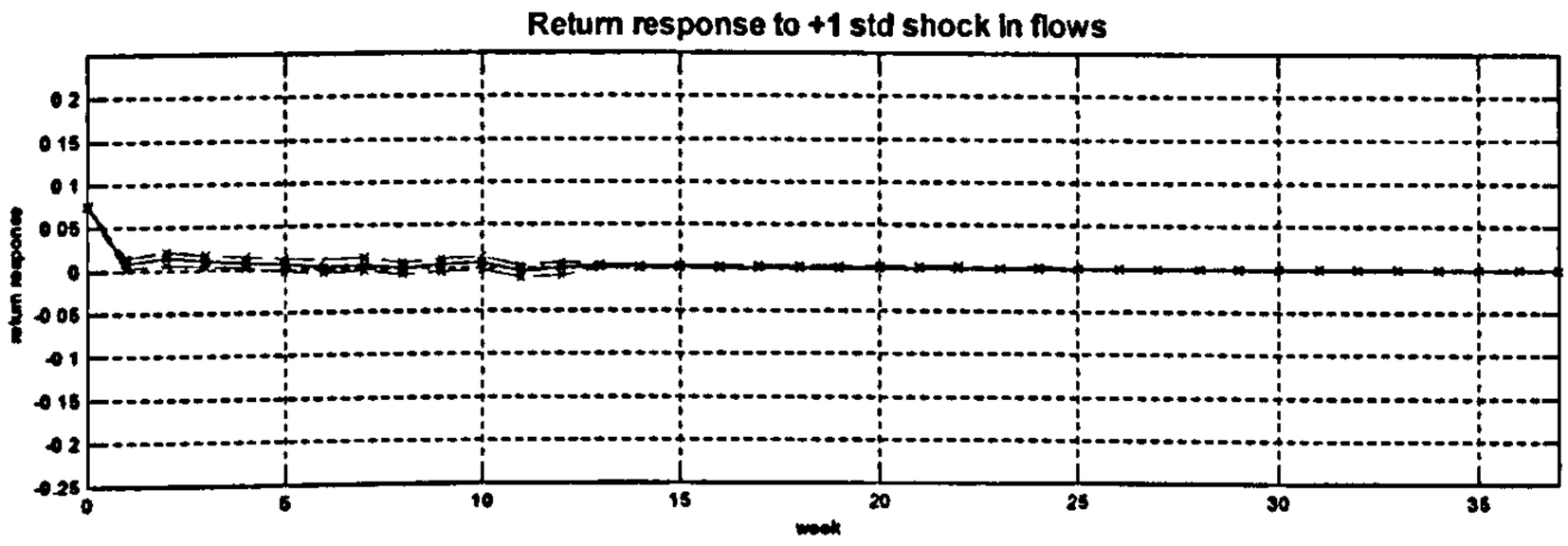
Panel B: Cut-off of Net flow = 0.25

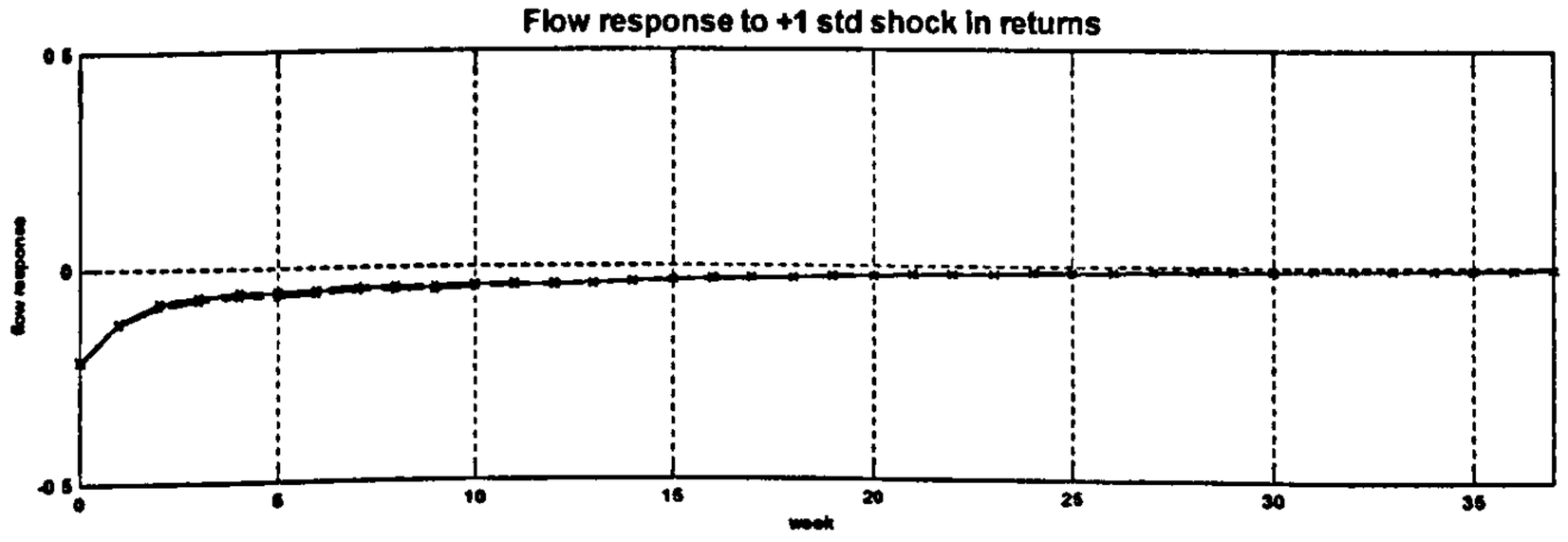
Broker	Portfolio	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10
Big	Long	31.2	30.0	23.9	28.4	38.9	52.0	0.2	5.5	11.8	11.4	5.7	13.3	10.8	10.5	12.7	8.8
	Short	-16.7	-18.8	-10.3	-16.8	-29.6	-46.9	9.9	7.1	2.8	4.4	-0.7	8.2	5.2	5.3	5.8	10.8
	Difference	47.9	48.8	34.1	45.2	68.6	98.8	-9.7	-1.6	8.9	6.9	6.5	5.2	5.6	5.2	6.9	-2.0
	t-stat	8.6	8.1	5.8	7.3	10.9	14.8	-1.8	-0.3	1.8	1.3	1.1	1.0	1.1	1.0	1.3	-0.4
Mix	Long	23.9	18.7	18.6	14.3	16.2	17.5	15.3	11.7	19.4	12.5	14.4	16.9	13.6	16.1	16.3	13.6
	Short	3.4	2.9	4.2	3.2	-5.7	-14.7	4.6	3.2	7.1	5.7	4.7	4.7	7.1	3.0	9.9	6.6
	Difference	20.5	15.8	14.4	11.1	21.9	32.2	10.7	8.5	12.3	6.8	9.7	12.2	6.5	13.1	6.4	6.9
	t-stat	4.6	3.6	3.5	2.4	5.4	7.1	2.4	1.9	2.8	1.6	2.2	2.7	1.5	3.0	1.5	1.7
Nor	Long	-7.1	-7.2	-15.2	-20.0	-40.5	-46.7	30.2	24.7	18.7	19.9	18.7	12.6	17.4	11.7	15.4	13.5
	Short	45.3	43.2	46.1	48.7	58.7	58.3	-0.2	7.6	12.9	12.8	12.6	16.7	13.5	18.5	14.0	13.5
	Difference	-52.4	-50.4	-61.3	-68.7	-99.2	-105.0	**30.4**	**17.1**	5.8	7.1	6.0	-4.1	3.9	-6.8	1.4	0.0
	t-stat	-14.6	-13.4	-16.5	-19.0	-24.7	-24.2	**7.9**	**4.4**	1.5	1.9	1.6	-1.1	1.0	-1.9	0.4	0.0
Int	Long	-14.4	-21.7	-24.3	-36.9	-67.7	-132.4	9.2	9.2	7.3	2.0	6.3	3.9	4.3	1.9	3.0	-2.6
	Short	52.6	53.1	55.6	66.1	95.2	150.9	12.5	17.8	17.1	18.5	17.1	19.7	19.5	21.2	15.2	21.5
	Difference	-66.9	-74.8	-79.8	-103.8	-163.9	-283.5	-2.7	-11.4	-9.5	-16.2	-11.3	-16.1	-15.1	-18.8	-12.1	-23.5
	t-stat	-12.9	-14.6	-15.1	-19.7	-27.0	-33.5	-0.5	-2.1	-1.9	-3.1	-2.0	-3.0	-2.9	-3.5	-2.4	-4.6

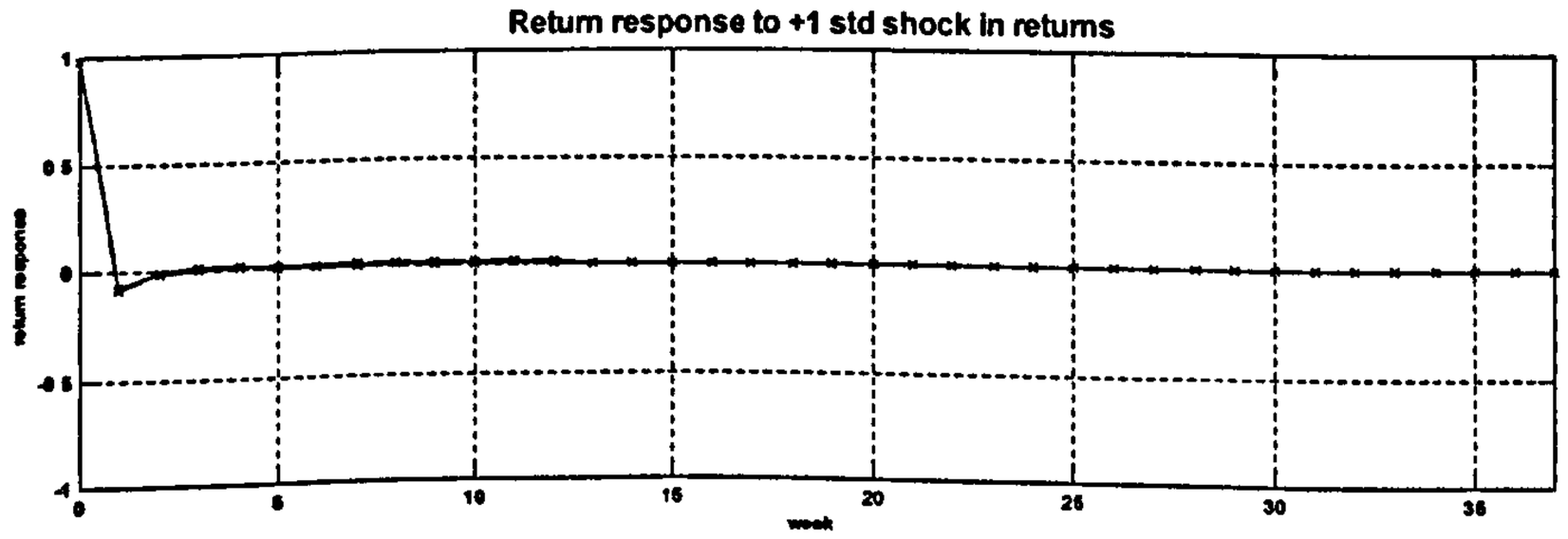
Figure 3.14: Impulse Response Functions for Cross-Sectional Flows and Returns

The upper two panels in this figure show the response of returns and flows to a +1 standard deviation shock in flows. The lower two panels show the response to a +1 standard deviation shock in returns. Both flows and returns are expressed in standard deviation units.









In summary section 3.5 shows there is a strong relationship between aggregated small investor flows and short-term returns. In contrast to models that assume individual investors are irrational, this section shows that the flows of these investors actually positively predict short-term returns at both the aggregate and cross-sectional levels. Further examination reveals that full-service brokerage clients predominantly drive this result.

3.6. Conclusion

This chapter uses Australian data over the period 1991 to 2002 to examine whether individual investors' trades contain systematic patterns that survive after aggregation. Aggregate individual trades reveal a strong negative relationship between net flows and returns indicating that individual investors are *negative feedback traders* at a weekly frequency. This is true for both aggregate market level flows (flows into and out of the market as a whole) and cross-sectional flows. The data also reveals strong persistence in flows, which is especially prevalent at the cross-sectional level. Small investors are net buyers of stocks with high recent volatility. Aggregate small investor trading activity is high in the first half of January and in the second half of June (before tax-year end).

Examining net flows at the cross-sectional level individually for 56 retail brokers reveals that net flows across independent firms are significantly positively related. The average correlation between weekly cross-sections of net flows for the typical internet brokerage firm pair is 0.23 and for full-service brokers is 0.16. For the top 50 stocks these correlations increase to 0.44 and 0.23 respectively. Additionally *every* single retail broker pair has a significantly positive correlation coefficient. These results confirm that the actions of individual investors are indeed highly pervasive across a large number of unrelated brokerage firms.

The final section examines the relationship between small investor flows and future returns. At the market level, I find that the net flows of full-service brokerage

clients predict future short-term aggregate market returns. This finding is consistent with short-term price pressure given evidence of subsequent return reversals. At the cross-sectional level I find that full-service clients' net trades positively predict returns for a period of up to 2 weeks with no evidence of subsequent reversals. This finding is consistent with small investors using valuable private information, or acting as liquidity providers allowing institutional brokers to unload inventory positions.

In summary, the main finding of this chapter is that individual investor trades *are* highly systematic across a large number of brokerage firms. This implies individual trading patterns aggregate strongly and have the potential to influence market dynamics. However the conjecture that individual investors are irrational does not be valid since the flows of full-service investors *positively* predict future short-term market and cross-sectional returns.

Future research could further examine the interaction between institutional and retail investors to distinguish liquidity provision from private information explanations. Studies using individual investors trades and other sources of disaggregated order flow have the potential to reveal much about the real-life behaviour of different types of investors and their impact on financial markets.

3.7. Appendix 1: Detrending of Market Level Flows

Market level buy, sell, and total flows exhibit substantial non-stationarity and positive skewness. To model these variables we need to transform them to achieve stationarity. Lo and Wang (2000) discuss various transformation methods that have previously been used for transforming volume data including removing linear or log linear time trends, taking differences, and using non-parametric kernel regressions.

Several obvious factors affect the low frequency dynamics of the flows in this sample including the growing market share of the brokers in the sample (see Figure 3.1), and the growth in overall market turnover and market capitalisation over the period. In order to determine the appropriate de-trending methodology, the log(buy), log(sell), and log(total) flows were first evaluated for the presence of a stochastic trend (unit root). An augmented Dickey-Fuller test with time trend rejects the presence of a stochastic trend for all three series at the 1% significance level[74]. To preserve the information in the levels (which is destroyed by taking differences), the data is de-trended using a deterministic polynomial function of order five[75] and the residuals are standardised by their mean and standard deviation. The fitted polynomial fitted trend looks similar to a linear time trend with some curvature to capture the rapid growth of the small brokerage volume in the 1997-2000 period and subsequent plateauing of this series. Detrended series are also standardised to have mean of zero and standard deviation of one.

[74] The lag length used in the unit root test was one, which was determined using the AIC criterion. Examining the autocorrelation function for first differences of the buy, sell, and total flow data shows no significant lags other than the first.

[75] The polynomial order was determined using the Schwartz information criterion. For polynomials of order > 5 the Schwartz criterion is reasonably flat (the AIC continues to fall). However, for too high an order polynomial, we are no longer removing low frequency information from the series, but start to remove the high frequency information, which defeats the purpose of the de-trending. The fifth order polynomial has enough degrees of freedom to fit the low frequency dynamics, without removing the high frequency information.

To assess the robustness of the buy, sell, and total flow results I repeat the analysis in Table 3.4 for buy, sell, and total flows using firstly raw flows and secondly flows detrended using a Hodrick-Prescott filter. This produces the same significant return, flow, volatility, and calendar effects. Repeating the analysis using log-differenced flows also produces the same effects for past returns, volatility, and calendar variables, but gives negative auto-correlations for the past flow variables due to the over-differencing induced by the transformation.

4. Noise Trader Risk Exists... but the noise traders are not who you think they are

4.1. Introduction

A long running and unresolved debate is whether behavioural finance 'matters'. While there is evidence at the micro and experimental level that individuals do not always behave in a fully rational manner, evidence that these errors aggregate or have real economic effects is relatively scarce. This chapter aims to contribute to this debate by using a unique database of individual investor trades to examine the empirical importance of a key behavioural theory, noise trader risk, in a broad equity market setting. The objective is to test whether noise trader risk exists and can be associated with a particular group of agents, namely individual investors.

'Noise trader risk' is defined in the sense of the seminal model of DeLong, Shleifer, Summers, and Waldmann (1990) (DSSW). In this model, rational investors with short horizons fail to fully offset the mispricing induced by correlated noise traders due to uncertainty about future noise trader demand. A non-fundamental factor (investor sentiment) exists that is systematic across assets and is priced in equilibrium. DSSW (1991) show that noise traders can earn higher returns than rational investors by bearing the extra priced risk that is generated by their own trading activities. This enables them to survive long-run extinction from financial markets.

Whether noise trader risk is economically important is a disputed issue. Shleifer (2000) suggests that noise trader risk is a fundamental limit to arbitrage in equity markets. Similarly DSSW (1989) argue that noise trader risk can lead to deleterious welfare effects including mispricing, excess volatility, and ultimately a dramatic reduction in the capital stock and consumption growth rate of an economy. Other

economists are sceptical of the empirical importance of the theory, for example West (1988) states "there is little direct evidence that trading by naïve investors plays a substantial role in stock price determination".

Existing empirical evidence on the significance of noise trader risk is equally mixed. Empirical tests have focussed mainly on closed-end-fund discounts as a proxy for noise trader sentiment. While this measure is quite clean, due to the measurability of closed-end-fund net asset values, it remains based on an economically unimportant sector of the equity market and whether it is the best way to capture noise trader risk is not clear. The main empirical evidence supporting the noise trader risk theory is by Lee, Shleifer, and Thaler (1991) who find that small stocks co-move with closed-end-fund discounts suggestive of a correlated sentiment factor associated with individual investors. Other researchers (often strongly) dispute the validity and significance of this result (Chen, Kan, and Miller (1993), Elton, Gruber, and Busse (1998)) while other authors find mixed results (Sias, Starks, and Tinic (2001) and Gemmill and Thomas (2003)[76]).

This chapter differs from previous empirical research by using a unique database of individual investor trading to examine noise trader risk in a broader equity market setting. By pinning noise trader risk down to a subset of investors I am able to construct an alternative measure of noise trader 'sentiment' and conduct several empirical tests designed to detect noise trader risk in equity returns[77]. Additionally

---

[76] Sias, Starks, and Tinic (2001) find closed-end-funds exhibit higher volatility than their underlying assets, but find no evidence that noise trader risk is priced. Gemmill and Thomas (2003) find that closed-end-fund discounts vary with an external measure of sentiment (open-end mutual fund flows), but also find no evidence that noise trader risk is priced.

[77] A related paper taking a similar approach is Brown et al (2002) who find suggestive evidence of a sentiment factor in US and Japan using a short sample of daily mutual fund flows. My paper differs from theirs by using a long sample of *direct* flows by individual investors rather than using mutual fund flows as a proxy for sentiment. In concurrent research, Kumar and Lee (2003) also construct a measure of small investor sentiment using the daily buy-sell imbalance from a large discount broker. My paper differs from theirs by using the *percentage of trading* accounted for by small investors in a stock to proxy for the level of noise trader presence, and I also use data from a significantly larger number of retail brokerage firms.

from a welfare perspective, finding the group of investors (if any) who generate noise trader risk may support regulation, taxation or education of these investors to ameliorate any adverse economic effects.

The unique database used in the study is relatively large compared to existing research that uses individual investor trading data. The database contains over 39 million retail investor trades from 47 retail brokers in Australia over a 7-year period[78]. I use this database to test the joint hypothesis that noise trader risk exists and is associated with individual investors. In looking for potential candidates for noise traders, small investors seem an obvious choice. It is difficult to dispute that individuals possess limited information relative to institutions and previous empirical evidence shows that individuals appear to exhibit less than fully rational behaviour in their stock trading decisions. For example it has been empirically shown that small investors hold undiversified stock portfolios, sell stocks that subsequently outperform the stocks they buy, exercise options in a clearly irrational manner, and trade too frequently[79]. In terms of size, individuals are a substantial group of investors, directly owning a large percentage of outstanding equity in many world markets. In 1999/2000 the ownership figures for individuals were 40% for the US, 16% for the UK, 20% for Japan, and 24% for Australia. Additionally, participation by individuals shows significant intertemporal[80] and cross-sectional variation. In summary small investors are a large group, appear to behave less than fully rationally

---

[78] By comparison, Barber and Odean (2000) use around 3 million trades in aggregate from a single broker, Barber and Odean (2002) use around 8.5 million trades from 3 retail brokers and Grinblatt and Keloharju (2000) use around 2 million trades in aggregate across all investors for a two year period. Note however that these papers have account level information, while this paper uses aggregate information only for each of the brokers.

[79] See Brennan (1995) for a discussion of individuals vs institutions. Blume and Friend (1975), Polkovnichenko (2002) and ASX (2001) present empirical evidence on the diversification of small investor portfolios. Odean (1998) examines the performance of sells vs buys. Barber and Odean (2000) show individuals seem to trade too often and finally Poteshman and Serbin (2001) present evidence on irrational option exercise.

[80] In the UK, the number of individual investors went from 4m to 11m between 1981 and 1990, in Japan the number went from 20m to 35m between 1990 and 2000. In Australia the number went from 1.2m in 1991 to 5.7m in 2000 (Source: Proshare 2000). Dramatic increases in the number of individual shareholders also occurred in the US between 1910 and 1925, and decreases occurred in the 1930's (Warshow (1923) and Froman (1934)).

and their participation in trading is variable. These features make them a plausible potential source of systematic noise trader risk.

Of course, small investors may *not* generate such a risk factor. If their actions are insufficiently correlated across stocks or are not driven by irrational sentiment, then no systematic risk will exist. Additionally, individuals may not be the sole source of noise trader risk in equity markets. We can easily broaden the definition of noise trader risk to include not only sentiment risk, but also any non-fundamental influence on the demand of a subset of investors that is sufficiently correlated across stocks. For example we can replace the notion of an aggregate psychological influence that affects individuals with various institutional frictions, causing institutions to behave in a correlated manner unrelated to fundamentals, to generate a similar effect. Non-fundamentally driven changes in institutional demand may be associated with factors such as performance related mutual fund flows, common investment strategies, style investing, herding due to career concerns, or by changes in demand induced by investment mandate restrictions (such as benchmark composition changes).

The first section of this chapter presents a simple model to clarify several predictions of the noise trader risk theory and motivate the empirical tests. There are two types of trader in the model, rational traders and noise traders. As the percentage of noise traders in a particular stock increases, the theory predicts that (all else equal) the volatility of the stock should rise (excess volatility) and the correlation of the stock with a portfolio of other high noise trader participation stocks should also rise (excess comovement). Additionally the correlation between any two stocks will be increasing in the product of the proportion of noise traders in each stock.

The remainder of the chapter empirically examines these predictions using data on small investor trading to assess whether individuals qualify as a likely source of noise trader risk. Trading, rather than ownership, data is used to measure the proportion of noise traders in each stock since, as noted by Black (1986), "noise traders must trade to have their influence". Additionally the model demonstrates a

clear monotonic relationship between trading and the proportion of sentiment investors associated with a stock. Three empirical tests are performed looking in turn for evidence of excess volatility and excess comovement, and for evidence of a noise trader risk factor in cross-sectional returns.

The results of the first test (excess volatility) show that when the proportion of trading volume accounted for by individuals in a given stock increases, then future volatility actually *decreases* rather than increases as predicted by the theory. This result controls for various fundamental factors that may influence volatility. Due to an adding-up constraint this result is consistent with Sias (1996) who shows that future volatility is higher after *institutional* ownership increases.

The second test (excess comovement) shows that increases in individual participation have no impact on correlations between a stock and a portfolio of high individual participation stocks, inconsistent with individuals generating a correlated sentiment effect. Intriguingly however, increases in institutional participation in a stock *do* lead to increased correlations of the stock with a portfolio of high institutional participation stocks. This suggests the possibility that institutions may be the source of a non-fundamental risk factor. This result is consistent with Barberis, Shleifer, and Wurgler (2002) who show inclusion of a stock into the S&P 500 index leads to an increase in the beta of the stock with respect to the market index.

Finally, factor tests are conducted to assess whether there is any evidence of a priced noise trader risk factor in cross-sectional returns. I find there *is* strong evidence of such a factor in returns and additionally the factor appears to be priced, consistent with the noise trader theory. The betas on the noise trader factor are highly significant for a number of test portfolios and also for a sample of individual stocks, even in the presence of other widely used factors including SMB (size), HML (value), and WML (momentum). Additionally a risk premium is associated with the noise trader factor that is robust across sub-periods and estimation methodology. However, consistent with the results of the previous two tests, the sign of this risk

premium suggests that *institutions* rather than individuals are the likely source of the related risk.

In summary, this chapter has two main conclusions. Firstly, there *is* evidence that noise trader risk exists. Evidence of excess volatility and excess correlations associated with changes in noise trader participation is found. Additionally, there is strong evidence of a priced noise trader risk factor in returns. Stocks with a high exposure to noise trader risk earn significantly higher excess returns. Secondly, individual investors do *not* appear to be the source of the noise trader risk. Institutional frictions are a much more plausible source of non-fundamental demand shocks than is an individual sentiment effect. Frictions including performance related mutual fund flows, herding due to career concerns, common investment strategies, and style investing can all generate non-fundamental influences on the demand of institutional investors leading to the existence of such a risk factor. Ultimately, whatever the source of the correlated non-fundamental demand shocks, the final effect on equilibrium returns remains the same; investors demand extra returns from stocks that are exposed to this factor, consistent with the noise trader risk theory.

The chapter is structured as follows. Section 4.2 builds a simple model of noise trader risk. Section 4.3 discusses the data used in the analysis. Section 4.4 presents the volatility results. Section 4.5 the correlation results. Section 4.6 conducts factor tests. Section 4.7 interprets the results, and Section 4.8 concludes.

4.2. Simple Model of the Affect of Noise Traders on Market Dynamics

The purpose of the simple model in this chapter is to motivate the empirical tests for excess volatility, excess correlations, and a noise trader risk factor in returns. The model also provides a rationale for the main empirical proxy used in the tests, namely retail share of turnover (RS). The model can be viewed as a stylised version of DSSW (1989).

There are two groups of trader in the model. The first group are subject to common sentiment shocks while the second group are fully rational. There are K stocks in the economy. Let:

μ_t^i = proportion of sentiment traders associated with stock i at time t.

δ_t = common sentiment (or non-fundamental) factor at time t.

P_t^i = price of stock i at time t.

F_t^i = fundamental value of stock i at time t[81].

Assume for simplicity that the fundamental value of each stock follows a random walk, with independent increments across stocks (ie no fundamental correlation between stocks).

$$F_t^i = F_{t-1}^i + \varepsilon_t^i$$

where $\varepsilon_i \sim 0, \sigma_\varepsilon^2$

The number of sentiment traders existing at time t is N while the number of rational traders at time t is M. In the model we will let $N, M \to \infty$ so both types of traders become individually atomistic. The entire mass of sentiment traders is given by μ_t^i, while the mass of rational traders is given by $1-\mu_t^i$.

Both types of trader are subject to random liquidity shocks at each point in time. For simplicity take these liquidity shocks as independent across time and across all traders. The purpose of this assumption is to generate trade in the model that is unrelated to the sentiment factor δ_t.

The demand function for a typical rational trader, j, per unit mass is given in reduced form as:

---

[81] This is the equilibrium value of the stock when the proportion of sentiment traders is equal to zero.

$$D_t^{h,Rat} = 1 + b(F_t^i - P_t^i) + z_t^{j,t}$$

where the constant b represents the slope of the demand curve capturing any limits to arbitrage facing the rational investors (for example risk aversion or agency effects arising from short evaluation horizons[82]). For $b < \infty$ the demand curve is downward sloping. As noted in Grinblatt and Han (2002), we can think of b as representing whatever generates the optimal demand function given a utility function. The rational trader also faces a liquidity shock z_t^j that is distributed $N(0, \sigma_z^2)$ and is iid across time and independent across traders.

The demand function for the typical sentiment trader, h, per unit mass is given as:

$$D_t^{h,Sent} = 1 + b(F_t^i + \delta_t - P_t^i) + z_t^{h,t}$$

where δ_t is a common sentiment factor affecting all sentiment traders at time t, across all stocks. For example $\delta_t > 0$ may represent systematic over-optimistic expectations by this class of investors. As with the rational investors, the sentiment traders also face a liquidity shock z_t^j that is distributed $N(0, \sigma_z^2)$ and is iid across time and other traders.

The supply of each stock is fixed and normalised to one unit. We can impose market clearing to determine the equilibrium price for each stock at time t as follows:

$$1 = \mu\left[\underset{N\to\infty}{\overset{a.s.}{\lim}} \frac{1}{N}\sum_{j=1}^{N} D_t^{i,j,Sent}\right] + (1-\mu)\left[\underset{M\to\infty}{\overset{a.s.}{\lim}} \frac{1}{M}\sum_{j=1}^{M} D_t^{i,j,Rat}\right]$$

---

[82] Rational traders who are risk averse will limit their demand in response to fundamental risk σ_ε^2. Also Shleifer (2000) discusses other limits to arbitrage that might limit rational investors.

The equilibrium price is given as[83]:

$$P_t^i = F_t^i + \mu\delta_t \qquad \text{a.s.}$$

The change in price is given by:

$$P_t^i - P_{t-1}^i = \varepsilon_t^i + \mu_t^i(\delta_t - \delta_{t-1}) \qquad \text{a.s.}$$

In this model, price changes are driven by either innovations to fundamental value F_t^i or innovations to the common sentiment factor δ_t.

From this result we can derive several testable hypotheses.

Hypothesis 1: *(Excess volatility) An increase in the proportion of noise traders, μ_t^i, in a stock will increase the volatility of the stock*[84].

$$\frac{\partial Var(P_t^i - P_{t-1}^i)}{\partial \mu_i^t} > 0$$

Hypothesis 2: *(Excess correlation) An increase in the proportion of noise traders, μ_t^i, in a stock will increase its correlation with the common sentiment factor δ_t.*

$$\frac{\partial Cov(P_t^i - P_{t-1}^i, \delta_t - \delta_{t-1})}{\partial \mu_t^i} > 0$$

Hypothesis 3: *(Noise Trader Risk Factor) The covariance between any two stocks will be increasing in the product of the proportion of noise traders in both stocks, μ_t^i and μ_t^j:*

---

[83] See Appendix 2 for the derivation of this result. This solution is analogous to the solution for the equilibrium price in DSSW (1989) equation (10).

[84] This hypothesis is directly analogous to the same result in DSSW (1989) equation (13). See Appendix 2 for derivations.

$$\frac{\partial Cov(P_t^i - P_{t-1}^i, P_t^j - P_{t-1}^j)}{\partial \mu_t^i \mu_t^j} > 0$$

In equilibrium the amount of stock held by the two groups of investor is given by:

Sentiment Investors: $\mu + \mu(1-\mu)b\delta$

Rational Investors: $(1-\mu) - \mu(1-\mu)b\delta$

We can see that a positive sentiment shock $\delta_t > 0$ results in rational investors selling to sentiment investors in response to the high price relative to fundamental value.

Total trading volume, Vol_t, in this model can also be calculated by comparing the optimal demand of a typical trader of a given type at time t with the optimal demand for the trader at time $t-1$. The absolute value of the difference will give the amount traded by this agent. We can then sum total traded volume across all agents to obtain the total volume traded (which we must divide by two to avoid double counting of buy and sell volume). Trading volume in this model occurs through both liquidity shocks and revisions to the sentiment factor δ_t. Total trading volume is given as:

$$2 * Vol_t = \mu \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \left| D_t^{i,j,Sent} - D_{t-1}^{i,j,Sent} \right| + (1-\mu) \frac{1}{M} \lim_{N\to\infty} \sum_{j=1}^{M} \left| D_t^{i,j,Rat} - D_{t-1}^{i,j,Rat} \right|$$

Simplifying this equation gives:

$$Vol_t = b(1-\mu)\mu \left| \delta_t - \delta_{t-1} \right| + \frac{\mu\sigma_z}{\sqrt{\pi}} + \frac{(1-\mu)\sigma_z}{\sqrt{\pi}}$$

The first term, $b(1-\mu)\mu\left|\delta_t - \delta_{t-1}\right|$, represents trading *between* sentiment traders and rational traders induced by changes in the common sentiment factor. The second and third terms represent liquidity motivated trading *within* the group of sentiment traders and the group of rational traders respectively.

We can now define our empirical proxy, retail share of turnover, as the percentage of total trading volume accounted for by the sentiment traders[85]. Retail share of turnover (RS) is defined as the ratio of trading by sentiment investors to total trading and is given as:

$$RS_t = \frac{0.5b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\mu\sigma_z}{\sqrt{\pi}}}{b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}}$$

As liquidity trading becomes more important relative to trading induced by changes in the sentiment factor, δ_t, then $RS_t \to \mu$. If liquidity trading is non-existent ($\sigma_z = 0$) then $RS_t \to 0.5$ (ie every trade by a sentiment investor is matched by an identical opposite trade by a rational investor). When both forms of trading exist, then RS_t lies between 0.5 and μ.

Finally we derive an explicit expression for the percentage of trading volume that takes place *within* groups vs the percentage of trading volume that takes place *between* groups. The proportion of total trading that is within-group is given as:

$$within_group_trading_t = \frac{\frac{\sigma_z}{\sqrt{\pi}}}{b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}}$$

One minus this expression gives the proportion of trading that occurs between groups. As liquidity trading becomes more important, within-group trading approaches 100%. Conversely, if liquidity trading is not present at all ($\sigma_z = 0$) then between-group trading approaches 100%. For fixed b, σ_z, δ_i, between-group trading is maximised when $\mu = 0.5$.

---

[85] In the empirical tests we employ the initial assumption that retail investors are equivalent to the sentiment investors in this model.

Using the trading volume results, we can now motivate the use of retail share (RS) as a valid empirical proxy for the unobservable variable μ_t^i.

Lemma 1: *There is a monotonic increasing one-to-one relationship between* μ_t^i *and RS. Ranking stocks based on RS will give identical orderings as ranking stocks on the basis of the true (unobserved)* μ_t^i. *Conversely, the relationship between holdings of sentiment investors (SH) and* μ_t^i *is not necessarily monotonic. Therefore ranking stocks on SH will not necessarily give the same orderings as ranking stocks based on* μ_t^i.

To clearly illustrate the relationship between retail share, holdings and μ_t^i Figure 4.1 presents plots of the three variables for two different cases. Both cases have b=1, $\delta_{t-1} = 0$, and $\sigma_z = 0.4$, but differ in the value of δ_t. In the case where $\delta_t = 0.4$ (Figure 4.1.A) there is a monotonic positive one-to-one mapping between the three variables. In the case where $\delta_t = +2$ (Figure 4.1.B) the retail share variable is monotonic in μ_t^i but the holding variable is not.

In summary, the role of the model is to motivate the three main empirical tests in this chapter. These tests allow us to evaluate the joint hypothesis that noise trader risk exists and individual investors are the source of the noise trader risk. The simple model presented here captures the broad predictions from existing noise trader risk models such as DSSW (1989) and DSSW (1990), namely noise trader risk will lead to excess volatility and excess correlations between stocks. In addition the model provides justification for using retail share (RS) as an empirical proxy for the unobservable variable μ_t^i. The next section discusses the unique data set used in this chapter that allows us to observe RS on a high frequency basis (weekly) over a 380-week period.

Figure 4.1: Relationship between RS, Sentiment Holdings and μ.

These figures plot the simulated relationship between retail share of turnover (RS), holdings of retail investors, and the true unobserved proportion of retail investors μ using the model in this chapter for two different parameter choices. The figures show RS is monotonic in μ while SH need not be.

Figure 4.1.A: Relationships for Parameter Set 1

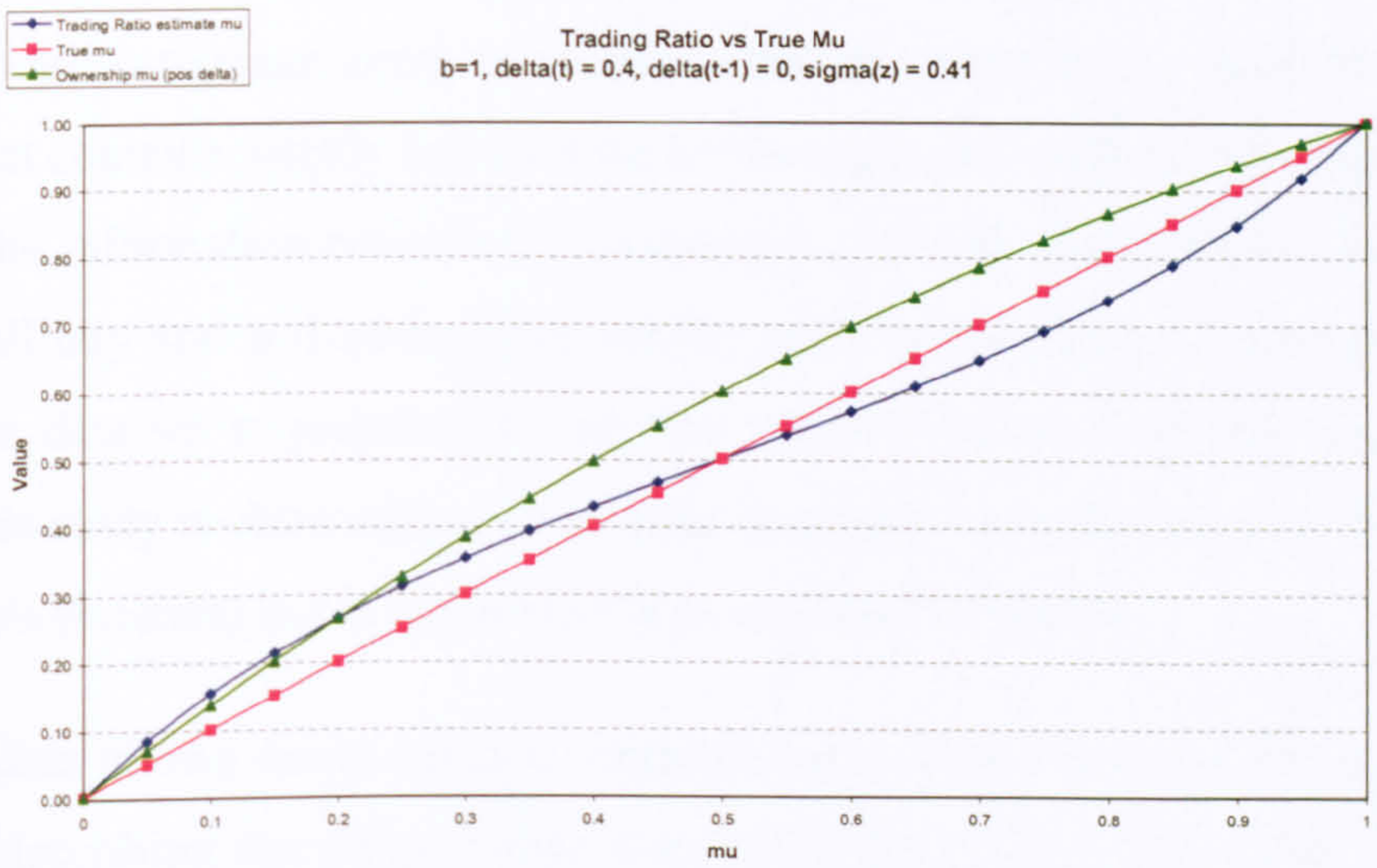


Figure 4.1.B: Relationships for Parameter Set 2

4.3. Data Description

This chapter uses a unique data set on the trading activity of individual investors in Australia from 12th December 1994 to 30th March 2002. This data set is relatively large compared to existing samples of retail investor activity and allows us to address the question of noise trader risk in a unique way. I collect data on individual trades from 47 retail brokerage firms in Australia for all stocks in the Australian market. The dataset contains weekly information on the aggregate trades of all clients of each broker. The information consists of the number of trades, value traded, and average price of all buy and sell trades executed by each broker during the week, for each stock. The data set is proprietary and was supplied by an Australian broker. Data used in this study is observable in real-time on a trade-by-trade basis by some market participants (brokers) but is not available to investors in general.

The data allows me to build a comprehensive record of retail flows. Summary statistics describing the dataset appear in Table 4.1. In aggregate over 39 million transactions are represented with an aggregate value of over $A386.6 billion. The average trade size for the small brokers is around $A9,800[86].

[86] Occasionally institutional trades appear in the records for the retail brokers. To remove the influence of these trades on the final results, stock weeks with an average trade size of >$A50000 for either buy or sell transactions were removed from the retail broker records. The number of records removed was small, 0.31% of Internet broker stock-weeks were removed, and 1.49% of full service broker trades were removed and placed in a separate category (miscellaneous trades), which were excluded from small investor aggregate trades.

Table 4.1: Summary Description of Aggregate Data (in $A)

This table presents summary statistics for the data sample. Results are shown for all 47 retail brokers, and split into full service, internet brokers and trades that were excluded due to large trade size. The sample period is December 1994 to March 2002.

	Broker Type			
Variable	**47 Retail Brokers**	**Full Service**	**Internet**	**Misc. Trades**
Total # Buys	20,210,410	11,371,988	8,466,553	371,869
Total # Sells	19,214,511	11,819,333	6,976,596	418,582
Total Buy Value ($M)	193,023,192,490	107,066,093,814	52,249,787,344	33,707,311,332
Total Sell Value ($M)	193,645,982,535	111,578,137,417	47,486,178,020	34,581,667,098
% Total Buy Value	100%	55%	27%	17%
% Total Sell Value	100%	58%	25%	18%
Average Buy Trade ($)	9,551	9,415	6,171	90,643
Average Sell Trade ($)	10,078	9,440	6,806	82,616
# Brokers	47	38	9	0
Stock-Weeks	3,198,235	2,411,745	760,936	25,554

To assess the likely impact of individual investors on market dynamics, the first question of interest is what percentage of total market equity individual investors own. Figure 4.2 displays the percentage ownership breakdown between individuals, foreigners, corporations, and various financial intermediaries. Individuals owned between 18% and 24% of all equity outstanding in Australia over the sample period. Foreigners owned between 35% and 45%, financial intermediaries between 25% and 35%, and banks, corporations, and government the remainder. Table 4.2 presents international comparisons for the US, UK, Japan, and Germany over a longer period. The aggregate direct holdings of individual investors in the UK, Japan, Germany, and Australia are quite similar at around 20% whilst the US has a higher level of around 45%. Also notable is the large decline in the share of equity owned directly by individuals over a longer historical period, for example in the US and UK the share of equity held directly by individuals has fallen from 69% and 47% respectively in 1969, to 42% and 15% in 1999. The similar level of individual ownership in Australia as in the UK, Japan, and Germany suggest that the effects found in this chapter are also likely to be of a similar magnitude in these countries.

A second question of interest is the percentage of *trading volume* (by value) accounted for by individuals and institutions respectively. Groups that trade heavily may have more impact on market dynamics than groups that do not participate in the market regularly. Figure 4.3 shows the proportion of total market turnover attributed to the 47 retail brokers over time. The average proportion is 6.58% for the full period, and 9.32% for the second half of the period for all index stocks. The percentage of trading accounted for by retail brokers is quite low, averaging less than 10% over the period despite their ownership being greater than 20%. This is consistent with findings in the UK where the London Stock Exchange (2000) finds that individuals account for around 8% of trading by value[87]. It is also consistent with the findings by various exchanges (eg ASX, NYSE) that show a large proportion of

[87] Transaction Survey 2000, London Stock Exchange.

individual investors trade infrequently, and a substantial proportion not at all. This result emphasises the small size of individual investors in terms of overall market turnover. In terms of the average index stock however, individuals are more important. For example Figure 4.4 shows that the proportion of turnover accounted for by individuals in the average index stock has risen from 15% to 40% over the sample period.

Figure 4.2: Breakdown by Ownership for Listed Australian Equities

This figure gives a breakdown by owner type for all Australian listed equities over time. The data is derived from ABS Financial Accounts, ABS Cat. no. 5232.0 using their standard definitions. It presents the proportion of the entire equity market owned by individuals (bottom), foreigners (second bottom), corporations (third), insurance and pension funds, financial intermediaries (mutual funds), government, and banks (top) over the sample period from 1995 to 2001.

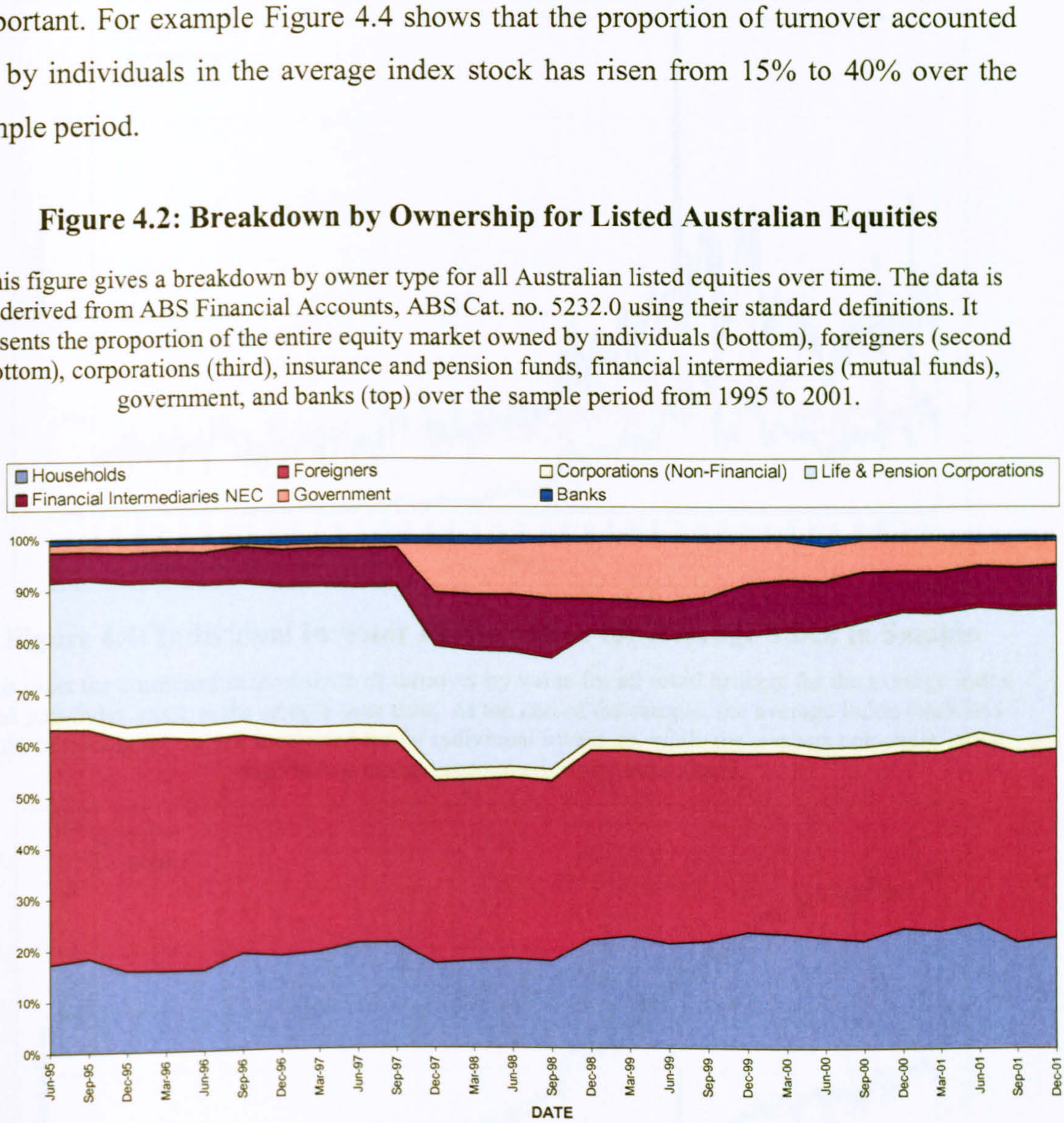


Figure 4.3: Aggregate market share of sampled brokers over time

This figure shows the market share of the brokers in the sample from December 1994 to March 2002. Market share is calculated as the value of trades executed by these brokers relative to the value of total market turnover.

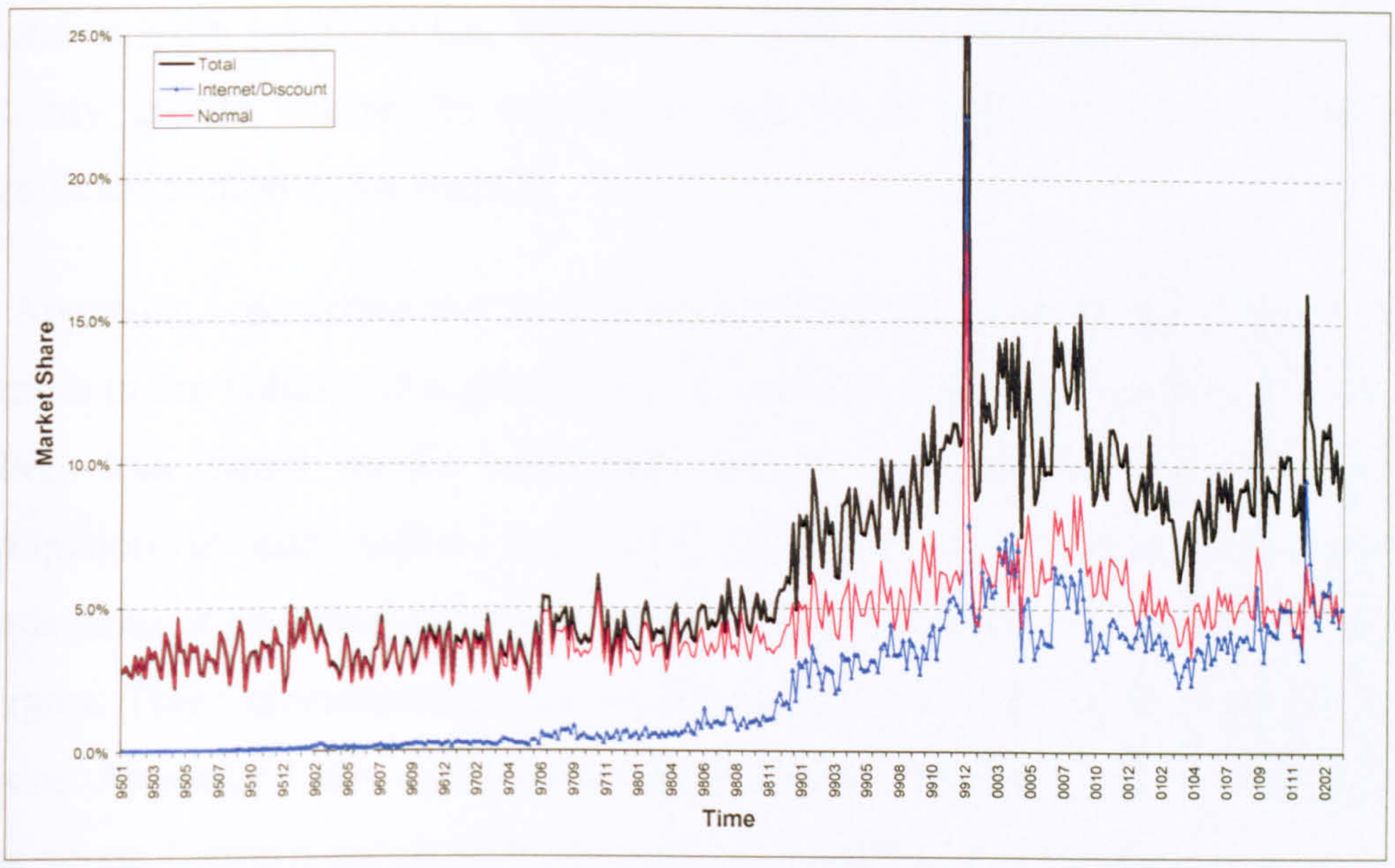


Figure 4.4: Individual Investor Participation for Average Stock in Sample

This plots the combined market share of turnover by value for all retail brokers for the average index and non-index stock in the sample over time. At the end of the sample, the average index stock had 40% of trading (by value) accounted for by individual investors, while the average non-index stock had 70% of trading accounted for by individuals.

This study focuses exclusively on stocks included in the main benchmark index (the All Ordinaries index to 2000 and the ASX300 subsequently). This gives between 250 and 350 stocks in each weekly cross-section. Index stocks capture around 90% of total market capitalisation. Non-index stocks are excluded due to extreme illiquidity and to ensure the results are not being driven by an economically insignificant portion of the market.

Appendix 1 describes the stock characteristics that explain the share of retail investors in the volume of a given stock. I find that retail participation is higher in smaller, less liquid stocks with high recent volatility and news exposure. Participation is also higher in stocks that have listed through government privatisations or demutualisations, and also in stocks that offer 'shareholder discount' programs. These characteristics are used as control variables in the following two sections. Appendix 1 also examines the proportion of trading that takes place within-group versus between-group for both retail and institutional investors.

Table 4.2.A: Breakdown of Australian Equity Ownership by Investor Type as percentage of all listed equity (in $A)

The data is derived from ABS Financial Accounts, ABS Cat. no. 5232.0 using their standard definitions.

Owner Type	Jun-95	Jun-96	Jun-97	Jun-98	Jun-99	Jun-00	Jun-01
Households (direct)	17%	16%	21%	18%	21%	22%	24%
Rest of World	46%	45%	39%	35%	37%	35%	36%
Non-Financial Corporations	2%	3%	3%	2%	2%	2%	2%
Life & Pension Corporations	26%	27%	27%	23%	21%	23%	24%
Financial Intermediaries	6%	6%	8%	11%	7%	10%	8%
Government	2%	2%	1%	10%	12%	7%	4%
Banks	1%	1%	1%	1%	1%	2%	1%
Total Equity Outstanding ($bn)	295,589	336,622	435,160	549,887	635,900	743,523	776,151

Table 4.2.B: International Comparison of Direct Individual Investor Equity Ownership % of all Listed equity

Percentage of Equity Owned Directly by Households (by Market Value)														
Country	1969	1981	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
US	69%	57%	51%	53%	53%	51%	49%	49%	45%	43%	41%	42%		
UK	47%	28%	20%	20%	20%	18%	20%			17%	17%	15%	16%	
Japan	38%	28%	20%	20%	21%	20%	20%	20%	19%	19%	19%	18%	19%	
Germany								21%	20%	19%	18%	17%	17%	
Finland								19%	17%	16%	13%		12%	
Australia								17%	16%	21%	18%	21%	22%	24%

Sources: US: Securities Institute Fact Book (www.sia.com), UK: National Statistics: Shareownership (www.statistics.gov.uk) , Japan: Tokyo Stock Exchange (www.tse org.jp), Germany and Finland: Proshare UK (www.proshare.org) Australia: Australian Bureau of Statistics (www abs gov.au).

4.4. Individual Participation and Stock Volatility

The first prediction from the simple noise trader model is that volatility should *increase* as the proportion of noise traders in a stock increases. This section empirically addresses this hypothesis under the assumption that individual investors are the relevant noise traders.

Previous research is divided on whether increased individual participation increases future return volatility. Friedman (1995) argues that increased concentration of decision making should lead to increases in volatility as *institutional* ownership increased. Sias (1996) finds empirical evidence supporting this argument by demonstrating a relationship between increased institutional ownership and increased future volatility in the US. Dennis and Strickland (2000) find that institutions engage in positive momentum trading on high absolute return days and conclude that institutions contribute to short-term market volatility[88]. Finally, in contemporaneous research, Bennett, Sias, and Starks (2003) find that increases in institutional ownership are associated with higher future firm-specific risk.

Alternatively, several empirical studies find that increases in the shareholder base lead to higher future volatility. Jayaraman, Shastri, and Tandon (1993), and Domowitz, Glen, and Madhavan (1998) show that listing of ADRs leads to an increase in shareholder base and a permanent increase in volatility. Also stock splits lead to an increase in shareholder base and a subsequent increase in volatility (Ohlson and Penman (1985)). Since increases in shareholder base are typically

---

[88] Also related is literature examining the effect of an increased shareholder base on liquidity-driven price volatility. Pagano (1989) predicts that an increase in the number of investors entering a market will lead to improved liquidity and reduced price volatility due to lower transactions related noise (lower market impact for trades). Allen and Gale (1994) extend and modify this argument, showing that an increase in investors with low liquidity preference may actually lead to higher price volatility, implying transactions related volatility depends not only on the number of investors in the market, but their liquidity preferences as well. The prediction from this strand of literature therefore depends on whether increases in individual participation leads to an increase or decrease in stock liquidity.

related to increased individual participation, these results are consistent with increased individual investor participation increasing future volatility. So it is not clear from previous research whether we should expect future volatility to increase when individual investor participation in a stock increases.

To test the prediction of Hypothesis 1, namely that an increase in retail share should lead to higher future volatility, I perform weekly Fama-Macbeth cross-sectional regressions on the universe of all stocks using the following specification.

$$\sigma^i_{t+1:t+12} = \alpha + \beta_1 \sigma^i_{t-12:t} + \beta_2 (RS^i_{t-12:t} - RS^i_{t-25:t-13}) + \beta_4 Liq^i_{t-12:t} + \beta_5 Size^i_t + \beta_6 NEWS^i_{t-12:t} + \beta_7 PRICE^i_t + \sum_{j=8}^{12} \beta_j D^i_j + e^i_t$$

where:

- σ^i_{t+k} = return volatility of stock from week t to $t+k$ for stock i[89].
- RS^i_{t+k} = % of turnover accounted for by retail investors from week t to $t+k$

Volatility next quarter[90] is regressed on volatility in the previous quarter (t-12:t), plus the change in retail participation from quarter (t-25:t-13) to quarter (t-12:t) and a set of ten control variables. The controls used include stock return $R^i_{t-12:t}$, liquidity $Liq^i_{t-12:t}$, and news story counts $NEWS^i_{t-12:t}$, each measured over the previous quarter. I also include end-of-quarter stock price $PRICE^i_t$, and market capitalisation $SIZE^i_t$. Finally I include five corporate event dummies, namely privatisations and demutualisations, IPOs, buybacks, splits, and consolidations to indicate whether a corporate event has taken place in the past three years. For ease of interpretation all

[89] Volatility is calculated as the standard deviation of daily raw stock returns from week t to t+k.

[90] See Sias (1996) for a similarly constructed test to assess the impact of increased institutional ownership on future stock price volatility. Using the change in institutional ownership reduces the dependence of institutional ownership on past volatility. To minimise the potential for data-snooping I use the same quarterly horizon as this previous research does.

variables in the regression are expressed as percentile ranks[91] or dummy variables. T-statistics are calculated using Newey-West standard errors to account for serial correlation in coefficients induced by the use of overlapping periods[92].

The results of the analysis appear in Table 4.3. Several of the control variables are significant. Volatility is highly persistent from one quarter to the next. Additionally, high past returns reduce future volatility consistent with the well-known leverage effect of Black (1976). Large and high-priced stocks have lower future volatility while stocks with high news intensity, share splits, and consolidations have higher future volatility (the effect of splits on future volatility is also found by Lamoureux & Poon (1987) and Kryzanowski & Zhang (1996)).

Finally, turning to the variable we are examining, we find that changes in retail participation are significantly *negatively* related to future volatility (with a t-statistic of -4.24 in the full regression specification). This indicates that increased retail participation leads to *lower* future return volatility.

---

[91] The economic conclusions are not affected when I use levels rather than percentile rankings.

[92] Using non-overlapping periods leads to virtually identical coefficient estimates, however the power of the test is somewhat reduced.

Table 4.3: Effect of Changes in Retail Participation on Future Stock Volatility (Quarterly)

The dependent variable is the volatility of the stock over the next quarter expressed as a percentile rank between zero and one. The independent variables are a) the volatility in the previous quarter b) the raw stock return in the previous quarter c) the change in retail participation from two quarters ago to the previous quarter d) turnover in the previous quarter e) size at the end of the previous quarter f) count of the number of news stories in the previous quarter g) stock price at the end of the previous quarter h) dummies indicating privatisations/demutualisations, ipos, buybacks, splits and consolidations. All x variables are expressed as percentile ranks or a dummy indicator of 0 or 1. The coefficients are estimated using Fama-Macbeth OLS and standard errors are adjusted for heteroskedasticity and autocorrelation using the Newey West adjustment. The period of the regression is from November-1995 to March 2002 and the regression is run on weekly cross-sections.

Panel A: Effect of Changes in Retail Participation on Future Stock Volatility (Quarterly): All Small Investors

Variable	Dependent Variable = Vol(Q)													
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*
Constant	0.50	0.58	0.50	0.70	0.50	0.72	0.50	0.79	0.50	1.77	0.51	3.21	0.51	2.32
Volatility (Q-1)	0.77	72.17	0.74	69.99	0.74	70.30	0.75	69.71	0.71	69.65	0.69	58.21	0.69	58.26
Return (Q-1)			-0.10	-8.68	-0.10	-8.85	-0.09	-8.34	-0.08	-7.08	-0.08	-5.97	-0.08	-5.84
Change in Retail Share of Turnover (Q-2 to Q-1)					**-0.02**	**-3.29**	**-0.02**	**-3.24**	**-0.02**	**-3.17**	**-0.02**	**-3.63**	**-0.02**	**-4.12**
Liquidity (turnover) (Q-1)							-0.01	-1.77	0.01	1.18	0.00	0.10	0.01	0.81
Size									-0.11	-10.77	-0.08	-6.71	-0.08	-6.77
News Count (Q-1)									0.02	2.08	0.03	3.00	0.03	3.00
Stock Price											-0.09	-4.78	-0.09	-4.57
Privatisation/Demutualisation													-0.02	-1.46
IPO (last three years)													0.01	0.73
Buyback (last three years)													-0.01	-1.37
Stock Split (last three years)													0.05	3.53
Consolidation (last three years)													0.04	2.29
Sample (each panel/total)	206	65299	205	65122	205	65122	204	64561	204	64561	203	64381	203	64381
Average R2	58.7%		60.01%		60.2%		60.9%		62.2%		62.9%		64.0%	

To assess the robustness of this result, I re-estimate the regressions in Table 4.3 for large stocks and small stocks separately. The leverage effect is noticeably weaker for larger stocks and quarterly volatility is less persistent. However there is no significant difference between the two groups for the coefficient describing the impact of retail participation on future volatility. For both groups an increase in retail participation leads to lower future volatility. Next I separate my sample into two time periods, 1995-1998 and 1999-2002. I find in both periods that the sign of the coefficient is negative, ie higher retail participation equals lower future volatility. Additionally the effect seems significantly stronger in the second period, inconsistent with the popular media view that small investors were responsible for a significant increase in volatility during the 'internet-bubble' period.

Based on these results, the first hypothesis of our noise trader model applied to individual investors does *not* appear to hold. Instead the result is consistent with the predictions of Friedman (1995) and the results of Sias (1996) who find that increases in institutional participation *increase* future volatility. The results are also consistent with the findings of Odean (1998), Grinblatt and Keloharju (2000), and Jackson (2002) who find that individuals are contrarian investors with respect to recent short-term returns, implying they are likely, if anything, to dampen short-term volatility through their trading actions.

4.5. Individual Participation and Excess Correlation

The second prediction from the simple noise trader model is that increases in the proportion of noise traders in a stock should increase the correlation of the stock with the common sentiment factor. This section empirically tests Hypothesis 2 of the model, assuming that individual investors are the relevant noise traders.

Since the sentiment factor is by its very nature unobservable, I proxy for this factor by constructing a factor mimicking portfolio. A priori, stocks that have the highest proportion of noise traders should have the highest correlation with the common sentiment factor. Thus to construct the mimicking portfolio I first sort stocks into five size quintiles at the start of each quarter, and then within each size group into three retail participation portfolios based on the average proportion of turnover accounted for by individuals over the last quarter. I then form an equally weighted portfolio of the stocks in each of the five high retail participation portfolios and hold this portfolio for the next quarter. The returns on this 'high retail' portfolio (HR), the mimicking portfolio, are taken as the proxy for the common sentiment factor. As a comparison I also construct a 'low retail' portfolio (LR) from the 5 size-sorted low retail participation portfolios and calculate its quarterly returns.

If stock returns were iid and cross-sectionally uncorrelated in the absence of a sentiment factor, then testing Hypothesis 2 would simply involve calculating the correlation of each stock with the mimicking portfolio return and comparing this with the level of retail participation in the stock. In reality however, stocks are correlated for fundamental reasons, for example cash flows of stocks in the same industry generally have a positive association. To examine Hypothesis 2, I regress *changes* in retail share on future correlations. Control variables are also included to adjust for other factors that may affect future correlations.

The regression specification is as follows:

$$\rho^{i,HR}_{t+1:t+12} = \alpha + \beta_1 \rho^{i,HR}_{t-12:t} + \beta_2 (RS^i_{t-12:t} - RS^i_{t-25:t-13}) + \beta_4 Liq^i_{t-12:t} + \beta_5 Size^i_t + \beta_6 NEWS^i_{t-12:t} + \beta_7 PRICE^i_t + \sum_{j=8}^{12} \beta_j D^i_j + e^i_t$$

where $\rho^{i,H}_{t:t+k}$ = correlation of stock i with HR portfolio between week t and $t+k$ [93].

The dependent variable is the correlation between the stock return and the high retail participation portfolio (HR) return. As a comparison, I also run the same regression using the correlation between the stock return and the low retail participation portfolio (LR) as the dependent variable. Correlations are calculated using daily returns for both portfolios over the previous three-month period, and are expressed as a percentile rank[94].

The results of this regression are presented in Table 4.4. Panel A presents the results where the dependent variable is the correlation of each stock with the *high* retail participation portfolio. Panel B presents the results where the dependent variable is the correlation with the *low* retail portfolio.

---

[93] Correlations are calculated using the daily returns of the individual stock and the high retail (HR) portfolio between weeks t and t+k. Dimson betas are used in the robustness checks to ensure the results are not sensitive to this specification.

[94] To ensure the results are not dependent on this specification, the analysis is also run using levels with economically similar results.

Table 4.4: Effect of Changes in Retail Participation on Future Stock Correlations (Quarterly)

The dependent variable is the average daily correlation of the stock with the high retail portfolio (panel A) and the low retail portfolio (panel B) over the next quarter expressed as a percentile rank between zero and one. The independent variables are a) the correlation in the previous quarter b) the volatility in the previous quarter c) the raw stock return in the previous quarter d) the change in retail participation from two quarters ago to the previous quarter e) turnover in the previous quarter f) size at the end of the previous quarter g) count of the number of news stories in the previous quarter h) stock price at the end of the previous quarter i) dummies indicating privatisations/demutualisations, ipos, buybacks, splits and consolidations. All x variables are expressed as percentile ranks or a dummy indicator of 0 or 1. The coefficients are estimated using Fama-Macbeth OLS and standard errors are adjusted for heteroskedasticity and autocorrelation using the Newey West adjustment. The period of the regression is from Nov-1995 to March 2002 and the regression is run on weekly cross-sections. Panel C shows the results of a robustness check for microstructure biases in the calculation of the correlation coefficients.

Panel A: Effect on correlation with HIGH retail portfolio: All Investors

Variable	All Small Investors, Dependent Variable = correlation with HIGH retail portfolio													
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*
Constant	0.00	2.25	0.00	2.22	0.00	2.26	0.00	2.26	0.00	1.81	-0.01	-3.62	-0.01	-3.19
Correlation (Q-1)	0.31	12.76	0.29	12.24	0.31	12.63	0.29	12.23	0.27	11.68	0.16	9.02	0.15	8.73
Volatility (Q-1)			0.00	-0.12			0.00	-0.14	-0.01	-0.44	0.11	4.82	0.10	4.79
Return (Q-1)			0.04	2.31			0.04	2.25	0.04	1.93	-0.01	-0.61	-0.01	-0.36
Change in Retail Share of Turnover (Q-2 to Q-1)					**-0.01**	**-1.20**	**-0.01**	**-0.93**	**0.00**	**-0.11**	**0.01**	**0.56**	**0.01**	**0.81**
Liquidity (turnover) (Q-1)									0.12	9.22	0.05	4.63	0.05	4.69
Size											0.27	6.60	0.26	6.30
News Count (Q-1)											0.04	3.00	0.03	2.17
Stock Price											0.07	3.19	0.08	3.18
Privatisation/Demutualisation													0.08	4.53
IPO (last three years)													0.01	0.52
Buyback (last three years)													-0.01	-0.63
Stock Split (last three years)													0.02	0.73
Consolidation (last three years)													-0.01	-0.35
Sample (each panel/total)	206	65278	205	65112	206	65278	205	65112	204	64560	203	64380	203	64380
Average R2	**11.2%**		**15.24%**		**11.7%**		**15.7%**		**17.7%**		**27.7%**		**29.9%**	

Panel B: Effect on correlation with LOW retail portfolio: All Investors

Variable	All Small Investors, Dependent Variable = correlation with LOW retail portfolio													
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*
Constant	0.50	4.57	0.50	3.32	0.50	4.39	0.50	3.30	0.50	2.47	0.49	-3 24	0.49	-3.63
Correlation with Low retail portfolio (Q-1)	0.30	8.34	0.28	8.81	0.29	8.21	0.28	8.70	0.25	8.37	0.13	7.51	0.14	7.91
Volatility (Q-1)			-0.13	-6.95			-0.12	-6.70	-0.13	-7.31	-0.02	-1.03	-0.02	-1.29
Return (Q-1)			0.05	3.21			0.05	2.94	0.04	2.54	0.00	-0.18	0.00	0.02
Change in Retail Share of Turnover (Q-2 to Q-1)					**-0.07**	**-8.08**	**-0.05**	**-6.18**	**-0.04**	**-5.32**	**-0.03**	**-5.66**	**-0.03**	**-5.27**
Liquidity (turnover) (Q-1)									0.13	7.22	0.08	5.01	0.08	5.09
Size											0.31	8.88	0.31	8.75
News Count (Q-1)											-0.01	-0.75	-0.01	-0.88
Stock Price											0.05	2.10	0.05	2.14
Privatisation/Demutualisation													0.02	1.11
IPO (last three years)													0.02	2.85
Buyback (last three years)													-0.01	-0.67
Stock Split (last three years)													-0.01	-0.42
Consolidation (last three years)													0.07	2.14
Sample Size (each panel/total)	206	65278	205	65112	206	65278	205	65112	204	64560	203	64380	203	64380
Average R2	12.2%		16.87%		13.2%		17.5%		20.2%		29.8%		31.6%	

Panel C: Robustness Check – probability of daily trading for stocks in each size group by retail participation

	Retail Participation Portfolio				
Size	Low	Medium	High	Difference (H-L)	t-stat
1 (smallest)	90.9%	92.0%	91.9%	1.1%	0.081
2	92.7%	94.4%	94.0%	1.3%	1.432
3	94.8%	96.8%	96.8%	2.0%	2.560
4	99.3%	98.1%	97.5%	-1.8%	-2.387
5 (largest)	98.5%	99.4%	98.2%	-0.4%	-0.704

We can see from Panel A that a change in retail participation has *no* significant effect on the correlation of the stock with the high retail portfolio. This finding is contrary to the prediction from the model of noise trader risk, assuming individuals are the relevant noise traders.

Interestingly however, Panel B shows a significant *negative* relationship between the change in retail participation and the correlation of the stock with the *low* retail portfolio. That is, decreases in retail participation lead to an increase in the correlation of the stock with the low retail portfolio. This has an intriguing interpretation. Decreases in retail participation imply increases in institutional participation (due to an adding up constraint) so therefore correlations between stocks and the high institutional (low retail) portfolio are increasing with institutional participation[95]. This is consistent with *institutions* being the source of noise trader risk rather than individuals. Furthermore, the result is consistent with the previous section; if volatility goes down when retail participation increases, it necessarily rises when institutional participation increases[96].

To check these results are robust I re-estimate the two regressions for two subsamples. The results are similar in both the first half of the data (1995-1998) and the second (1999-2002). In both cases the negative relationship between participation and the low retail portfolio is highly significant, and the relationship between participation and the high retail portfolio is negligible[97]. Similarly, estimating the

---

[95] This result is consistent with Barberis, Shliefer, and Wurgler (2002) who find that betas of stocks rise when the stock enters the S&P 500 and also with the results of Pindyck and Rotemberg (1993) who find that residual returns between stocks in different industries are more highly correlated when institutional ownership is higher (for a sample of 42 stocks).

[96] The adding up constraint implies the proportion of trading accounted for by individuals and institutions sums to one. By 'institutions' I mean any non-retail investor eg mutual funds, foreign investors, banks etc. My data does not allow me to separate these various institutional investors.

[97] The coefficient on retail participation for the low portfolio is –0.037 (t-stat –3.71) in the first half of the data and –0.030 (t-stat –3.63) in the second half. For the high portfolio the coefficient is –0.011 (t-stat –0.93) and 0.022 (t-stat 1.90) in the first and second half of the sample respectively.

regression separately for small and large stocks shows that the negative relationship is significant for both groups of stocks, although it is slightly stronger for smaller stocks[98].

As an additional robustness check against potential microstructure biases in the excess correlation tests, the analysis was repeated using Dimson (1979) betas in place of return correlations. The beta of each stock with respect to the HR and LR portfolios was calculated using daily returns, including lead and lag coefficients to capture non-synchronous trading effects[99]. The results of the analysis lead to exactly the same conclusions. Namely, increases in retail share do not increase the beta of the stock with respect to the HR portfolio, while increases in retail share *do* lead to a significant reduction of the beta with respect to the LR portfolio[100].

More generally the analysis revealed that non-synchronous trading was not a particular concern in the data. Table 4.4 Panel C shows the average stock in each of the 15 size/retail portfolios has a daily trading probability exceeding 90%. There is no evidence that high retail stocks have more non-trading days than low-retail stocks as shown in the t-statistics comparing the mean probabilities for each size group. Three of the five size groups reveal no difference in means, size group 3 shows a higher trading probability for high retail stocks while size group 4 shows a lower trading probability for high retail stocks. As such, the excess correlation findings remain robust with respect to these potential microstructure effects.

---

[98] The coefficient on retail participation for the low portfolio is –0.044 (t-stat –4.01) for small stocks (Ex-100) and –0.029 (t-stat –2.44) for large stocks. For the high portfolio the coefficient is 0.005 (t-stat 0.37) and 0.008 (t-stat 0.49) for small and large stocks respectively.

[99] The Dimson betas were calculated using lead and lag returns of 1, 3 and 5 days all leading to similar results. Beta coefficients were aggregated across the lead and lag periods as outlined in Dimson (1979).

[100] In the full specification, the coefficient on retail participation is –0.02 (t-stat –4.24) for the LR beta regression, and –0.005 (t-stat –1.41) for the HR beta regression, ie the results are not affected using this specification.

4.6. Testing for a Noise Trader Risk Factor In Returns

This section more formally tests whether there is any evidence of a priced noise trader risk factor in returns. Lee, Shleifer, and Thaler (1991) show that percentage changes in closed-end-fund discounts are a significant factor in explaining cross-sectional returns, and survive the inclusion of the market portfolio and various macroeconomic factors. This section has a similar goal, ie to assess whether a common sentiment factor exists. However a more direct proxy for the common sentiment factor is used. In addition, rather than simply showing that the betas on the sentiment factor are significant, this section attempts to estimate the size of the risk premium (if any) associated with the sentiment factor[101].

To test the joint hypothesis that there is a priced noise trader risk factor in returns and that retail investors (or institutions) are the noise traders, a zero investment portfolio is constructed. This portfolio is long the portfolio of high retail participation stocks, and short the portfolio of low retail participation stocks discussed in section 4.5. The weekly returns of this portfolio (HRLR) are used to proxy for the noise trader risk factor. To ensure this factor is not simply capturing previously known influences on returns (for example those identified in Fama and French (1992)), several other factors are included in the specification. It is assumed that excess returns (over the risk free rate) are generated from an exact multifactor model:

$$R_{it}^{ex} = \alpha + \beta' f_t + \varepsilon_{it}$$

$$f_t = [R_{mkt,t}^{ex} \quad SMB_t \quad HML_t \quad WML_t \quad HRLR_t]'$$

$$\beta_i = [\beta_{i,MKT} \quad \beta_{i,SMB} \quad \beta_{i,HML} \quad \beta_{i,WML} \quad \beta_{i,HRLR}]'$$

$$\varepsilon_t \sim N(0,\Sigma),\ E(\varepsilon_t \varepsilon_t^T) = \Sigma,\ E[(f-\mu_f)(f-\mu_f)'] = \Omega,\ \mu_f = E(f)$$

---

[101] As noted by Shleifer and Summers (1990) if investor sentiment affects a broad range of assets in the same way, ie it contains a systematic component, then this systematic risk will be priced. They note that "assets subject to the whims of investor sentiment should yield higher average returns than similar assets not subject to such whims".

Each of the factors is a return on a portfolio of traded securities and I assume that a risk free asset exists. The MKT, SMB, and HML factors are constructed in a similar manner to Fama and French (1992) however the earnings to price ratio (EP) is used in place of the book to market ratio due to data availability. The WML (momentum) factor is constructed from a portfolio that is long six-month winners and short six-month losers. Correlations between the factors are presented in Table 4.5.A. This shows that the HRLR factor has relatively low correlations with the other factors. To assess whether the HRLR is a linear combination of the other factors, regressions are run separately for each factor using the remaining factors as explanatory variables and the R2 statistics are presented. The results in Table 4.5.B show the R2 for the regression where the HRLR factor return is the dependent variable is lower than that of all other factors apart from the WML factor. These descriptive statistics imply the HRLR factor is not just a combination of existing known factors in disguise.

Table 4.5: Relationship Between Factor Time Series

This table presents the correlations between the factors used in the pricing model. The HRLR factor is constructed as the returns on a portfolio that is long high retail participation stocks and short low retail participation stocks.

Panel A: Correlation Matrix

	MKT	SMB	HML	WML	HRLR
MKT	1	-0.357	-0.363	-0.120	0.256
SMB		1	-0.208	-0.272	0.253
HML			1	0.112	-0.372
WML				1	-0.284
HRLR					1

Panel B: Linear Dependence between factors

R2 of a linear regression of each factor on other 4 factors and a constant	
MKT	39.88%
SMB	37.85%
HML	29.54%
WML	15.74%
HRLR	25.25%

Four groups of test portfolios are used to assess the pricing performance of the factor model, 10 equally weighted size portfolios, 25 size and E/P portfolios, 10 momentum portfolios, and 15 size and retail participation portfolios.

Since all the factors are portfolio returns, I initially use the single-pass regression approach of Gibbons et al (1989) to assess the pricing ability of the model and the significance of the HRLR factor. Weekly returns over the full period are used to maximise the power of the test to detect significant beta coefficients. The estimated beta coefficients and t-statistics for the four different test portfolio groups are presented in Table 4.6.

Initially to test whether the HRLR factor is important in pricing assets we should test whether the estimated beta coefficients on the HRLR factor are jointly different to zero. This approach is advocated in Kan and Zhang (1999) as a way to avoid including spurious factors in the pricing model. The null hypothesis is that the HRLR factor is insignificant, ie $H0:\beta_{HRLR}=0$. To test this hypothesis, assuming normality of errors, we can use a statistic analogous to the GRS (1989) statistic for testing whether alphas are jointly equal to zero. The equivalent statistic for testing whether the HRLR betas are jointly equal to zero is given as follows:

$$J=\frac{T-N-K}{N}\hat{\sigma}^2_{HRLR}\hat{\beta}_{HRLR}\hat{\Sigma}^{-1}\hat{\beta}_{HRLR}\sim F(N,T-N-K)$$

The distribution of this statistic is derived in Appendix 3. If errors do not conform to the normal iid assumption Mackinlay and Richardson (1991) advocate using a GMM based estimator to assess the joint significance of the alpha coefficients. Again, we can form a similar GMM statistic to test the joint significance of the HRLR betas. The statistic is given as:

$$J_{GMM}=T\hat{\beta}'_{HRLR}\left[R\left[D_T'S_T^{-1}D_T\right]^{-1}R'\right]^{-1}\hat{\beta}_{HRLR}\sim\chi^2_N$$

The results for both tests are presented together with the results of the four test portfolio groups in Table 4.6. We can see that both these statistics strongly reject the

null that $\beta_{HRLR} = 0$. For each group of test portfolios the betas on the HRLR factor are jointly significant at the 1% level. This indicates that the betas are useful in capturing variation in returns not captured by the other factors.

As a specification check the two statistics are also calculated for various combinations of alternative factors. The results are presented in Table 4.7 and show that the HRLR betas remain significant after including them in a market factor model, a three-factor model (MKT, SMB, HML), and a market and momentum model (MKT, WML).

Table 4.6: Test Results for Four-Factor Model

This table presents the results of the factor tests for four groups of test portfolios. Firstly 10 size portfolios, then 10 momentum portfolios, then 25 size/EP portfolios, and finally 15 size/retail participation portfolios. The coefficients for each portfolio are presented with the individual t-statistics. Additionally, the GRS F-test and a GMM-test result are presented, testing if the alphas are jointly equal to zero. Finally a similar F-test and GMM-test statistic is presented for the hypothesis that the betas for the HRLR factor are jointly equal to zero. The period of the regression is January 1995 to March 2002 with weekly returns used for all index stocks.

Table 4.6.A: 10 Size and 10 Momentum Test Portfolios with HRLR factor

Pricing Tests: Weekly returns from Jan 1995 to Mar 2002										
10 Equally Weighted Size Portfolios										
Item	1	2	3	4	5	6	7	8	9	10
constant	0.0006	-0.0011	0.0000	0.0000	0.0018	0.0007	0.0003	-0.0004	0.0010	0.0001
beta mkt	0.78	0.86	0.83	0.84	0.89	0.88	0.89	0.90	0.86	0.97
beta SMB	0.73	0.77	0.76	0.81	0.84	0.48	0.40	0.32	0.14	0.02
beta HML	-0.04	0.21	0.14	0.12	0.04	0.15	0.05	0.21	0.14	0.07
beta WML	-0.73	-0.13	-0.10	-0.04	-0.11	-0.09	-0.03	-0.02	-0.03	-0.02
beta HRLR	0.76	0.22	0.22	0.05	0.12	0.20	0.13	-0.02	0.04	-0.04
t(constant)	0.43	-1.38	0.00	-0.04	3.40	1.00	0.52	-0.68	1.81	0.18
t(beta mkt)	8.49	15.53	19.37	20.79	23.78	17.55	20.34	21.15	22.79	45.82
t(beta SMB)	6.44	11.27	14.40	16.17	18.38	7.75	7.55	6.16	3.00	0.84
t(beta HML)	-0.45	3.51	3.01	2.79	1.06	2.75	0.99	4.60	3.48	2.86
t(beta WML)	-9.46	-2.82	-2.87	-1.18	-3.50	-2.05	-0.78	-0.54	-1.11	-1.09
t(beta HRLR)	7.16	3.40	4.41	0.98	2.82	3.46	2.62	-0.38	0.94	-1.44
R2 (raw)	58%	56%	67%	66%	76%	58%	64%	61%	68%	90%

Tests on Alpha Coefficient	GRS Statistic	1.9998	p(GRS)	0.0325	GMM Stat	17.8813	p(GMM)	0.0570
Tests on b(HRLR) Coefficient	J Statistic	12.1056	p(J)	0.0000	GMM Stat	81.8045	p(GMM)	0.0000

10 Equally Weighted Momentum Portfolios										
Item	1	2	3	4	5	6	7	8	9	10
constant	-0.0014	-0.0011	-0.0003	0.0000	0.0014	0.0005	0.0006	0.0007	0.0006	0.0025
beta mkt	1.02	0.94	0.87	0.83	0.80	0.79	0.75	0.81	0.85	0.95
beta SMB	0.97	0.51	0.44	0.33	0.35	0.38	0.40	0.39	0.37	0.51
beta HML	0.03	0.12	0.13	0.16	0.13	0.27	0.21	0.18	0.14	-0.05
beta WML	-1.12	-0.50	-0.32	-0.23	-0.08	-0.02	0.06	0.11	0.25	0.46
beta HRLR	0.62	0.15	0.11	-0.10	-0.07	0.00	-0.05	-0.04	0.07	0.31
t(constant)	-1.18	-1.32	-0.46	-0.04	2.58	0.88	1.11	1.37	0.98	3.76
t(beta mkt)	12.11	16.54	19.04	19.08	20.67	20.86	18.16	21.37	20.80	20.49
t(beta SMB)	9.31	7.33	7.82	6.25	7.38	8.13	7.95	8.52	7.37	8.92
t(beta HML)	0.36	1.89	2.61	3.50	3.06	6.57	4.66	4.36	3.07	-1.03
t(beta WML)	-15.85	-10.65	-8.48	-6.22	-2.50	-0.74	1.75	3.60	7.22	11.82
t(beta HRLR)	6.34	2.25	2.06	-1.92	-1.60	-0.04	-0.99	-0.88	1.54	5.70
R2 (raw)	74%	66%	67%	61%	62%	60%	52%	60%	62%	70%

Tests on Alpha Coefficient	GRS Statistic	3.2136	p(GRS)	0.0006	GMM Stat	49.9429	p(GMM)	0.0000
Tests on b(HRLR) Coefficient	J Statistic	10.0417	p(J)	0.0000	GMM Stat	45.5534	p(GMM)	0.0000

Table 4.6.B: Test Portfolios: 25 size and EP portfolios, with HRLR factor

Pricing Tests: Weekly returns from Jan 1995 to Mar 2002										
	EP Ratio Quintiles									
Size	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
	constant					t(constant)				
Small	-0.0004	-0.0011	0.0001	-0.0003	0.0016	-0.19	-0.87	0.08	-0.26	1.16
2	0.0020	0.0005	0.0000	-0.0004	0.0007	1.65	0.52	0.01	-0.58	0.78
3	0.0006	0.0009	0.0012	0.0005	0.0012	0.42	0.73	1.42	0.67	1.23
4	-0.0006	-0.0002	-0.0003	-0.0006	0.0005	-0.43	-0.25	-0.35	-0.89	0.65
Big	-0.0010	0.0005	0.0007	0.0006	0.0017	-1.18	0.74	0.99	1.02	2.75
	beta mkt					t(beta mkt)				
Small	0.74	0.91	0.99	0.83	0.99	5.43	9.88	13.01	9.50	10.01
2	1.09	0.88	0.79	0.86	0.92	12.55	13.13	15.49	16.49	14.64
3	0.96	0.94	0.86	0.73	0.76	10.24	10.95	14.83	14.13	11.38
4	0.95	0.96	0.88	0.93	0.87	9.92	15.32	15.25	19.20	16.57
Big	0.95	0.83	0.87	0.93	0.98	15.72	16.58	18.04	21.18	22.66
	beta SMB					t(beta SMB)				
Small	0.61	0.83	0.90	0.80	1.04	3.63	7.38	9.54	7.50	8.60
2	1.24	0.80	0.73	0.71	0.79	11.66	9.64	11.70	11.14	10.25
3	0.71	0.53	0.53	0.34	0.41	6.17	5.07	7.47	5.40	4.96
4	0.46	0.29	0.32	0.29	0.27	3.92	3.79	4.47	4.91	4.15
Big	-0.09	0.11	0.03	0.14	0.04	-1.17	1.86	0.48	2.54	0.66
	beta HML					t(beta HML)				
Small	-0.44	0.08	0.40	0.37	0.50	-2.99	0.77	4.79	3.92	4.72
2	-0.46	0.00	0.25	0.46	0.42	-4.89	-0.03	4.52	8.11	6.13
3	-0.34	0.03	0.17	0.33	0.28	-3.34	0.28	2.73	5.98	3.81
4	-0.15	0.01	0.32	0.35	0.41	-1.40	0.21	5.13	6.65	7.25
Big	-0.45	-0.05	0.06	0.31	0.49	-6.87	-0.86	1.12	6.54	10.40
	beta WML					t(beta WML)				
Small	-0.55	-0.24	-0.12	-0.03	-0.31	-4.76	-3.08	-1.92	-0.44	-3.77
2	-0.09	-0.12	0.05	0.00	-0.19	-1.22	-2.07	1.20	-0.04	-3.56
3	-0.21	-0.13	-0.06	-0.09	-0.19	-2.72	-1.83	-1.14	-2.09	-3.34
4	0.11	0.15	-0.04	-0.06	-0.17	1.35	2.78	-0.89	-1.44	-3.77
Big	0.10	0.00	-0.14	-0.05	-0.13	2.05	0.02	-3.41	-1.24	-3.57
	beta HRLR					t(beta HRLR)				
Small	0.72	0.18	0.01	-0.04	0.32	4.59	1.67	0.11	-0.44	2.81
2	0.29	-0.08	-0.10	-0.08	0.10	2.88	-1.08	-1.65	-1.33	1.33
3	0.64	0.27	0.00	-0.06	0.15	5.85	2.70	0.02	-0.95	1.94
4	0.32	0.05	-0.05	-0.19	-0.05	2.92	0.67	-0.74	-3.48	-0.90
Big	0.11	-0.01	-0.11	-0.07	0.00	1.54	-0.24	-2.06	-1.40	0.05
	R2									
Small	34%	36%	39%	25%	36%					
2	60%	45%	44%	46%	47%					
3	52%	39%	44%	39%	35%					
4	35%	47%	43%	54%	48%					
Big	64%	55%	58%	62%	67%					
	Tests on Alpha Coefficient					Tests on Beta HRLR Coefficient				
	F stat	1.0865				J stat	6.5707			
	Critical 5% value	1.5388				Critical 5% value	1.5388			
	p-value	0.3557				p-value	0.0000			
	GMM alpha stat	32.6705				GMM b(HRLR) stat	141.0686			
	Critical 5% value	37.6525				Critical 5% value	37.6525			
	p-value	0.1395				p-value	0.0000			

Table 4.6.C: 15 Size / Retail Participation Portfolios with HRLR

Pricing Tests: Weekly returns from Jan 1995 to Mar 2002							
Retail Portfolio	**constant**	**beta mkt**	**beta SMB**	**beta HML**	**beta WML**	**beta HRLR**	**R2**
1 (Small/Low Retail)	0.0013	0.95	0.95	-0.12	-0.52	-0.92	41%
2	0.0010	0.71	0.70	0.02	-0.38	0.54	42%
3 (Small/High Retail)	0.0001	0.81	0.61	0.22	-0.45	1.02	54%
4	0.0003	0.79	0.75	0.15	-0.07	-0.30	52%
5	0.0014	0.86	0.70	0.14	-0.26	0.11	55%
6	0.0006	0.79	0.68	0.07	-0.03	0.60	60%
7	0.0013	0.69	0.60	0.03	-0.03	-0.24	45%
8	0.0000	1.00	0.74	0.26	-0.11	0.20	59%
9	0.0011	1.01	0.86	-0.11	-0.15	0.50	73%
10	0.0001	0.84	0.30	0.18	-0.10	-0.36	56%
11	-0.0006	0.89	0.36	0.22	-0.01	0.00	53%
12	-0.0005	0.95	0.47	0.05	-0.07	0.59	68%
13 (large/Low Retail)	0.0000	0.96	0.04	0.05	-0.03	-0.20	78%
14	0.0011	0.95	0.07	0.07	-0.09	0.00	73%
15 (large/High Retail)	0.0007	0.78	0.12	0.20	-0.05	0.22	60%
tstats	**t(constant)**	**t(MKT)**	**t(SMB)**	**t(HML)**	**t(WML)**	**t(HRLR)**	
1 (Small/Low Retail)	0.96	9.84	8.01	-1.13	-6.42	-8.26	
2	0.70	7.27	5.80	0.14	-4.68	4.77	
3 (Small/High Retail)	0.10	8.60	5.21	2.10	-5.73	9.42	
4	0.49	16.77	12.97	2.95	-1.66	-5.59	
5	1.65	14.32	9.50	2.13	-5.14	1.65	
6	0.69	13.02	9.08	1.08	-0.64	8.52	
7	1.83	14.27	10.09	0.64	-0.73	-4.29	
8	0.04	17.31	10.43	4.20	-2.27	2.99	
9	1.31	17.58	12.15	-1.75	-3.15	7.60	
10	0.08	18.98	5.52	3.66	-2.82	-7.06	
11	-0.85	17.82	5.86	4.12	-0.30	0.08	
12	-0.69	17.18	6.95	0.77	-1.53	9.34	
13 (large/Low Retail)	-0.05	28.69	1.03	1.28	-1.17	-5.27	
14	1.99	24.49	1.47	1.78	-2.90	-0.07	
15 (large/High Retail)	1.13	17.93	2.27	4.25	-1.31	4.37	

Tests on Alpha	GRS Stat	1.35	5% Crit Val	1.70	p-value	0.1680
Coefficients	GMM alpha stat	24.79	5% Crit Val	25.0	p-value	0.0529
Tests on Beta HRLR	J stat	1478	5% Crit Val	1.70	p-value	0.0000
Coefficients	GMM b(HRLR) stat	34555	5% Crit Val	25.0	p-value	0.0000

Table 4.7: Significance Tests for HRLR Beta

This table presents test statistics for the joint significance of the HRLR beta factor for various groups of test assets. The value of the F statistic is given by J, together with its critical value and p-value. The value of the GMM statistic is JGMM. Results are presented separately for each group of test portfolios.

Significance tests for HRLR beta factor						
10 size portfolios						
Additional Factors	J	5% crit	p(J)	JGMM	5% crit	p(GMM)
MKT	18.7	1.86	0.000	81.3	18.3	0.000
MKT, SMB, HML	13.0	1.86	0.000	61.2	18.3	0.000
MKT, WML	16.3	1.86	0.000	97.5	18.3	0.000
MKT, SMB, HML, WML	12.1	1.86	0.000	81.8	18.3	0.000
10 momentum portfolios						
Additional Factors	J	5% crit	p(J)	JGMM	5% crit	p(GMM)
MKT	17.9	1.86	0.000	63.3	18.3	0.000
MKT, SMB, HML	11.5	1.86	0.000	46.2	18.3	0.000
MKT, WML	15.2	1.86	0.000	55.3	18.3	0.000
MKT, SMB, HML, WML	10.0	1.86	0.000	45.6	18.3	0.000
25 size/EP portfolios						
Additional Factors	J	5% crit	p(J)	JGMM	5% crit	p(GMM)
MKT	10.0	1.54	0.000	213.3	37.7	0.000
MKT, SMB, HML	7.1	1.54	0.000	164.0	37.7	0.000
MKT, WML	8.9	1.54	0.000	163.0	37.7	0.000
MKT, SMB, HML, WML	6.6	1.54	0.000	141.1	37.7	0.000
15 size/retail portfolios						
Additional Factors	J	5% crit	p(J)	JGMM	5% crit	p(GMM)
MKT	1412.2	1.70	0.000	45735.5	25.0	0.000
MKT, SMB, HML	1481.8	1.70	0.000	34530.0	25.0	0.000
MKT, WML	1406.5	1.70	0.000	47991.9	25.0	0.000
MKT, SMB, HML, WML	1477.7	1.70	0.000	34555.3	25.0	0.000

As a further check that the betas are not significant due to an individual investor preference for the variables that we are sorting on, the results of a beta significance test similar to Elton, Gruber, and Busse (1998) are presented. Firstly all 93 stocks that exist in the index for the full period are selected. The returns of these stocks are regressed against the market factor plus the HRLR factor. The percentage of the HRLR betas that are significant is found to be 27.9% with an F-statistic of 2.93.

To assess whether this result is significant the empirical distribution of these two statistics are found via bootstrapping. The bootstrapping procedure creates a new random factor with the same time series properties as the HRLR factor for each stock. Then a long-short portfolio is constructed by double sorting on size and this random factor, and the returns of the 93 stocks are regressed against this new factor. The percentage of significant betas, and the F-statistic, is calculated for each random factor, and this process is repeated 1000 times. The empirical distribution is presented in Figure 4.5. The random factor has an average of 8.12% of betas significant (with a standard deviation of 3.18%) and an average F-statistic of 1.23 (with a standard deviation of 0.21). From the figure we can see that the HRLR factor (with 27.9% of betas significant and an F-statistic of 2.93) is highly significant, with none of the 1000 draws producing a value as extreme. These results suggest the betas on the HRLR factor are indeed significant.

We next move to the question of whether the HRLR factor is priced and estimate the size of the risk premium associated with the factor. Using the single pass time series approach this test is particularly simple, with the estimate of the risk premium for each factor simply being the time series average of the factor return.

$$\hat{\mu}_k = \frac{1}{T}\sum_{t=1}^{T} f_t$$

Figure 4.5: Empirical Distribution of % Significant Betas and F-statistics

These figures plot the empirical bootstrap distributions of the % betas significant and F-statistic produced using a random factor, against the statistics produced using the HRLR factor for the 93 stocks that are present for the full sample period. The HRLR factor is highly significant for both statistics, lying well outside the most extreme of the 1000 simulated statistics.

Figure 4.5.A: Empirical Distribution of the % betas significant statistic

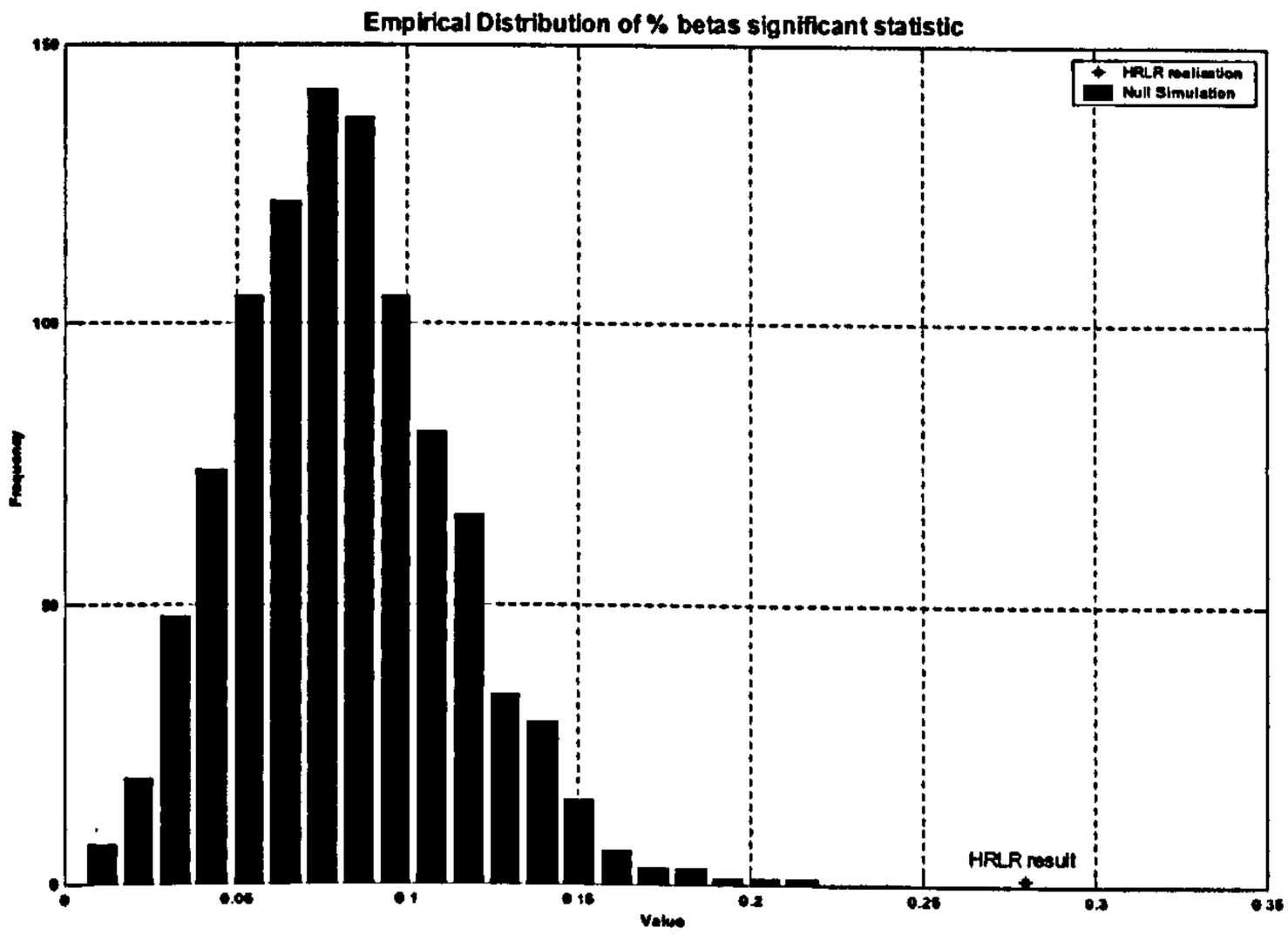


Figure 4.5.B: Empirical Distribution of the F statistic

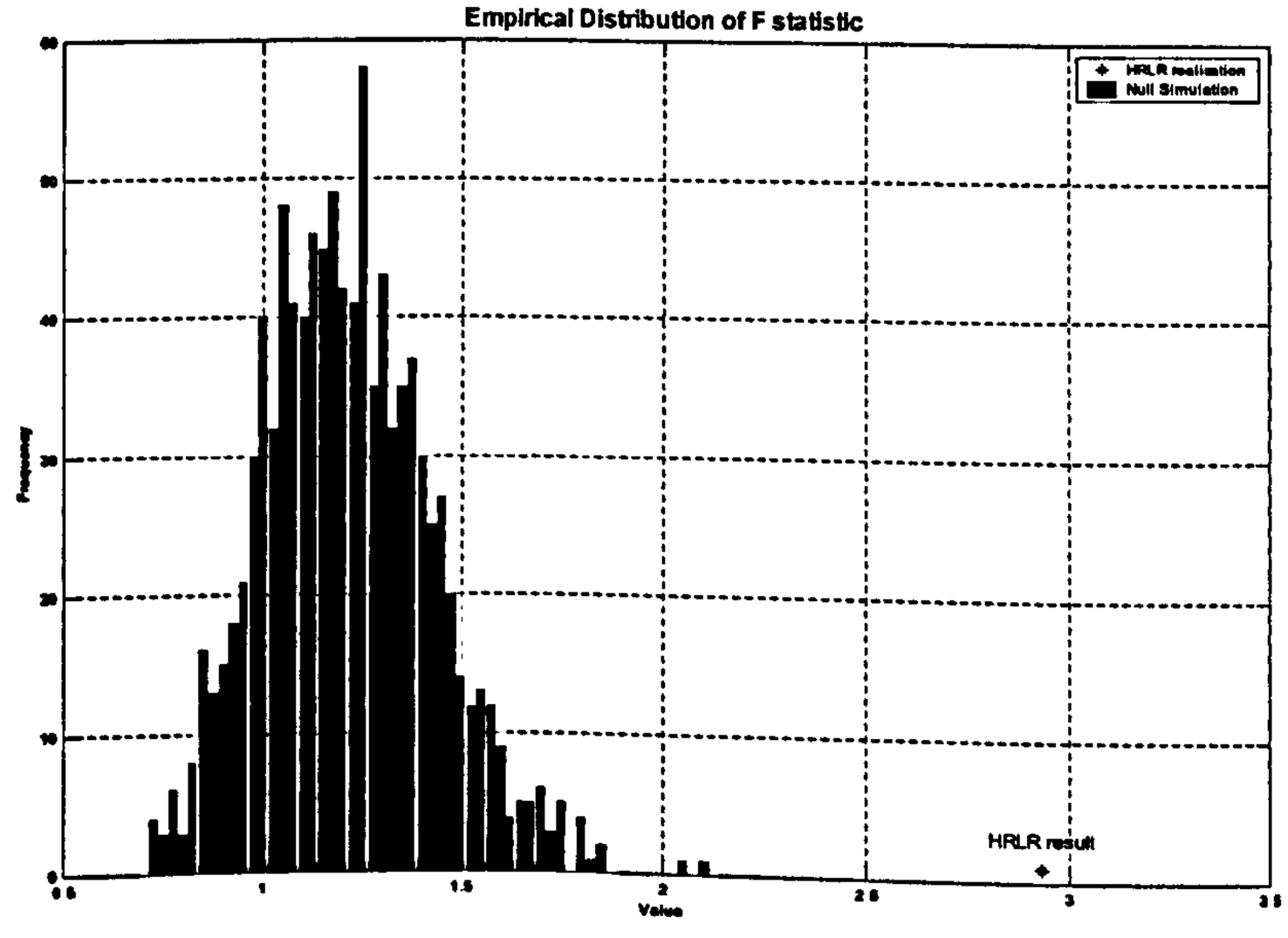


Table 4.8: Time Series Estimate of Factor Risk Premia

This table presents estimates of the factor risk premia for the MKT, SMB, HML, WML and HRLR factors. The significance levels are presented assuming firstly that factor returns are iid, and secondly adjusting for autocorrelated factor returns using Newey West standard errors.

Table 4.8A: Full Period (1995-2002) risk premia estimates

Factor	Annual premium	T-Stat (iid)	T-Stat (Newey-West)
MKT	7.56%	1.65	1.68
SMB	-7.27%	-1.98	-1.75
HML	13.20%	3.40	2.97
WML	21.12%	4.61	3.70
HRLR	-8.70%	-2.47	-2.05

Table 4.8B: Split Period risk premia estimates for Kan and Zhang joint subperiod test

Factor	First Half Risk Premium	First Half T-stat	Joint Test 1% Critical t-value	Second Half Risk Premium	Second Half T-stat	Joint Test 1% Critical t-value
MKT	7.67%	1.26	1.48	7.52%	1.11	1.48
SMB	-4.75%	-1.05	-1.48	-10.88%	-1.89	-1.48
HML	10.41%	2.53	1.48	17.01%	2.58	1.48
WML	14.19%	3.15	1.48	27.33%	3.43	1.48
HRLR	-4.87%	-1.76	-1.48	-12.53%	-1.92	-1.48

Table 4.9: Two Stage Estimates of Factor Risk Premia

This table presents two-stage cross-sectional estimates of the factor risk premia for the MKT, SMB, HML, WML and HRLR factors. The significance levels are presented assuming firstly that factor returns are iid, and secondly adjusting for autocorrelated factor returns using Newey West standard errors.

Table 4.9A: Full Period (1995-2002) risk premia estimates using 10 size portfolios

Factor	Annual premium	T-Stat (iid)
Constant	7.10%	0.55
MKT	5.01%	0.37
SMB	-8.15%	-1.61
HML	1.39%	0.12
HRLR	-7.33%	-0.72

Table 4.9B: Full Period (1995-2002) risk premia estimates using 25 size/EP portfolios

Factor	Annual premium	T-Stat (iid)
Constant	6.37%	0.90
MKT	5.89	0.80
SMB	-11.20%	-2.55
HML	11.61%	2.39
HRLR	-8.64%	-1.63

Table 4.9C: Full Period (1995-2002) risk premia estimates using 15 size/retail portfolios

Factor	Annual premium	T-Stat (iid)
Constant	12.34%	1.62
MKT	3.72%	0.43
SMB	-13.31%	-2.47
HML	11.20%	1.42
HRLR	-9.19%	-2.28

Table 4.9D: Full Period (1995-2002) risk premia estimates using all stocks

Factor	Annual premium	T-Stat (iid)
Constant	14.68%	4.22
MKT	-4.32%	-1.26
SMB	-7.92%	-2.79
HML	12.82%	3.82
HRLR	-6.77%	-2.31

The factor risk premium estimates, together with significance levels assuming both iid factor returns, and autocorrelated factor returns are presented in Table 4.8. For the full period, the HRLR factor is associated with a significantly *negative* risk premium of –8.64% per annum with a t-stat of 2.45 (assuming iid factor returns) and 2.05 (using Newey West standard errors). The sign of the HRLR risk premium is particularly interesting as it confirms the results of section 4.5. That is, there *does* appear to be evidence of a priced noise trader risk factor in the return data. However, due to the negative risk premium, the risk appears to be associated with *institutional participation* rather than with individuals.

As a robustness check Kan and Zhang (1999) advocate using sub-period tests to ensure that the risk premium in different sub-periods is of the same sign. The results of this subperiod test are presented in Table 4.8.B. They show that the risk premium on the HRLR factor is negative in both the first and second half of the data, and the size of the risk premium exceeds the critical 5% values for the joint sub-period test.

As a further robustness check on the estimate of the risk premia, a two-stage cross-sectional factor test is performed. In the first stage, for the 10 size portfolios, estimates of the betas are obtained from time-series regressions of weekly portfolio returns on factor returns over a rolling window of 52 weeks. In the second stage, these beta estimates are used in Fama-Macbeth cross-sectional regressions of weekly returns to estimate the risk premia attached to the various factors. Standard errors for the gamma estimates are calculated using Newey-West standard errors. The results of this test appear in Table 4.9. Again, a negative risk premia (–7.33% pa) is associated with the HRLR risk factor however the significance of the premium is low (t-stat = 0.15) due to the small number of test portfolios (in fact none of the factors are significant for this choice of test portfolios). Repeating this test for alternative test portfolios yield estimates for the HRLR risk premium of -8.64% (t-stat = 1.63) for the 25 size/EP portfolios, –6.77% (t-stat = 2.31) for the 15 size/retail portfolios and -9.19% (t-stat = 2.28) for individual stocks. The size of the estimated risk premia using the two-pass approach compare reasonably well with the original time series

estimate of –8.64%, ensuring the original estimate is not an artefact of the estimation methodology.

In summary, the results of this section indicate that the HRLR factor has highly significant betas in the pricing model for a variety of different specifications and test portfolios. Furthermore the HRLR factor appears to be priced. The sign of the risk premium associated with this factor confirms the finding in the previous section that the factor appears to be associated with *institutional* investors, rather than individual investors.

4.7. Discussion and Interpretation

The result that noise trader risk appears to be generated by institutions rather than individuals is particularly interesting since previous research has generally pointed at individuals as being the likely source of such risk.

At first appearance, it seems illogical that institutions could be noise traders. However, noise trader risk in its most general form refers simply to factors independent of fundamentals that generate correlated demand shocks across a group of investors. These shocks need not be generated by common sentiment shocks, but may be generated by other means such as institutional frictions. These frictions can be entirely rationally based, arising for example from principal-agent problems between institutions and their investors.

Various features of the institutional structure of the investment management industry may lead to correlated demand shocks that are unrelated to stock fundamentals. Next I briefly discuss several possible frictions including style investing, investment universe restrictions, herding, and common investment strategies.

In the 'style investing' model of Barberis and Shleifer (2003), a group of investors (switchers) reallocate money between investment styles based on past

relative performance. This generates excess co-movement between stocks in the same style category in addition to cycles of style performance over time. Barberis and Shleifer (2003) argue that institutional investors, such as pension fund trustees, participate in style investing in order to fulfil their fiduciary obligations to create systematic investment strategies. Additionally a form of 'fiduciary window dressing' is suggested where trustees switch to outperforming styles that are easier to justify ex-post to their principals[102]. We can relate their model to the reduced form noise trader model in the current chapter by replacing our sentiment factor δ with a function of recent past returns. Their model also generates excess correlation and excess volatility predictions consistent with my empirical results. Namely the empirical results in this chapter show there is evidence of excess correlation related to institutional ownership, and evidence of a common factor in returns related to institutional participation and unrelated to fundamentals.

A related explanation of the results in this chapter is due to investment mandates that typically restrict institutions to invest in certain segments of the market, for example value and growth specialist managers, S&P500 index managers, and fixed interest funds required to hold investment grade debt. These restrictions are likely to affect institutional demand in a correlated manner when a stock enters or leaves a given investment universe. Consistent with this interpretation, Barberis, Shleifer, and Wurgler (2002) find evidence that index inclusions (which induce no changes in the fundamentals of a firm) lead to increased co-movement of stocks with market benchmark indices. After a stock enters the S&P500 index its beta with respect to the other S&P500 stocks increases, and after it is excluded from the index its beta falls, despite inclusion and exclusion having no fundamental effect on the company itself.

An alternative explanation for why institutional investors may have non-fundamental demand shocks is due to common investment strategies and/or herding.

---

[102] An interesting demonstration of this effect occurred in the late 1990's in several markets (eg UK and Australia) where 'value' managers lost substantial levels of funds under management after 2-3 years of poor style performance. This led to correlated selling of value stocks by these managers.

Lynch and Musto (2003) argue that investment managers persist with strategies that have recently performed well, and discard strategies (and portfolio managers) that have recently performed poorly. Thus strong performance of a particular investment style may lead to subsequent imitation by other managers. Other sources of common investment strategies may be due to academic research findings (eg small cap research in the early 1980's), or common education (eg managers educated through MBA programs). Common investment strategies can generate correlated non-fundamental shocks to demand[103]. This has the potential to generate excess volatility, excess correlations, and non-fundamental return factors related to institutional participation. Evidence on manager herding also supports this conjecture[104].

Finally, it can be argued that correlated demand shocks are far more likely to emanate from institutions than from atomistic individuals. Friedman (1995) argues that institutional decision-making is highly concentrated which may lead to excess volatility due to random demand factors. It is plausible that a small number of large institutional investors may be subject to a common factor (eg non-informative broker macroeconomic forecasts) or friction (eg asset consultant pressure) that leads their demands to be correlated. Institutions also have high levels of turnover relative to individuals that may amplify their importance in creating such correlated non-fundamental effects. We saw in the data discussion that individuals account on average for less than 10% of total trading volume, and similar results have been found in the UK where individuals account for only 8% of trading volume.

---

[103] Bennett, Sias, and Starks (2003) find evidence that institutional investor preferences exhibit slow moving shifts in time, eg exhibiting a broad shift in preferences from large to small cap stocks over the 1990's.

[104] Both Wermers (1999) and Sias (2002) find evidence of institutional investor herding using US data.

4.8. Conclusion

If individuals generate noise trader risk in equity markets, we *should* find that increases in individual participation in a given stock generates higher volatility and higher correlations between the stock and a portfolio of other high individual participation stocks. Similarly models of noise trader risk predict that if sentiment of such investors is sufficiently correlated and investors are risk averse, then a risk premium will be demanded for exposure to the sentiment factor.

This chapter shows that increases in individual participation lead to *lower* future stock volatility, inconsistent with the noise trader prediction. Similarly there is no evidence of excess correlations being generated by increases in individual participation.

However the chapter *does* find evidence of non-fundamental risk associated with *institutional* participation. Increases in institutional participation increase future volatility and the correlation of a stock's returns with a high institutional participation portfolio. Additionally evidence is found of a significant priced noise trader risk factor in returns. The sign of the risk premium suggests that institutions, not individuals, are the source of noise trader risk in equity markets.

The traditional interpretation of noise trader risk argues that it is driven by irrational correlated sentiment of small investors. However noise trader risk in its more general sense can be generated by *any* factor that leads to non-fundamentally induced demand shocks. A more useful interpretation of noise trader risk may instead be that institutional frictions in the investment management industry lead to correlated demand shocks unrelated to fundamentals. Frictions such as common investment strategies, performance related mutual fund flows, and career concerns can generate such demand shocks, rather than irrational sentiment.

Future research exploring the sources and implications of these frictions is likely to be a fruitful endeavour.

4.9. Appendix 1: Control Variable Discussion

This appendix defines the control variables used throughout the chapter. Additionally the stock characteristics that explain the share of retail turnover in a stock are presented. The data sources for the control variables are described as follows:

- Prices (adjusted and unadjusted), dividends, market capitalisation, and volume traded were obtained from Datastream.

- Index composition was obtained from the ASX Index Almanac. Throughout the chapter 'index stocks' are defined to be those in the All Ordinaries index up to April 2000 and the ASX300 index subsequently[105].

- Newspaper stories were obtained from the Fairfax F2 on-line searchable archive using a web-scraper. This archive covers stories published in the Sydney Morning Herald and the Australian Financial Review over the full period of this study[106]. News counts are obtained by summing the number of stories where a particular stock was mentioned.

- Corporate events were obtained from the SDC Australia/New Zealand database for IPOs, SEOs, buybacks, demutualisations, and privatisations.

- Stock split and consolidation data was collected manually from Bloomberg.

---

[105] On the 3rd April 2000 the construction of the All Ordinaries index was changed significantly, with the number of stocks in the index going from around 266 to 500. The ASX300 post April 2000 is the closest market index to the All Ordinaries prior to the changes. Source: ASX Factbook 2001.

[106] These two papers had a combined readership in 2000 of over 1 million readers on weekdays and 1.5 million on Saturdays (compared a total population of 19.1 million and a labour force of 9.1 million). Source: Newspaper readership statistics were obtained from: http://www.business.vu.edu.au/bho2250/Stats/Circnews_data.htm. Population statistics were obtained from the Australian Bureau of Statistics.

These control variables explain a significant proportion of the variation in the retail share of turnover across stocks (average cross sectional adjusted R squared values of up to 47% are obtained for weekly data). Table 4.10 presents the results of Fama-Macbeth panel regressions with retail share as the dependent variable, and the various control variables discussed above as the explanatory variables.

We can see from Table 4.10 that stock size is the most important variable determining the retail share of volume. Smaller stocks have a significantly higher percentage of trading volume accounted for by individuals. Individuals also account for a higher proportion of volume for stocks with high past six-month volatility, high past news counts, and low past liquidity. Corporate events significantly affect individual trading intensity, with privatised stocks having higher intensity and stocks undertaking buybacks or stock consolidations having lower intensity. Attention related variables such as previous one-week volatility and news count lead to higher participation. Also high priced stocks have lower retail intensity. Finally stocks that have a shareholder discount program have significantly higher retail participation.

An expanded version of this appendix (available on request) relates the various factors discussed in this section to underlying economic sources including high information acquisition costs for individuals relative to institutions, and also institutional frictions such as 5% reporting requirements, and higher liquidity requirements for institutional investors.

Finally, we examine the percentage of trading that is within-group versus between-group. From the model in section 4.2 the percentage of within-group trading is given by:

$$within_group_trading_t = \frac{\frac{\sigma_z}{\sqrt{\pi}}}{b(1-\mu)\mu\left|\delta_t - \delta_{t-1}\right| + \frac{\sigma_z}{\sqrt{\pi}}}$$

Using the data we can estimate this percentage using the net order flow of individuals divided by the total trading volume for different levels of μ. Figure 4.6 shows the empirical distribution plotted against μ (proxied by RS). This estimate is obtained using a linear kernel estimator. Between-group trading is maximised at a level of 32% when $\mu = 0.5$. Figure 4.6.B shows the model prediction for b = 1, δ_t = 0.4 and δ_t = 0.41. The shape of the empirical estimate closely matches the model prediction for these parameters (among others). The implications of this result will be examined more fully in subsequent research.

Figure 4.6: Distribution of Within-Group and Between-Group Trading

The first figure plots the empirical percentage of trading that occurs within-group (within institutions and individual investors respectively) vs between-group. The line is estimated using a linear kernel estimator. The second figure plots the prediction from the simple noise trader risk model for $b = 1$, $\delta = 1$ and $\sigma_z = 1$.

Figure 4.6.A: Empirical Distribution of within and between group trading

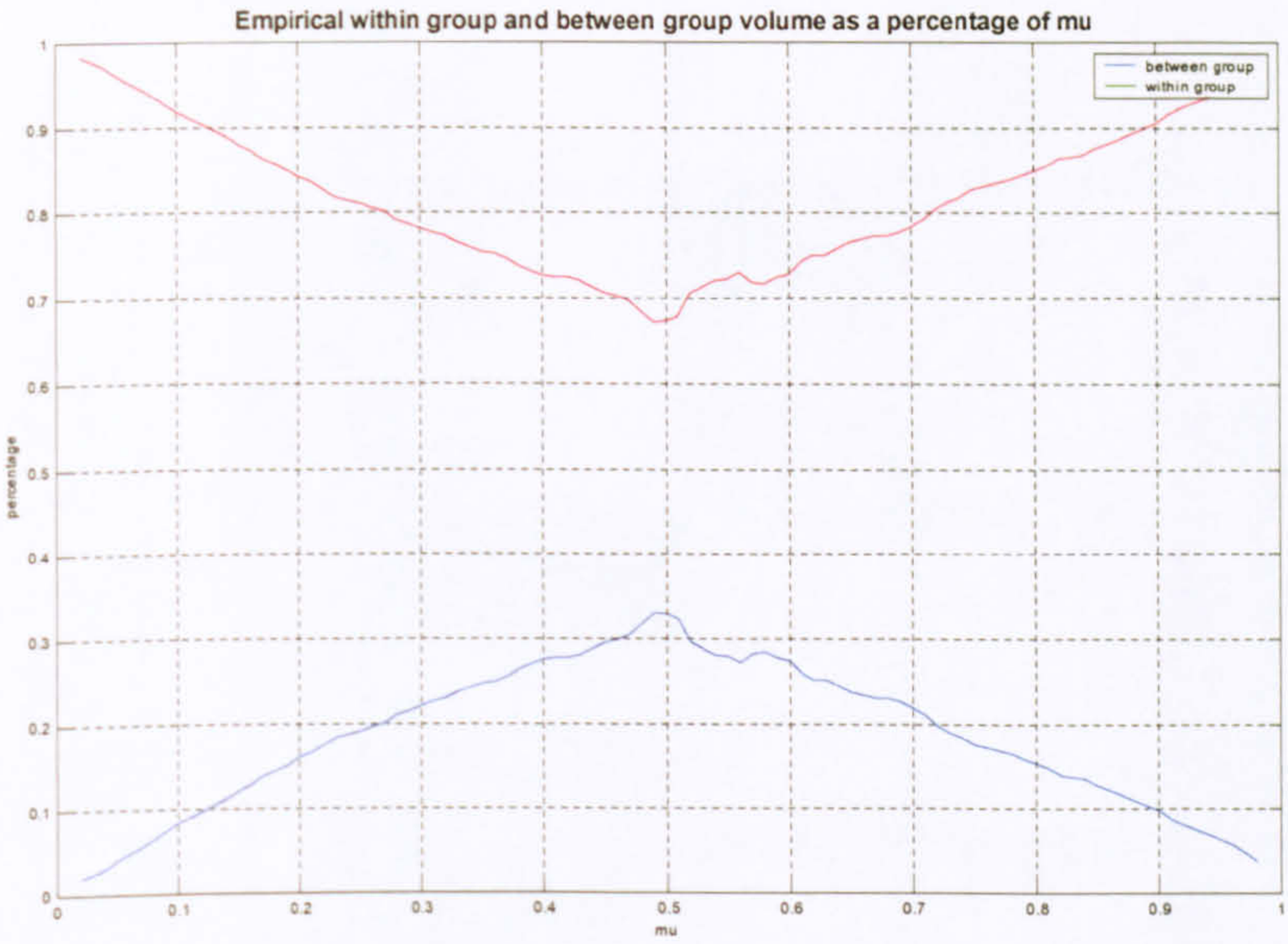


Figure 4.6.B: Model Prediction of within and between group trading

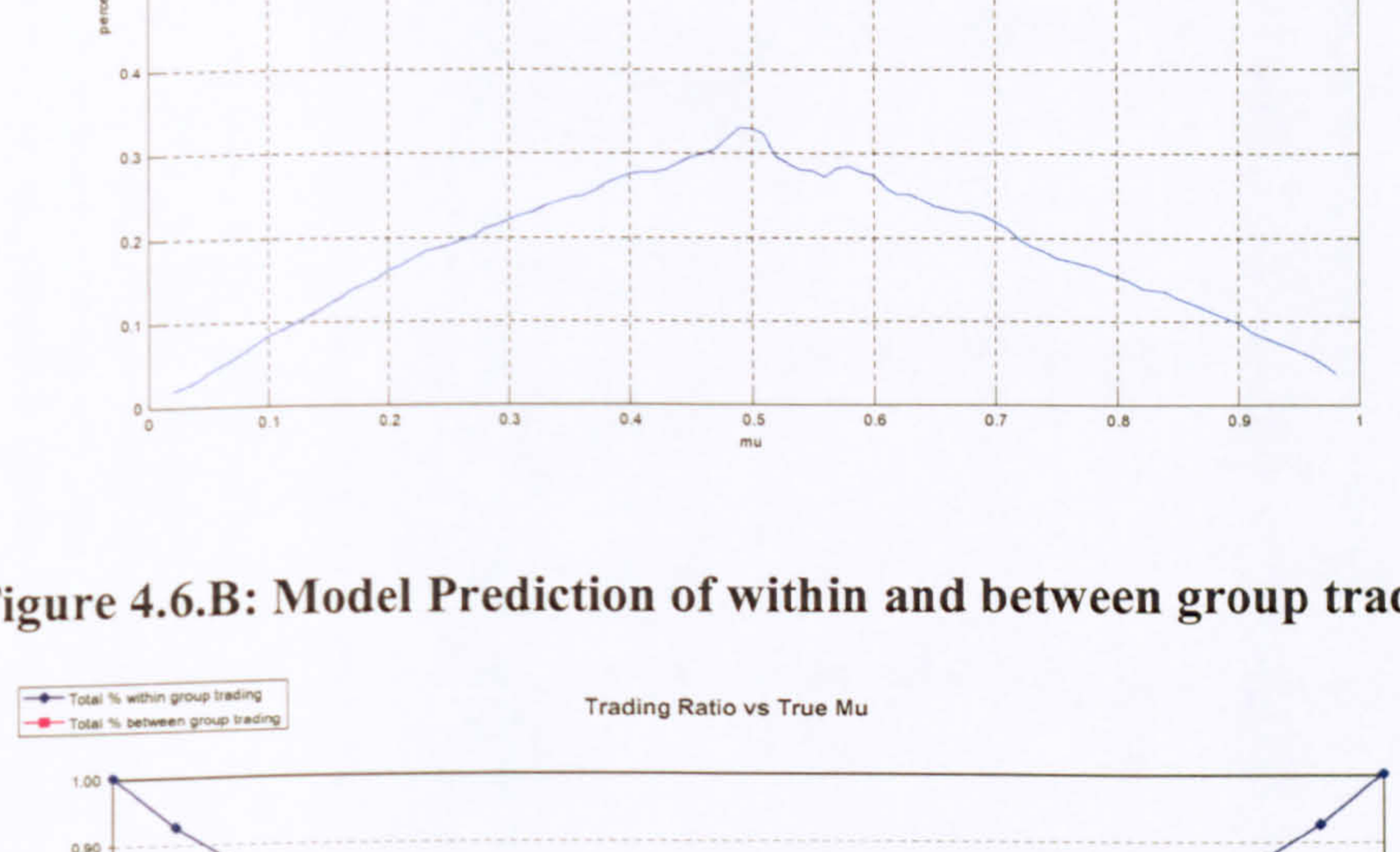


Table 4.10: Characteristics Explaining The % Retail Share of Total Volume Traded – All Small Investors

The dependent variable is the share of retail turnover in each stock expressed as a percentile rank between zero and one. The independent variables are a) size = market value of the firms equity b) count of the number of news stories in the last six months c) the past six months raw return d) past six month volatility e) liquidity (turnover) over past 10 weeks f) dummies indicating privatisations/demutualisations, ipos, buybacks, splits and consolidations g) last week's return, last week's volatility and count of news stories in the previous week e) dividend yield f) stock price g) dummy for shareholder discount program. All x variables are expressed as percentile ranks or a dummy indicator of 0 or 1. The coefficients are estimated using Fama-Macbeth OLS and standard errors are adjusted for heteroskedasticity and autocorrelation using a Newey West Filter. The period of the regression is from November 1995 to March 2002 and the regression is run on weekly cross-sections.

Variable	All Retail Investors													
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Coef.	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*	*Coef.*	*T-stat*
Constant	0.50	27.62	0.50	4.17	0.50	3.93	0.51	4.01	0.50	2.72	0.50	-1.02	0.50	0.85
Size	-0.56	-18.44	-0.52	-19.56	-0.49	-17.10	-0.59	-20.28	-0.60	-16.41	-0.56	-21.32	-0.54	-17.22
Past Six Months News Count			0.06	3.90	0.09	6.32							0.04	2.96
Past six month return			-0.02	-1.37	-0.02	-1.62							-0.01	-1.02
Past six month volatility			0.18	13.31	0.20	13.37							0.17	15.30
Liquidity					-0.14	-15.03							-0.15	-19.48
Privatisation/Demutualisation							0.15	19.39					0.14	16.11
IPO in last three years							-0.06	-5.62					-0.05	-5.51
Buyback in last three years							-0.01	-0.70					0.00	-0.60
Stock Split (last three years)							0.07	4.67					0.01	0.46
Consolidation (last three years)							-0.04	-2.40					-0.08	-5.22
Past one week return									-0.01	-1.36			0.01	1.35
Past one week volatility									0.12	15.47			0.05	10.21
Past one week News Count									0.12	6.41			0.11	5.42
Stock Price											0.00	-0.20	-0.02	-2.16
Dividend Yield											-0.03	-1.53	0.02	1.21
Shareholder Discount											0.05	7.83	0.06	7.78
Sample (each panel/total)	255	84067	209	66963	207	66346	255	84067	212	67752	253	83435	206	65954
Average R2	32.8%		38.9%		46.0%		35.4%		38.2%		34.1%		47.0%	

4.10. Appendix 2: Derivation of Model Propositions

4.10.1 Equilibrium Price

Market clearing implies that aggregate supply equals aggregate demand, namely:

$$1=\mu\left[\lim_{N\to\infty}^{a.s.}\frac{1}{N}\sum_{j=1}^{N}D_t^{i,j,Sent}\right]+(1-\mu)\left[\lim_{M\to\infty}^{a.s.}\frac{1}{M}\sum_{j=1}^{M}D_t^{i,j,Rat}\right]$$

Substituting the demand functions for each type of investor gives:

$$1=\mu(1+bF_t^i+b\delta_t-bP_t^i)+\mu\left[\lim_{N\to\infty}^{a.s.}\frac{1}{N}\sum_{j=1}^{N}z_t^j\right]+(1-\mu)(1+bF_t^i-bP_t^i)+(1-\mu)\left[\lim_{M\to\infty}^{a.s.}\frac{1}{M}\sum_{j=1}^{M}z_t^j\right]$$

Since $E(z_t^j)=0<\infty$ and z_t^j are iid then by Kolmogorov's law of large numbers we have:

$$\lim_{N\to\infty}^{a.s.}\frac{1}{N}\sum_{j=1}^{N}z_t^j=0$$

Therefore the equilibrium price is given as:

$$P_t^i=F_t^i+\mu\delta_t \qquad \text{a.s.}$$

The change in price is similarly given by:

$$P_t^i-P_{t-1}^i=\varepsilon_t^i+\mu_t^i(\delta_t-\delta_{t-1}) \text{ a.s.} \qquad \text{QED.}$$

4.10.2 Hypothesis 1-3 derivation

Hypothesis 1: Given that ε,δ are independent:

$$Var(P_t^i-P_{t-1}^i)=Var(\varepsilon_t^i)+(\mu_t^i)^2Var(\delta_t-\delta_{t-1})$$

$$\frac{\partial Var(P_t^i-P_{t-1}^i)}{\partial\mu_t^i}=2\mu_t^iVar(\delta_t-\delta_{t-1})>0 \qquad \text{QED.}$$

Hypothesis 2:

$$Cov(P_t^i - P_{t-1}^i, \delta_t - \delta_{t-1}) = \mu_t^i Var(\delta_t - \delta_{t-1})$$

$$\frac{\partial Cov(P_t^i - P_{t-1}^i, \delta_t - \delta_{t-1})}{\partial \mu_t^i} = Var(\delta_t - \delta_{t-1}) > 0 \quad \text{QED.}$$

Hypothesis 3:

$$Cov(P_t^i - P_{t-1}^i, P_t^j - P_{t-1}^j) = Cov(\varepsilon_t^i, \varepsilon_t^j) + \mu_t^i \mu_t^j Var(\delta_t - \delta_{t-1})$$

$$\frac{\partial Cov(P_t^i - P_{t-1}^i, P_t^j - P_{t-1}^j)}{\partial \mu_t^i \mu_t^j} = Var(\delta_t - \delta_{t-1}) > 0 \quad \text{QED.}$$

4.10.3 Volume Derivation

To prove that

$$2 * Vol_t = \mu \overset{a.s.}{\lim_{N\to\infty}} \frac{1}{N} \sum_{j=1}^{N} \left| D_t^{i,j,Sent} - D_{t-1}^{i,j,Sent} \right| + (1-\mu) \frac{1}{M} \overset{a.s.}{\lim_{N\to\infty}} \sum_{j=1}^{M} \left| D_t^{i,j,Rat} - D_{t-1}^{i,j,Rat} \right|$$

simplifies to

$$Vol_t = b(1-\mu)\mu \left| \delta_t - \delta_{t-1} \right| + \frac{\mu \sigma_z}{\sqrt{\pi}} + \frac{(1-\mu)\sigma_z}{\sqrt{\pi}}$$

we first note the following lemma.

Lemma C.1: If x~N(0,1), then $E(|X|) = \frac{2}{\sqrt{2\pi}}$.

Proof: $E(|X|) = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = \int_{0}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx + \int_{-\infty}^{0} -x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$

Consider the first integral, making the substitution $z = x^2$ gives the integral:

$$Integral1 = \int_{0}^{\infty} \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{\frac{-z}{2}} dz = \left[\frac{-e^{\frac{-z}{2}}}{2\sqrt{2\pi}\left(\frac{1}{2}\right)} \right]_0^{\infty} = \frac{1}{\sqrt{2\pi}}$$

Consider the second integral, making the substitution $z=-x^2$ gives the integral:

$$Integral2=\int_{-\infty}^{0}\frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{\frac{-z}{2}}dz=\left[\frac{e^{\frac{-z}{2}}}{2\sqrt{2\pi}\left(\frac{1}{2}\right)}\right]_{-\infty}^{0}=\frac{1}{\sqrt{2\pi}}$$

Therefore $E(|X|)=\frac{2}{\sqrt{2\pi}}$. QED.

Using Lemma C.1, we can derive a related result:

Lemma C.2: If x~N$(0,\sigma_x^2)$, then $E(|X|)=\frac{2\sigma}{\sqrt{2\pi}}$

Proof: $E(|X|)=\int_{-\infty}^{\infty}|x|\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}dx=\int_{0}^{\infty}x\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}dx+\int_{-\infty}^{0}-x\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}dx$

Making the substitutions $z=x^2$ and $z=-x^2$ as in the first proof yields the required result. QED.

Next, we can calculate the expected value of the absolute difference between two normal iid random variables as follows:

Lemma C.3: If $x_1,x_2\sim N(0,\sigma_{x_i}^2)$ and x_1 & x_2 are independent then $E(|x_1-x_2|)=\frac{2\sigma_x}{\sqrt{\pi}}$

Proof: Let $z=x_1-x_2$, then $E(z)=0$, $2\sigma_x^2$ and $z\sim N(0,2\sigma_x^2)$. From Lemma C.2 we can conclude that:

$$E(|z|)=\frac{2(\sigma\sqrt{2})}{\sqrt{2\pi}}=\frac{2\sigma}{\sqrt{\pi}}.$$ QED.

Next we combine Lemma C.3 with the Kolmogorov strong law of large numbers to get Lemma 4.

Lemma C.4: Let the aggregate trading volume generated by liquidity trades of the sentiment investors be $ZT_{N,t}^{Sent} = \frac{1}{N}\sum_{j=1}^{N}\left|z_t^j - z_{t-1}^j\right|$ where $z_t^j \sim N(0,\sigma_z^2)$ and iid.

Then $\lim_{N\to\infty}^{a.s.} ZT_{N,t}^{Sent} = E(ZT_{N,t}^{Sent}) = \frac{2\sigma_z}{\sqrt{\pi}}$, ie $ZT_{N,t}^{Sent} = \frac{2\sigma_z}{\sqrt{\pi}}$ almost surely.

Proof: From Lemma C.3 we know that $E\left(\left|z_t^j - z_{t-1}^j\right|\right) = \frac{2\sigma_z}{\sqrt{\pi}} < \infty$. Since the expectation is finite, and the normal random variables are iid, then by Kolmogorov's strong law of large numbers we have that $\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\left|z_t^j - z_{t-1}^j\right| = \frac{2\sigma_z}{\sqrt{\pi}}$. QED.

This result can be applied to both the sentiment and rational investors for the liquidity parts of their trading volume.

Now we can proceed to simplify our original expression:

$$2*Vol_t = \frac{\mu}{N}\sum_{j=1}^{N\to\infty}\left|D_t^{i,j,Sent} - D_{t-1}^{i,j,Sent}\right| + \frac{(1-\mu)}{M}\sum_{j=1}^{M\to\infty}\left|D_t^{i,j,Rat} - D_{t-1}^{i,j,Rat}\right|$$

Substituting in the demand functions gives:

$$2*Vol_t = b(1-\mu)\mu\left|b(1-\mu)(\delta_t - \delta_{t-1})\right| + \frac{\mu}{N}\sum_{j=1}^{N\to\infty}\left|z_t^j - z_{t-1}^j\right| + (1-\mu)\left|b\mu(\delta_t - \delta_{t-1})\right| + \frac{(1-\mu)}{M}\sum_{j=1}^{M\to\infty}\left|z_t^j - z_{t-1}^j\right|$$

$$2*Vol_t = 2*b(1-\mu)\mu\left|\delta_t - \delta_{t-1}\right| + \frac{\mu}{N}\sum_{j=1}^{N\to\infty}\left|z_t^j - z_{t-1}^j\right| + \frac{(1-\mu)}{M}\sum_{j=1}^{M\to\infty}\left|z_t^j - z_{t-1}^j\right|$$

Using the results from Lemma C.4 gives:

$$2*Vol_t = 2*b(1-\mu)\mu\left|\delta_t - \delta_{t-1}\right| + \frac{\mu 2\sigma_z}{\sqrt{\pi}} + \frac{(1-\mu)2\sigma_z}{\sqrt{\pi}}$$

and finally:

$$Vol_t = b(1-\mu)\mu\left|\delta_t - \delta_{t-1}\right| + \frac{\mu\sigma_z}{\sqrt{\pi}} + \frac{(1-\mu)\sigma_z}{\sqrt{\pi}} \quad \text{QED.}$$

4.10.4 Lemma 1 Derivation

$$\frac{\partial RS_t}{\partial \mu} = \frac{0.5b(1-2\mu)|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}}{b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}} - \frac{\left(0.5b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\mu\sigma_z}{\sqrt{\pi}}\right)(1-2\mu)b|\delta_t - \delta_{t-1}|}{\left(b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}\right)^2}$$

$$\frac{\partial RS_t}{\partial \mu} = \frac{\frac{\sigma_z}{\pi^2} + b(1-\mu)\mu|\delta_t - \delta_{t-1}|\frac{\sigma_z}{\sqrt{\pi}} + \frac{\sigma_z}{\sqrt{\pi}}b|\delta_t - \delta_{t-1}|(1-2\mu)(0.5-\mu)}{\left(b(1-\mu)\mu|\delta_t - \delta_{t-1}| + \frac{\sigma_z}{\sqrt{\pi}}\right)^2}$$

The first two terms of the numerator are positive, and the denominator is positive. The third term of the numerator can be written as:

$$\frac{\sigma_z}{\sqrt{\pi}}b|\delta_t - \delta_{t-1}|(1-2\mu)(0.5-\mu) = \frac{\sigma_z}{\sqrt{\pi}}b|\delta_t - \delta_{t-1}|\left(\frac{1}{\sqrt{2}} - \sqrt{2}\mu\right)^2 > 0$$

Therefore, $\frac{\partial RS_t}{\partial \mu} > 0$ for all μ_t^i

Looking at holdings of sentiment investors:

$$\frac{\partial SH}{\partial \mu} = 1 + (1-2\mu)b\delta$$

This expression can be positive or negative depending on the values μ, b, δ. QED.

4.11. Appendix 3: Small Sample Distribution for Beta Test

The model of excess returns is as follows:

$$R_{it}^{ex} = \alpha + \beta' f_t + \varepsilon_{it}$$

$$f_t = [R_{mkt,t}^{ex} \quad SMB_t \quad HML_t \quad HRLR_t]', \ \beta_i = \left[\beta_{i,MKT} \quad \beta_{i,SMB} \quad \beta_{i,HML} \quad \beta_{i,HRLR}\right]'$$

$$\varepsilon_t \sim N(0,\Sigma),\ E(\varepsilon_t\varepsilon_t^T)=\Sigma,\ E\left[(f-\mu_f)(f-\mu_f)'\right]=\Omega,\ \mu_f=E(f)$$

The maximum likelihood estimators are the OLS estimators. The information matrix of the estimators is as follows:

$$I(\alpha,\beta)=T\Sigma^{-1}\otimes\begin{bmatrix}1 & \mu_K\\ \mu_K & \Omega+\mu_K\mu_K'\end{bmatrix}$$

Inverting this gives the covariance matrix for the parameters:

$$V(\alpha,\beta)=\frac{1}{T}\begin{bmatrix}1+\mu_K'\Omega^{-1}\mu_K & -\mu_K\Omega^{-1}\\ -\Omega^{-1}\mu_K & \Omega^{-1}\end{bmatrix}\otimes\Sigma$$

This implies the distribution of the parameters is given as:

$$\hat{\alpha}\sim N\left(\alpha,\frac{1}{T}\left(1+\mu_K'\Omega^{-1}\mu_K\right)\Sigma\right)$$

$$\hat{\beta}\sim N\left(\beta,\frac{1}{T}\Omega^{-1}\otimes\Sigma\right)$$

$$Cov(\hat{\alpha},\hat{\beta})=-\frac{1}{T}\mu_K'\Omega^{-1}\otimes\Sigma$$

To test the null hypothesis that an individual element of β is equal to zero, ie $H0:\beta_i=0$, we can use the theorem of Muirhead (1983) outlined in Campbell, Lo, and Mackinlay (1996 pp193) to show that the finite sample distribution of the following test statistic is:

- $$J=\frac{T-N-K}{N}\hat{\sigma}_{fi}^2\hat{\beta}_i\hat{\Sigma}^{-1}\hat{\beta}_i\sim F(N,T-N-K)$$

5. References

5.1. Chapter 2: Trade Generation, Reputation, and Sell-Side Analysts

Abarbanell, J and Lehavy, R (2003), Biased Forecasts or Biased Earnings? The Role of Reported Earnings in Explaining Apparent Bias and Over/Underreaction in Analysts' Earnings Forecasts, *Working Paper*, Michigan.

Aitken, M, Muthuswamy, J and Wong, K (2001), The Impact of Brokers' Recommendations: Australian Evidence, *Pacific Basin Journal of Finance*, forthcoming.

Almazan, A, Brown, K, Carlson, M and Chapman, D (2003), Why Constrain you mutual fund manager, *Journal of Financial Economics*, forthcoming.

Anderson, T, Bollerslev, T, Diebold, F, Labys, P (2003), Modeling and Forecasting Realized Volatility, *Working Paper*, Duke University.

Barber B, Lehavy R, McNichols M and Trueman B (2001), Can Investors Profit from the Prophets? Consensus Analyst Recommendations and Stock Returns, *Journal of Finance*, 56(2), 531-563.

Barber B, Lehavy R, McNichols M and Trueman B (2003), Reassessing the returns to analysts' stock recommendations, *Financial Analysts Journal*, 59(2), 88-96.

Benabou, R and Laroque, G (1992), Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility, *Quarterly Journal of Economics*, 107(3), 921-958.

Boni, L and Womack, K (2002), Solving the Sell-Side Research Problem: Insights from Buy-Side Professionals, *Working Paper*, Tuck School of Business.

Brown, L (1991), Forecast Selection When all Forecasts Are Not Equally Recent, *International Journal of Forecasting*, 7(3), 349-356.

Brown, L (1993), Earnings Forecasting Research: Its Implications for Capital Markets Research, *International Journal of Forecasting*, 9, 295–320.

Brown, L, Griffin, P, Hagerman, R and Zmijewski, M (1987), An Evaluation of Alternative Proxies for Market's Assessment of Unexpected Earnings, *Journal of Accounting and Economics*, 11, 159–93.

Chen, Q, Francis, J and Jiang, W (2003), Investor Learning About Analyst Predictive Ability, *Working Paper*, Columbia.

Clement, M (1999), Analyst Forecast Accuracy: Do ability, resources, and portfolio complexity matter?, *Journal of Accounting and Economics*, 27, 285-303.

Cooper, R, Day, T and Lewis, C (2001), Following the leader: a study of individual analysts' earnings forecasts, *Journal of Financial Economics*, 61, 383-416.

D'Avolio, G (2002), The Market for Borrowing Stock, *Journal of Financial Economics*, 66, 271-306.

Das, S, Levine, C, Sivaramakrishnan, K (1998), Earnings predictability and bias in analysts' earnings forecasts, *Accounting Review*, 73, 277-294.

DeJong, D, Forsythe, R, Lundholm, R and Uecker, W (1985), A Laboratory Investigation of the Moral Hazard Problem in an Agency Relationship, *Journal of Accounting Research*, 23, 81-120.

Diamond, D (1989), Reputation Acquisition in Debt Markets, *Journal of Political Economy*, 97(4), 828-862.

Dimson, E and Marsh, P (1984), An Analysis of Brokers' and Analysts' Unpublished Forecasts of UK Stock Returns, *Journal of Finance*, 39(5), 1257-1292.

Dugar, A and Nathan, S (1995), The Effect of Investment Banking Relationships on Financial Analysts' Earnings Investment Recommendations, *Contemporary Accounting Research*, 12, 131–60.

Eames, M, Glover, S and Kennedy, J (2002), The association between trading recommendations and broker-analysts' earnings forecasts, *Journal of Accounting Research*, 40(1), 85-104.

Eaton, J. and Gersovitz, M (1981), Debt with Potential Repudiation: Theoretical and Empirical Analysis, *Review of Economic Studies*, 48, 289-309.

Ellis, K, Michaely, R and O'Hara, M (2000), When the Underwriter is the Market Maker: An Examination of Trading in the IPO Aftermarket, *Journal of Finance*, 60(3), 1039-1074.

Ellis, K, Michaely, R and O'Hara, M (2002), The Making of a Dealer Market: From Entry to Equilibrium in the Trading of Nasdaq Stocks, *Journal of Finance*, 57(5), 2289-2316.

Fama, E (1980), Agency Problems and the Theory of the Firm, *Journal of Political Economy*, 88(2), 288-307.

Francis, J and Philbrick, D (1993), Analysts' Decisions as Products of a Multi-task Environment, *Journal of Accounting Research*, 31(2), 216-230.

Gomes, A (2000), Going Public without Governance: Managerial Reputation Effects, *Journal of Finance*, 55(2), 615-646.

Gu, Z and Wu, J (2003), Earnings Skewness and Analyst Forecast Bias, *Journal of Accounting and Economics*, forthcoming.

Hayes, R (1998), The impact of trading commission incentives on analysts' stock coverage decisions and earnings forecasts, *Journal of Accounting Research*, 36(2), 299-320.

Hayward, M and Boeker, W (1998), Power and Conflicts of Interest in Professional Firms: Evidence from Investment Banking, *Administrative Science Quarterly*, 43(1), 1-22.

Holmstrom, B (1982), Managerial Incentive Problems: A Dynamic Perspective, Reprinted in: *Review of Economic Studies* (1999), 66, 169-182.

Hong, H and Kubik, J (2003), Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts, *Journal of Finance*, 58(1), 313-351.

Irvine, P (2001), Do analysts generate trade for their firms? Evidence from the Toronto Stock exchange, *Journal of Accounting and Economics*, 30, 209-226.

Irvine, P (2003), Analysts' Forecasts and Brokerage-Firm Volume, *Working Paper*, Emory.

Jackson, A (2003), Noise Trader Risk Exists...but the Noise Traders are not who you think they are, *Working Paper*, London Business School.

John, K and Nachman, D (1985), Risky Debt, Investment Incentives, and Reputation in a Sequential Equilibrium, *Journal of Finance*, 40(3), 863-878.

Johnson, T (2003), Forecast Dispersion and the Cross-Section of Expected Returns, *Working Paper,* London Business School.

Kyle, A (1985), Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-1335.

Keane, M and Runkle, D (1998), Are Financial Analysts' Forecasts of Corporate Profits Rational?, *Journal of Political Economy*, 106(4), 768-805.

Khorana, A and Servaes, H (2002), An Examination of Competition in the Mutual Fund Industry, *Working Paper,* London Business School.

Kreps, D and Wilson, R (1982), Reputation and Imperfect Information, *Journal of Economic Theory*, 27, 253-279.

Lamont, O (2003), Going down fighting: Short-Sellers vs Firms, *Working Paper*, Chicago.

Leach, C (1992), Repetition, Reputation, and Raiding, *Review of Financial Studies*, 5(4), 685-708.

Leone, A and Wu, J (2002), What Does it Take to Become a Superstar? Evidence from Institutional Investor Rankings of Financial Analysts, *Working Paper*, Rochester.

Lim, T (2001), Rationality and Analysts' Forecast Bias, *Journal of Finance*, 56(1), 369-385.

Lin, H and McNichols, M (1998), Underwriting Relationships, analysts' earnings forecasts and investment recommendations, *Journal of Accounting and Economics*, 25, 101-127.

Lys, T and Sohn, S (1990), The Association Between Revisions of Analysts' Earnings Forecasts and Security-Price Changes, *Journal of Accounting and Economics*, 13, 341–63.

Michaely, R and Womack, K (1999), Conflict of Interest and the Credibility of Underwriter Analyst Recommendations, *Review of Financial Studies*, 12(4), 653-686.

Milgrom, P and Roberts, J (1982), Predation, Reputation, and Entry Deterrence, *Journal of Economic Theory*, 27, 280-312.

Morgan, J and Stocken, P (2003), An Analysis of Stock Recommendations, *RAND Journal of Economics*, 34(1), 183-203.

O'Brien, P (1988), Analysts' Forecasts as Earnings Expectations, *Journal of Accounting and Economics*, 10, 53–83

Ottaviani, M and Squintani, F (2002), Non-Fully Strategic Information Transmission, *Working Paper*, London Business School.

Rajan, R and Servaes, H (1997), Analyst Following of Initial Public Offerings, *Journal of Finance*, 52(2), 507-529.

Scharfstein, D and Stein, J (1990), Herd Behaviour and Investment, *American Economic Review*, 80(3), 465-479.

Spitzer, E (2002), United States Senate Committee on commerce, science and technology, Testimony of New York state attorney general Eliot Spitzer, URL: http://www.senate.gov/~commerce/hearings/062602spitzer.pdf

Stickel, S (1992), Reputation and performance among security analysts, *Journal of Finance*, 47(5), 1811-1836.

Trueman, B (1994), Analyst Forecasts and Herding Behaviour, *Review of Financial Studies*, 7(1), 97-124.

Wall Street Journal (2002), Wall Street Has an Unlikely New Cop: Spitzer --- State Office, Used to Policing Junk Mail, Finds Fertile Ground in Stock Research, 25-April-2002.

White, H (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48, 817–830.

Williams, R (2000), A note on robust variance estimation for cluster-correlated data, *Biometrics*, 56, 645–646.

Womack, K (1996), Do Brokerage Analysts' Recommendations Have Investment Value?, *Journal of Finance*, 51(1), 137-167

Zitzewitz, E (2001), Opinion-producing agents: career concerns and exaggeration, *Working Paper*, Stanford GSB.

5.2. Chapter 3: The Aggregate Behaviour of Small Investors

Aitken, M, Kua, A, Brown, P, Walter, T and Izan, H (1995), An Intraday Analysis of the Probability of Trading on the ASX at the Asking Price, *Australian Journal of Management*, 20(2), 115-154.

Baker, M and Stein, J (2002), Market Liquidity as a Sentiment Indicator, *NBER Working Paper*, No. 8816.

Barber, B and Odean, T (2000), Trading Is Hazardous To Your Wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance*, 55(2) 773-806.

Barber, B and Odean, T (2002), All That Glitters: The Effect of Attention and News on the Buying Behaviour of Individual and Institutional Investors, *Working Paper*, University of California.

Barber, B, Odean, T and Zhu, N (2003), Systematic Noise, *Working Paper*, University of California.

Barberis, N and Thaler, R (2001), A Survey of Behavioural Finance, *Working Paper*, University of Chicago.

Benartzi, S (2001), Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock?, *Journal of Finance*, 56(5), 1747-1764.

Choe, H, Kho, B and Stulz, R (2000), Do domestic investors have more valuable information about individual stocks than foreign investors, *NBER Working Paper.*

Cohen, R (1999), Asset Allocation Decisions of Individuals and Institutions, *Working Paper,* Harvard University.

Conrad, J, Gultekin, M and Kaul, G (1997), The Profitability of Short-Term Contrarian Strategies: Implications for Market Efficiency, *Journal of Business and Economic Statistics*, 15(3), 379-386.

De Long, J, Shleifer, A, Summers, L and Waldmann, R (1990), Noise trader risk in financial markets, *Journal of Political Economy*, 98, 703-738.

Dhar, R and Kumar, A (2001), A Non-Random Walk Down the Main Street: Impact of Price Trends on Trading Decisions of Individual Investors, *Working Paper*, Yale.

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

Enders, W (1995), *Applied Econometric Time Series*, Wiley: New York.

Epps, T and Epps, M (1976), The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, *Econometrica*, 44, 305-321.

Fong, K, Madhavan, A and Swan, P (2001), Why do Markets Fragment? A Panel-Data Analysis of Off-Exchange Trading, *Working Paper*, University of Sydney.

Froot, K, O'Connell, P and Seasholes, M (2001), The portfolio flows of international investors, *Journal of Financial Economics*, 59, 151-193.

Gallant, R, Rossi, P and Tauchen, G (1992), Stock Prices and Volume, *Review of Financial Studies*, 5(2), 199-242.

Griffin, J, Harris, J and Topaloglu, S (2003), The Dynamics of Institutional and Individual Trading, *Journal of Finance*, forthcoming.

Grinblatt, M and Keloharju, M (2000), The investment behaviour and performance of various investor types: a study of Finland's unique data set, *Journal of Financial Economics*, 55, 43-67.

Grinblatt, M and Keloharju, M (2001), What Makes Investors Trade, *Journal of Finance*, 56(2) 589-616.

Hamilton, J (1994), *Time Series Analysis*, Princeton, New Jersey.

Hong, H and Stein, J (1999), A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance*, 54(6), 2143-2184.

Jegadeesh, N and Titman, S (1993), Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48(1), 65-91.

Kahneman, D and Tversky, A (1982), *Judgement under Uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge.

Karpoff, J (1987), The Relation between Price Changes and Trading Volume: A Survey, *Journal of Financial and Quantitative Analysis*, 22(1), 109-126

Kelly, M (1997), Do Noise Traders Influence Stock Prices?, *Journal of Money, Credit and Banking*, 29(3), 351-363

Lee, C, Shleifer, A and Thaler, R (1991), Investor sentiment and the closed-end fund puzzle, *Journal of Finance*, 46, 75-110.

Lo, A and Wang, J (2000), Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory, *Review of Financial Studies*, 13(2), 257-300.

Newey, W and West, K (1987), A Simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

Odean, T (1998), Are investors reluctant to realise their losses?, *Journal of Finance*, 53(5), 1775-1798.

Pan, J and Poteshman, A (2003), The Information in Option Volume for Stock Prices, *Working Paper,* MIT No.4275-03.

Poteshman, A and Serbin, V (2003), Clearly Irrational Financial Market Behavior: Evidence from the Early Exercise of Exchange Traded Stock Options, *Journal of Finance*, 58 (1), 37-70.

Ranguelova, E (2001), Disposition Effect and Firm Size: New Evidence on Individual Investor Trading Activity, *Working Paper*, Harvard.

Ritter, J (1988), The Buying and Selling Behaviour of Individual Investors at the Turn of the Year, *Journal of Finance*, 43(3), 701-717.

Shapira, Z and Venezia, I (2000), Patterns of behaviour of professionally managed and independent investors, *Journal of Banking and Finance*, 25(8), 1573-1587.

Shefrin, H and Statman, M (1985), The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance*, 40, 777-790.

Shiller, R (1998), Human Behaviour and the Efficiency of the Financial System, *Working Paper*, Cowles Foundation.

Shleifer, A (2000), *Inefficient Markets*, Oxford: Oxford University Press.

Sims, C (1980), Macroeconomics and Reality, *Econometrica*, 48, 1-49.

Vuolteenaho, T (2002), What Drives Firm Level Stock Returns, *Journal of Finance*, 57(1), 233-264.

Warther, V (1995), Aggregate Mutual Fund Flows and Security Returns, *Journal of Financial Economics*, 39, 209-235.

5.3. Chapter 4: Noise Trader Risk

Allen, F and Gale, D (1994), Limited Market Participation and Volatility of Asset Prices, *American Economic Review*, 84(4) 933-955.

ASX (2001), Share Ownership Survey. Available from: http://www.asx.com.au/

Barber, B and Odean, T (2000), Trading Is Hazardous To Your Wealth: The Common Stock Investment Performance of Individual Investors, *Journal of Finance,* 55(2) 773-806.

Barber, B and Odean, T (2002), All that glitters: The effect of attention and news on the buying behaviour of individual and institutional investors, *Working Paper*, UC Berkeley.

Barberis, N and Shleifer, A (2002), Style Investing, *Journal of Financial Economics*, forthcoming.

Barberis, N, Shleifer, A and Wurgler, J (2002), Comovement, *NBER Working Paper* No.8895.

Bennett, J, Sias, R and Starks, L (2003), Greener Pastures and the Impact of Dynamic Institutional Preferences, *Review of Financial Studies*, forthcoming.

Black, F (1976), Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association,* 177-181.

Black, F (1986), Noise, *Journal of Finance,* 41(3), 529-543.

Blume, M and Friend, I (1975), The Asset Structure of Individual Portfolios and Some Implications for Utility Functions, *Journal of Finance*, 30(2), 585-603.

Brennan, M (1995), The Individual Investor, *Journal of Financial Research*, 18(1), 59-74.

Brown, S, Goetzmann, W, Shiraishi, N, Watanabe, M (2002), Investor Sentiment in Japanese and US Daily Mutual Fund Flows, *Working Paper*, Yale.

Campbell, J, Lo, A, Mackinlay, A (1996), The Econometrics of Financial Markets, Princeton: Princeton University Press.

Chen, N, Kan, R and Miller, M (1993), Are the Discounts on Closed-End Funds a Sentiment Index?, *Journal of Finance*, 48(2), 795-800.

De Long, J, Shleifer, A, Summers, L and Waldmann, R (1989), The Size and Incidence of the Losses from Noise Trading, *Journal of Finance*, 44(3), 681-696

De Long, J, Shleifer, A, Summers, H and Waldmann, R (1990), Noise trader risk in financial markets, *Journal of Political Economy*, 98, 703-738.

De Long, B, Shleifer, A, Summers, L and Waldmann, R (1991), The Survival of Noise Traders in Financial Markets, *Journal of Business*, 64(1), 1-19.

Dennis, P and Strickland, D (2003), Who Blinks in Volatile Markets, Individuals or Institutions, *Journal of Finance*, 57(5), 1923-1950.

Dimson, E (1979), Risk Measurement When Shares are Subject to Infrequent Trading, *Journal of Financial Economics*, 7, 197-226.

Domowitz, I., J. Glen and A. Madhavan (1998), International Cross-Listing and Order Flow Migration: Evidence from an Emerging Market, *Journal of Finance*, 53(6), 2001-27.

Elton, E, Gruber, M, Busse, J (1998), Do Investors Care about Sentiment?, *Journal of Business*, 71(4), 477-500.

Fama, E and French, K (1992), The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47(2), 427-465.

Friedman, B (1995), Economic Implications of Changing Share Ownership, *Journal of Portfolio Management*, 2(3), 59-70.

Froman, L (1934), The increase in the number of stockholders in American corporations since 1929, *Journal of Political Economy*, 42(6), 783-787.

Gemmill, G and Thomas, D (2002), Noise-Trading, Costly Arbitrage, and Asset Prices: Evidence from Closed End Funds, *Journal of Finance*, 57(6), 2571-2594.

Gibbons, M, Ross, S and Shanken, J (1989), A Test of the Efficiency of a Given Portfolio, *Econometrica*, 57(5), 1121-1152.

Grinblatt, M and Keloharju, M (2000), The investment behaviour and performance of various investor types: a study of Finland's unique data set, *Journal of Financial Economics*, 55, 43-67.

Grinblatt, M and Han, B (2002), The Disposition Effect and Momentum, *Journal of Financial Economics*, forthcoming.

Jayaraman, N, Shastri, K and Tandon, K (1993), The Impact of International Cross Listings on Risk and Return: The Evidence from American Depository Receipts, *Journal of Banking and Finance*, 17(1), 91-104.

Jackson, A (2002), The Aggregate Behaviour of Individual Investors, *Working Paper*, London Business School.

Kan, R and Zhang, C (1999), Two-Pass Tests of Asset Pricing Models with Useless Factors, *Journal of Finance*, 54(1), 203-235.

Kryzanowski, L and Zhang, H (1996), Trading Patterns of Small and Large Traders around Stock Split Ex-Dates, *Journal of Financial Research*, 29(1), 75-90.

Kumar, A and Lee, C (2003), Individual Investor Sentiment and Comovement in Small Stock Returns, *Working Paper*, Cornell.

Lamoureux, C and Poon, P (1987), "The Market Reaction to Stock Splits", *Journal of Finance*, 42(5), 1347-1370.

Lee, C, Shleifer, A and Thaler, R (1991), Investor Sentiment and the Closed-End Fund Puzzle, *Journal of Finance*, 46(1), 75-109.

London Stock Exchange (2000), Transaction Survey. Available from: http://www.londonstockexchange.com.

Lynch, A and Musto, D (2003), How Investors Interpret Past Fund Returns, *Journal of Finance*, forthcoming.

Mackinlay, C and Richardson, M (1991), Using Generalized Method of Moments to test Mean-Variance Efficiency, *Journal of Finance*, 46(2), 511-526.

Muirhead, R (1983), Aspects of Multivariate Statistical Theory, New York: John Wiley and Sons.

Odean, T (1998), Are investors reluctant to realise their losses?, *Journal of Finance*, 53(5), 1775-1798.

Ohlson, J and Penman, S (1985), Volatility Increases Subsequent to Stock Splits: An Empirical Aberration, *Journal of Financial Economics*, 14(2), 251-266.

Pagano, M (1989), Endogenous Market Thinness and Stock Price Volatility, *Review of Economic Studies*, 56, 269-288.

Pindyck, R and Rotemberg, J (1993), The Comovement of Stock Prices, *Quarterly Journal of Economics*, 108(4), 1073-1104.

Polkovnichenko, V (2002), Household Portfolio Diversification, *Working Paper*, University of Minnesota.

Poteshman, A and Serbin, V (2001), Clearly Irrational Financial Market Behavior: Evidence from the Early Exercise of Exchange Traded Stock Options, *Journal of Finance*, 58(1), 37-71.

Shleifer, A (2000), *Inefficient Markets: An Introduction to Behavioural Finance*, Clarendon Lectures in Economics, Oxford University Press.

Shleifer, A and Summers, L (1990), The Noise Trader Approach to Finance, *Journal of Economic Perspectives*, 4(2), 19-33.

Sias, R (1996), Volatility and the Institutional Investor, *Financial Analyst Journal*, 52(2), 13-20.

Sias, R (2002), Institutional Herding, *Working Paper*, Washington State.

Sias, R, Starks, L and Tinic, S (2001), Is Noise Trader Risk Priced?, *Journal of Financial Research*, 24(3), 311-329.

Warshow, H (1923), The Distribution of Corporate Ownership in the United States, *Quarterly Journal of Economics*, 39(1), 15-38.

West, K (1988), Bubbles, Fads and Stock Price Volatility Tests: A Partial Evaluation, *Journal of Finance*, 43(3) 639-656.

Wermers, R (1999), Mutual Fund Herding and the Impact on Stock Prices, *Journal of Finance*, 54(2), 581-622.