Sentiment order imbalance and co-movement

Chelley-Steeley, Patricia

DOI:
10.1111/eufm.12146

Sentiment, Order Imbalance and Co-movement. An Examination of Shocks to Retail and Institutional Trading Activity.

ABSTRACT

Using order flow imbalance as a measure of sentiment we show that positive and negative shocks to sentiment captured by the Smooth Transition Conditional Correlation GARCH model (STCC GARCH model) lead to lower co-movement between portfolio and market returns in the post-shock period. We find an asymmetry is present as positive shocks to sentiment have less impact on co-movement changes than negative shocks. We also find that shocks to retail sentiment and the sentiment of two types of institutional investors leads to a reduction in co-movement. Positive shocks to institutional order flow imbalance lead to smaller reductions in co-movement than associated with retail shocks. These effects exist even after we control for firm specific and market-wide news.

Keywords: Order flow shock and sentiment, co-movement, smooth transition model
JEL: G12, G14

Sentiment, Order Imbalance and Co-movement. An Examination of Shocks to Retail and Institutional Trading Activity[1].

Introduction

Compelling evidence shows that the mood or sentiment of the stock market can play an important role in influencing the extent to which individual stock returns co-move with the rest of the market. A pioneering paper by Kumar and Lee (2006) suggests that order flow imbalance (buying activity relative to selling activity) is a useful measure of sentiment. This connection exists because an optimistic mood encourages more buying activity and less selling activity while a pessimistic mood encourages more selling and less buying activity. Kumar and Lee (2006) use a measure of order flow imbalance adjusted for common factors to capture sentiment and show that sentiment increases co-movement.

A premise of Kumar and Lee (2006) is that market sentiment is captured by the sentiment of retail investors because these investors do not have access to the information resources that institutions have so they make more irrational investment decisions. However, there is growing evidence to suggest that institutional investors also act irrationally and are influenced by sentiment, see for example Brown and Cliff (2005), Barberis et al (2005), Bagnoli et al (2009) and DeVault, Sias and Starks (2016). Institutional investors become sentiment traders because they are influenced by "reputational trading" which encourages institutions to trade in the direction of sentiment to avoid their performance standing out from the average while the consistent short term predictability of sentiment strategies and the impediments to corrective low cost arbitrage cause institutions to take advantage of short term predictability driven by sentiment because it is profitable. As a result the order flow of institutional investors reflect sentiment. Moreover, recent analysis of the order flow of retail and institutional investors by DeVault, Sias and Starks (2016) shows that institutions are more driven by sentiment than retail investors. We are therefore motivated to extend the analysis of

[1] We would like to thank an anonymous referee for the generosity of their time and comments which have allowed us to improve the paper considerably.

Kumar and Lee (2006) and consider whether the sentiment of retail *and* institutional investors can influence co-movement.

One of the key results presented in Kumar and Lee (2006) shows that the portfolio returns of small firms is positively influenced by the order flow imbalance or sentiment of retail investors. They do not find that there is a relationship between the portfolio returns of larger firms and retail sentiment. Large institutional investors invest more heavily in large firms while retail investors are more concentrated in small firms. This suggests that for large firms to reflect investor sentiment the measure of sentiment may need to be broadened to include institutional sentiment as these investors are more likely to have an influence over large firms. Moreover, the results of Kumar and Lee (2006) do not provide direct and conclusive evidence of the relationship between sentiment and co-movement only that sentiment influences returns and therefore indirectly must influence co-movement.

Using an adaptation of the Lee and Radhakrishna (2000) algorithm we identify the trades associated with three types of traders. Small trades are classified as retail trades, medium trades are classified as institutional informed trades or stealth trades and the largest trades are classified as large institutional trades. For each of these three investor groups we use the traders daily order flow imbalance to capture their sentiment. For each type of investor this is measured as a ratio of their dollar value of buyer initiated trades to seller initiated trades adjusted for common factors. We also measure the total sentiment of the market using all trades. Over the period we examine we find that sentiment to the different investor groups on average across all stocks is optimistic but varies over time.

Empirically, we find that firm characteristics are also important in determining sentiment as some firm characteristics are associated with average order imbalances greater than unity which suggests greater optimistic sentiment. These characteristics, are firms that are S&P500 index constituents, larger, more liquid, have low book-to-market ratios, have higher prices, are older firms, have higher levels of institutional ownership, have lower earnings-to-price ratios and higher earnings growth because on average over time buying pressure for these stocks out weighs selling pressure. Large institutional traders on average display positive sentiment across all stock characteristics we examine, but have the greatest positive sentiment for smaller firms, stocks with low illiquidity and firms with high prices and earnings growth. Retail investors have pessimistic sentiment for small stocks, low priced stocks and young stocks but the most positive sentiment for high priced stocks, firms with low earnings per share ratios and firms with high earnings growth. Stealth traders generally have positive sentiment for each firm characteristic which is especially strong for large firms, and stocks with, high prices, high institutional ownership, low EPS and high earnings growth.

We calculate the average pairwise correlation between changes in the order flow imbalance of one stock and another to identify whether changes to order flow are systematically correlated, an indication that trading decisions are coordinated and therefore influenced by sentiment. Our results suggest that changes to the overall order flow imbalances across all stocks are not highly correlated but changes amongst retail investors appear to be correlated. We also find that firm characteristics influence the average pairwise correlations and contribute to higher levels of co-ordination amongst firms with some shared characteristics. Multivariate analysis identifies which characteristics have an independent influence over correlations.

We next examine whether the sentiment level of the investor group is related to portfolio excess returns. This is similar to some previous analysis undertaken by Kumar and Lee (2006). We extend their analysis by regressing portfolio returns in excess of the risk free rate against a set of market factors and the sentiment of our distinctive types of investors. We find that total sentiment, retail and large institutional sentiment is correlated with portfolio returns even after controlling for market risk factors. This confirms that the sentiment of retail investors influences co-movement but also shows that the sentiment of institutional investors matters.

To our knowledge, we are the first paper to comprehensively examine the impact that shocks to sentiment, measured as order flow imbalance shocks have on co-movement. The study of shocks rather than levels offers a number of advantages. First, shocks capture unanticipated changes in order flow imbalance so reflect the element of order flow imbalance or sentiment that represents an innovation or change in behavior so may have a different impact on returns and co-movement to order flow imbalance levels. We also show that order flow imbalance levels are non-stationary but changes to order flow imbalance are stationary which provides an additional motivation to focus on shocks. Within the paper we consider two types of shocks. The first is the change in the order flow imbalance level between consecutive periods. This is a useful measure as it adequately captures the concept of a shock and can be easily calculated.

Using this concept of shock we estimate the average pairwise correlation between the market excess return and changes in sentiment for each of the investor groups. This analysis is similar to some that Kumar and Lee (2006) undertook but we extend their analysis to also include the sentiment of stealth and institutional traders. We find that the change in total order flow imbalance and retail order flow imbalance are negatively correlated with changes in the market excess return. This suggests that changes in order flow imbalance are associated with reduced co-movement and motivates us to examine the impact of order flow shocks on stock market correlation further.

A weakness of the analysis undertaken by Kumar and Lee (2006) is that a direct link between co-movement and order flow imbalance is not established. This motivates us to examine a second type of shock to order flow imbalance which is derived from the Silvennoinen and Teräsvirta (2005) and Berben and Jansen (2005) Smooth Transition Conditional Correlation GARCH (STCC GARCH model). Estimation of this model allows us to capture the conditional return correlation between individual portfolio returns and the market portfolio in periods prior and subsequent to an order flow imbalance shock. This allows us to establish within the confines of the model how sentiment shocks influence conditional correlation and therefore co-movement. Within this framework we are also able to control for the effects of firm specific and market-wide information on returns and order flow imbalance to ensure we control for the effects of new information on pre and post shock estimates of conditional correlation.

A particular advantage of the way shocks are identified in this model, as noted by Forbes and Rigobon (2002) and Boyer et al (2008) is that a self-selection bias is avoided because shocks to order flow imbalance are identified endogenously. Moreover, the model can capture non-linearities allowing us to examine separately within the context of the model how positive and negative shocks influence correlations. This attribute is important as Hong, Lim and Stein (2000), Vuolteenaho (2002), Chan (2003) and Kothari, Lewellen, and Warner (2006) found that investors respond asymmetrically to good and bad news. Moreover, Gemmill (1996), Keim and Madhavan (1996), Conrad, Johnson and Wahal (2001) have all shown that increased buying pressure leads to a permanent appreciation in price but an increase in selling pressure does not. This suggests that positive and negative changes to sentiment may have diverse effects on correlation. Another advantage of the STCC GARCH model is that shocks are only identified as shocks if they are higher than the threshold, this makes these shocks less noisy that changes per se as they will not reflect minor variations in order flow imbalance.

Our estimation of the smooth transition model provides direct evidence that shocks to order flow imbalance lead to a reduction in co-movement as average post shock conditional correlations between portfolios and the market are lower than in the pre-shock period. We find that in general positive shocks (shocks that are optimistic) lead to smaller average reductions in post shock conditional correlations than negative shocks (shocks that are pessimistic). We find that when pre-shock sentiment is optimistic positive shocks lead to larger reductions in co-movement than negative shocks but when pre-shock sentiment is already pessimistic negative shocks cause larger changes to conditional correlation than positive shocks. We find that positive shocks have similar impacts on conditional correlation changes for portfolios that comprise only of S&P500 index constituents when compared to those drawn from non S&P500 index constituents but negative shocks lead to larger changes in conditional correlation for portfolios of non S&P500 constituents. When we compare the effects of shocks to the sentiment of the three investor groups we find that shocks to the order flow of retail investors lead to greater declines in post shock conditional correlations or co-movement than is associated with stealth or large traders. The smallest post shock changes in conditional correlation are associated with large institutional investors.

In a final piece of analysis we examine the impact that firm characteristics have on order flow changes. We do this to examine whether changes(shocks) to order flow and therefore the sentiment of different investors is influenced by the characteristics of the firm. We find that positive and negative changes to total order flow imbalance are only influenced by earnings per share values and how long a firm has been listed on the exchange.

The remainder of this paper is as follows. Section 2 provides a review of the literature on how sentiment and information influences retail and institutional investors. Section 3 describes in detail the data used. Section 4 describes some summary statistics and provides some preliminary analysis. Section 5 describes the smooth transition model. Section 6 provides the results of the smooth transition model. Section 7 examines the relationship between sentiment changes and firm characteristics, Section 8 discusses some sub-period analysis and explores the estimation of the smooth transition model over different return horizons. Section 9 offers some conclusions to the paper.

2. Information, Sentiment and Institutional and Retail Order Flow.

In this section we discuss how information and sentiment influences stock returns. We also explore the evidence that seeks to establish whether retail and institutional traders are information or sentiment traders.

2.1 Order Flow Imbalance and Information

Two distinct theoretical approaches have been developed to describe rational price formation in the market microstructure literature and how price changes are connected to order flow. These distinct approaches are the Grossman-Stiglitz (1980) noisy rational expectations equilibrium framework (REEF) and the Baysian-Nash approach of the Kyle model (1985). In the Grossman-Stiglitz REEF model informed and uninformed traders participate in a market for a single security. Informed traders observe a noisy signal about the end of period asset value while uninformed traders learn about information from market prices which are determined by a convex combination of the demands of the informed and uninformed traders. An implied auctioneer sets prices so that overall excess demand and supply is eliminated and the market clears. Within this rational expectations framework prices reflect new information conveyed through the order flow.

A distinctive feature of price formation in the Kyle (1985) model is that informed traders can act strategically which is not possible within the rational expectations framework. The Kyle model describes a market in which noise traders (uninformed traders), informed traders with an information signal and market makers (the role of the auctioneer is made explicit through the actions of the market maker) participate. The market maker observes the net order flow (difference between submitted buy trades and sell trades) but cannot ascertain whether informed or noise traders have placed orders. The order flow information is therefore a noisy signal of the stock's value and is used by the market maker to form a conditional expectation of the stock's value. Price changes determined by the market maker are a linear function of the net order flow so reflect the perceived information content of the order flow[2].

[2] More recently, using the Kyle (1985) framework Odders-White and Ready (2008) derive a model which allow the information content of the order flow imbalance to be broken down into two types of information i.e. information that is private and information that is publically available.

In the Hasbrouck (1991) VAR model it is shown that signed *unexpected* order flow (surprise or shock) has a large and permanent impact on returns[3]. This suggests that it is the unexpected component of signed order flow that leads to price impacts. The role of unexpected changes in order flow imbalance on stock returns is also an issue examined by Chordia et al (2015) who finds that both the level and the shock to order flow imbalance strongly predict the cross-section of expected stock returns[4].

Usually theoretical market microstructure models such as those that we have discussed document the relationship between transaction returns and the order flow at transaction time *t.* Order flow imbalance connects the transaction returns and the net order flow imbalance associated with the transaction to returns and order flow measured over a specified interval such as 15 minutes, one hour, one day etc. This is the case as the order flow imbalance measured over intervals such as a day is the cumulative net order flow of transactions over that interval and the returns over that interval are the cumulative transaction returns for that interval. If there is a relationship between transaction returns and order flow within an interval then there must be a relationship between the order flow imbalance of the interval and the returns measured over the interval.

A range of papers have attempted to gauge the strength of the connection between the level of order flow imbalance and stock returns (price impact) although no single measure of order flow imbalance has been employed. Blume et al (1989) uses order flow imbalance levels to show that there is a strong connection between order flow imbalance and stock returns during the stock market crash of 1987. Cushing and Madhavan (2000) examine order imbalance levels at the end of the trading day and find that order imbalance

[3] Hasbrouck constructs a bi-variate VAR model to explain transaction returns and signed transactions (transactions that have a minus sign if they were seller initiated but a positive sign if they were buyer initiated).
[4] Chordia et al (2015) measures the shock to order flow imbalance (measured as buyer initiated shares traded less seller initiated shares traded scaled by total number of shares traded and as total buy trades less sell trades scaled by all trades. The standard deviation of these constructed measures is then calculated on a monthly basis from daily values. The shock is measured as the difference between the month t standard deviation of order flow imbalance and a moving average of the standard deviation of order flow imbalance from month t-1.

is strongly correlated with the overnight and the next days return. Using a longitudinal sample Chordia and Subrahmanyam (2004) show that stock returns in excess of the market return are explained by the stock's order flow imbalance level[5].

2.2 Retail and Institutional Information Trading

A connection between stock returns and the level of order flow imbalance or changes in order flow imbalance for institutional and retail investors has been used to suggest that the order flow of both institutions and retail investors contain information. Klemkosky (1977) showed that quarterly institutional trading imbalances[6] are positively associated with contemporaneous abnormal returns. Kraus and Stoll (1972) and Mikkelson and Partch (1985) show that block trades initiated by institutional traders lead to significant price changes as the market absorbs the change in order flow while Griffin, Harris and Topaloglu (2003) find a strong contemporaneous relationship between changes in institutional ownership and daily stock returns. Recent evidence that specific types of institutional investors are informed is provided by Ülkü and Weber (2013) who shows that the trading activity of investment trusts, securities firms, pension funds and merchants lead to price impacts so must contain information. This contrasts with the trading activity of private funds and banks which do not contain information.

Although traditionally it is institutional investors that are assumed to be information traders because they have economies of scale in information acquisition more recently evidence has emerged to suggest that the order flow of retail investors may also contain information about future returns. Kaniel et al (2008) show that changes made to the order flow of individual investors is correlated with future short term returns so that the most heavily purchased stocks considerably outperform the most heavily sold stocks the following month.

[5] Blume et al (1989) uses as a measure of order flow imbalance the dollar value of buy orders less the dollar value of sell orders, Madhavan and Cushing (2000) use buyer initiated volume less seller initiated volume scaled by overall volume. Chordia and Subrahmanyam (2004) use the number of buyer initiated trades less the number of seller initiated trades scaled by the number of trades as well as the volume of buyer initiated trades less the number of seller initiated trades scaled by volume.

[6] Klemkosky (1977) uses as a measure of order flow imbalance the difference between the dollar value buying and selling activity of the ten largest trades identified in surveys of investment companies.

Kelley and Tetlock (2013) find that the limit and market order imbalances of retail investors are positively associated with future returns at horizons of up to 20 days. Moreover, they show that imbalances in market orders predict the tone of news stories indicating that these types of orders contain information[7].

2.3 Sentiment and Order Flow

An alternative view of how returns are connected to order flow is provided by the sentiment view of order flow imbalance which suggests that sentiment encourages coordinated trading activity in response to the market mood. An optimistic market mood encourages more buying activity and less selling activity encouraging an order flow imbalance in favor of buying while a pessimistic mood encourages more selling activity and less buying activity and therefore a balance in favor of selling. One of the advantages of capturing sentiment in this way is that analysis of order flow shows what investors have actually done through their trading rather than trying to predict what they might do from investor surveys or confidence indexes. Moreover, the relationship between sentiment indexes and order flow may not be synchronous. Order flow imbalance captures when investors actually respond to sentiment changes but a change in sentiment does not necessarily lead to changes in trading behavior unless the change in sentiment is believed to last. Some changes to sentiment indexes may be transitory which does not motivate investors to change their portfolios. The study of order flow imbalance focuses on the components of sentiment that matter.

Shleifer and Vishney (1997) argue that rational arbitragers do not force prices to intrinsic values immediately because of impediments to the process of arbitrage caused by trading costs and short selling failures[8] which allows the influence of sentiment to persist. As a result stock prices can persistently move away from their fundamental value and the correlations between order flow and stock returns become inflated. In this context, order flow imbalance is a reflection of the optimism or pessimism in the market and does not reflect

[7] Order imbalances are measured by Kelley and Tetlock (2010) as the number of shares bought less the number sold scaled by buys and sells.

[8] Constraints associated with short selling make it impossible to arbitrage away the position of sentiment driven traders. Some institutions for example can not take short positions; Shleifer and Vishny (1997) argue that some institutional arbitragers avoid volatile arbitrage positions.

information. This link between order flow imbalance and sentiment is exploited by Kumar and Lee (2006) who suggest that the level of order flow imbalance associated with retail investors, the investor group they believe best reflects sentiment, will act as an approximation to the amount of sentiment present in the market.

Baker and Wurgler (2007) suggest that sentiment does not affect all stocks equally. A rise in optimistic sentiment will cause some investors to reduce their investments in interest bearing securities and accounts and invest more heavily in the stock market while some investors will be encouraged to invest in more risky stock market investments than they held previously. The effects of these changes should be arbitraged away. However, because there are limits to the arbitrage process due to an inability to short sell and because arbitrage is costly and not equal across different stocks the effect of sentiment persists and influences stocks in an asymmetric way. Stock market characteristics associated with more costly arbitrage are stocks that are young, have a short earnings record, have high intangible values, are currently unprofitable but have potential for high growth. These characteristics impinge on the valuation process allowing irrational factors such as sentiment to have a more pervasive influence on their value. This means that when sentiment is optimistic there is a tendency for buying activity in hard to value stocks to rise but selling activity to rise when sentiment is pessimistic. This is noted in the work of Antoniou, Doukas and Subrahmanyam (2015) who show that when the market is optimistic high beta stocks become overpriced due to the influence of sentiment traders causing the security market line to become downward sloping. However, when the market is not optimistic less noise trading causes the security market line to be upward sloping.

A second reason why some stocks may respond differently to sentiment is that investors react differently to sentiment depending on their habitat preferences which causes them to have a preference for stocks with specific characteristics. This will cause asymmetric responses in different stocks when sentiment changes. For example, if retail investors have a preference for high liquidity stocks when sentiment rises individual investors will alter their investments in favour of stock market investments. However, if retail investors have a preference for low illiquidity securities, when these portfolio adjustments are made a disproportionately high amount of the investment will go into low illiquidity stocks. Thus the portfolio adjustments investors make in response to sentiment changes will depend upon their natural preferences or habitat.

As suggested by Barberis et al (2005) a third influence on the extent to which stocks are characterized by sentiment is the fraction of ownership held by different investors. This is important because some investors may be more likely to succumb to sentiment than others. For example, Kumar and Lee (2006) argue that retail investors do not have access to the same resources as institutional investors so high retail ownership may be associated with more sentiment motivated trading. When an investor group such as retail traders prefer a stock with particular characteristics so that a given type of investor holds a higher proportion of these stocks than other investors it is described as the investors natural habitat. The natural habitat of an investor can influence the level of sentiment if the type of investor associated with a habitat is less effective at arbitraging away inefficiencies.

2.4 Sentiment Trading by Retail and Institutional Investors

There is evidence to believe that both retail and institutional investors are influenced by sentiment. Although Kelley and Tetlock (2013) showed that the trading activity of retail investors predict short term returns over longer horizons individual investors have also been shown to underperform the market, an indication their trading is influenced by sentiment. For example, Odean (1999) finds that stocks bought by individuals underperform stocks sold by individuals over a twelve month horizon while Hvidkjaer (2008) shows that for small and medium sized firms the underperformance of purchases persists for up to three years[9]. Barber, Odean and Zhu (2008) show that retail purchases (sales) earn higher (lower) returns the following week but over a one year horizon purchases underperform sales. Barberis et al (2008) develop a model which explains these short run and long inconsistencies. They show that in the short run the buying activity of investors

[9] Hvidkjaer (2008) shows that volume from small trades to examine the relationship between retail investor trading behaviour and the cross-section of future stock returns and finds that stocks with the most seller-initiated small trade volume, measured over the previous several months, outperform stocks with the greatest buyer initiated small-trade volume. This return difference accrues during the first month and continues for two years but continues for three years among small and medium sized firms.

provides higher returns than sales because in the short run the effects of sentiment causes momentum but over longer periods markets adjust leading to return reversals and underperformance of purchases relative to sales.

Institutional investors also appear to be influenced by sentiment. Institutional investors represent a broad group and include diverse institutions such as hedge funds, pension funds, mutual funds as well as banks and insurance companies as well as independent investment advisors. One cause of institutional sentiment trading stems from reputational effects such as those discussed by Scharfstein and Stein (1990) who argue that institutional trading behavior is co-ordinated because one investment manager will not wish to behave very differently from another as they will not wish their performance to stand out[10]. In their model they show that managers will trade in the same direction as other managers even if they have opposing fundamental information rather than risk their reputation by performing below average. The use of sentiment allows managers to perform close to the average so that their performance does not stand out. Graham (1999) shows that when analysts have a high reputation in the industry, have low ability or if public information contradicts their private information analysts will prefer to trade in the same direction as other analysts rather than risk standing out from the crowd and being wrong.

Institutional investors invest on behalf of their clients and many of these will be individuals and smaller investors. Typically, institutional investors work to maintain their client base and prevent them from moving to rival institutions. If clients are themselves influenced by sentiment and the manager does not follow an investment strategy based on sentiment the manager may face declining revenue or have their position terminated by their underlying clients who are influenced by sentiment. These reputational effects cause institutional traders to employ sentiment.

[10] The idea is that if a fund manager does not use sentiment and their performance is below that of other fund managers they will stand out as underperforming managers even if they sometimes perform better than other managers. The strategy for fund managers to adopt therefore is to aim for the performance of average fund managers and not stand out because the reputational effects are so great.

Another reason why some institutions may adopt sentiment based trading is that in some circumstances it may be rational. We know that noise trading such as that induced by sentiment is positively correlated with short term returns as discussed above. This suggests that institutional investors that trade for short term gains are likely to benefit from trading on retail sentiment even when they are not exposed to reputational effects from their clients. They trade on sentiment because it is profitable to do so. Even if they wish to arbitrage away the predictability due to sentiment they are unable to because of the limits to implementing cost effective arbitrage strategies. As a result it is more profitable to trade on sentiment than arbitrage against the sentiment. Evidence of "riding the bubble" is provided by Brunnermeier and Nagal (2004) who examine the trading behavior of hedge funds during the technology bubble of 2000. They find that hedge funds were heavily invested in technology stocks rather than being a correcting force and arbitraging out of technology stocks. Griffin et al (2011) also examined trading patterns during the technology bubble but for a range of different institutions. They found that mutual funds and independent investment advisors as well as hedge funds invested even more heavily in technology stocks than retail investors during the bubble but divested at the point stock values began to decline unlike retail investors who continued investing even after institutional investors had shifted their positions. This suggests using sentiment can be profitable.

Further, evidence of sentiment influencing institutional investors is provided by Brown and Cliff (2005) who use survey data from "*Investor's Intelligence"* to track the number of bullish or bearish newsletters. Using this as a proxy for sentiment they find that during periods of optimistic sentiment stocks are overpriced. Bagnoli et al (2009) shows that institutional investors exhibit sentiment by documenting that analysts make more favorable recommendations when market sentiment is also favorable. Moreover, DeVault et al (2016) finds that it is the behavior of institutional investors that drives the relationship between stock returns and sentiment not retail investors. They also find that when market sentiment is optimistic institutional investors hold higher concentrations of speculative or hard to value stocks an indication that the behavior of institutions is influenced by sentiment.

2.5 Information, Sentiment and Co-movement

As we showed earlier changes in order flow can arise for two reasons, the arrival of new information or changes in sentiment. New information can be market-wide or firm specific. Roll (1988) argues that price impacts such as those described by Kyle (1985) are primarily caused by the arrival of new firm specific information which causes return correlation between stocks to be low as return variations are primarily firm specific and not shared by other firms. The arrival of new firm specific information will reduce return correlation amongst securities as firm specific factors exert a stronger influence over returns reducing the extent to which firm returns covary with the market. The arrival of new market-wide information will increase return correlation between stocks and the market as the effects of this news will influence stocks in general and therefore introduce common factors.

Sentiment increases co-movement as optimistic or pessimistic moods cause investors to undertake coordinated buying or selling activity which raises co-movement. Evidence exists that both retail investors and institutional investors contribute to co-movement as a result of sentiment. Barberis, Schleifer and Wurgler (2005) find that entry into the S&P500 index leads to an increase in beta for firms newly indexed which indicates that their co-movement with other indexed firms rises. They argue that since index entry has a neutral effect on fundamentals the observed rise in co-movement must be due to a rise in sentiment. They argue that this sentiment must be related to institutional sentiment because stock holdings of S&P500 index constituents are concentrated within the portfolios of institutional investors.

For retail investors Lee, Shleifer and Thaler (1991) find that the returns of stocks with low levels of institutional ownership move together more closely suggesting that a lack of institutional investors exacerbates co-movement[11]. Dorn et al (2008) show that the brokerage trading records of individual investors are correlated. Kumar and Lee (2006) show that buying activity of retail investors within low capitalization stocks is positively correlated with the buying activity in other small stocks and that the buying activity of one individual investor is correlated with the buying activity of other individual investors. This shows that the trading decisions of retail investors are coordinated. They also find that when stock holdings reflect higher concentrations of retail investment there are higher levels of order flow imbalance and sentiment and therefore higher levels of co-movement.

3. Data

We construct a sample of all NYSE/AMEX ordinary common stocks listed on the CRSP/COMPUSTAT merged database between the period January 1993 and December 2011[12]. From CRSP/COMPUSTAT, we extract daily security price and shares outstanding information along with CRSP value weighted market returns, volume information and data on firm characteristics.

3.1 Defining Buyer and Seller Initiated Trades

There is no single database or construction method that has been used to capture order flow imbalance. In this section we describe the method we use and why. From the Trades and Quotes (TAQ) database we obtain tick-by-tick data for NYSE/AMEX stocks. Using this information we identify buyer or seller initiated trades by applying the Lee and Ready (1991) algorithm which Lee and Radhakrishna (2000) show is highly accurate at separating buyer and seller initiated trades in equity markets. This algorithm has been used extensively in previous studies to capture order flow information, see for example Chordia et al (2002), Barberis et al (2005), Harford and Kaul (2005) or Barber, Odean and Zhu (2008).

[11] They show this through the analysis of closed-end fund discounts. They show that closed-end funds and small stocks tend to be held by individual investors, and that the discounts on closed-end funds narrow when small stocks do well.
[12] We begin in 1993 to coincide with availability of data on the TAQ database. Ordinary common stocks are identified using the CRSP share codes 10 and 11. This sample reflects a much longer sample than is usually studied by work that examines the influence of order flow.

Application of the algorithm requires comparison of transaction price to the contemporaneous quotes of the stock to ascertain whether a buyer or seller initiated trade has taken place. Seller initiated trades are those that take place below the mid-point price while buyer initiated trades take place above the mid-point price. In cases where this trade-quote comparison cannot be undertaken, the algorithm classifies buy trades as those that take place on an uptick and sells as those that take place on a downtick. Some trades such as those that take place during auctions cannot be classified. On average, each month, there are about five million trades associated with sample stocks. Using all intraday assignable trades, we calculate the aggregate dollar value of buyer initiated and seller initiated trades each day associated with each stock. The dollar value of all buyer initiated trades as a ratio to the dollar value of all seller initiated trades associated with a given stock each day is a measure of the total order flow imbalance level.

The ratio we employ has the advantage of automatically scaling the absolute size of the order flow to ensure it is the order flow imbalance rather than the absolute dollar value of order flow that is the focus of our investigations[13]. This is an important feature because small firms on average will naturally have less transactions and transactions of a smaller dollar value than large firms. The ratio is therefore similar but not identical to that used by Kumar and Lee (2006) who calculate the ratio of dollar value buy trades minus the dollar value of sell trades and scale by the value of total buys and sells [14].

3.2 Defining Trader Types

To partition retail and institutional trades we analyse each trade on TAQ and classify the trade using the Lee and Radhakrishna (2000) algorithm who show that trades below \$2,500 are retail while those above \$20,000 are institutional. Barber, Odean and Zhu (2009) provide evidence to show that the algorithm is efficient at

[13] When order flow is large buys are likely to be large and so are sells while when order flow is small buys and sells are likely to be small. This means the ratio of buys to sells is scaled.

[14] Their measure of order flow imbalance for day t is therefore $\frac{\sum_1^N (VB_{it} - VS_{it})}{\sum_1^N (VB_{it} + VS_{it})}$ where VB and VS are the dollar volume of buying and selling activity on day t calculated from trades 1 to N on day t. This is modified here as we construct and utilise ratios based on a daily frequency rather than a monthly.

identifying retail trades as they find that brokerage trades from retail investors are highly correlated with trades below the lowest cut-off. We calculate the daily order-flow imbalance level associated with retail trades by calculating the aggregate dollar value of buyer initiated trades to the aggregate dollar value of seller initiated trades using trades below the smallest cut-off.

There are likely to be two types of institutional traders active in the market, those that are informed and those that are uninformed. As shown by strategic trader models such as those of Easley and O'Hara (1987) institutional traders that are informed are likely to trade smaller quantities than preferred to disguise that they are informed. Empirical evidence of such "stealth trading" by informed institutional traders is provided by Barclay and Warner (1992) and Chakravarty (2001) who show that medium sized trades move prices more than any other trade size indicating they contain higher levels of information. Those initiating the very largest trades are likely to be uninformed institutional traders as they make no attempt to disguise their trading activity. Moreover, the recent study of Ülkü and Weber (2013) showed that some institutions can be informed while others can be uninformed which suggests that institutional trades are not necessarily a homogeneous group.

We calculate the order flow imbalance for large institutional traders as the ratio of aggregate dollar value of buyer initiated trades to the aggregate dollar value of seller initiated trades above the largest cut-off size. Trades between the smallest and largest cut-offs we designate as being from stealth traders as they will reflect the trades of institutions that are more likely to be informed. We calculate the order flow imbalance ratio of stealth traders as the daily aggregate dollar value of buyer initiated trades to the aggregate dollar value of seller initiated trades for trades on day *t* that are above the lowest cut-off but below the largest cut-off.

Kumar and Lee (2006) calculate their retail order imbalance ratios from buying and selling activity undertaken at a large U.S brokerage house over the period 1991-1996 which allows them to definitively assign trades as retail. Our use of TAQ data rather than brokerage reports or 13-F filings which are also sometimes used to calculate order imbalances (see for example Campbell et al (2009)) offers a range of advantages. TAQ data records all trades undertaken on NYSE/AMEX so our sample of retail investor trades is more representative of retail activity than a brokerage report from one firm. As noted by Kumar and Lee (2006) for some stocks thin trading was a problem which caused their order imbalance ratios to be noisy but with the wider range of trades captured by TAQ thin trading is not an issue of concern. The quantity of information available from TAQ data also allows us to study a longer time horizon than is usually possible or available from brokerage reports. Moreover, in some of the analysis we undertake we wish to study order flow over relatively short intervals. 13-F filings only facilitate an analysis over a quarterly interval period and are therefore not suitable for this analysis. As noted by Barberis et al (2005) 13-F filings do not offer a comprehensive coverage of investors as wealthy individuals and small institutions do not file 13-Fs. Finally, the analysis of investor holdings and brokerage reports does not provide information about who initiates a trade only that a trade or a change in holdings takes place. Sentiment is likely to be more closely associated with the party that initiates a trade which provides an advantage to using TAQ data and applying the Lee-Ready algorithm.

In 2001 decimalisation of quotations took place which reduced trading costs and increased volume, see for example Bessembinder (2003). Some of this increase in volume was from high frequency trading although a 2014 congressional report suggests that after 2008 high frequency trading has declined substantially. High frequency trades are often small which causes these institutional trades to be incorrectly assigned as retail using the Lee-Radhakrishna algorithm. We recognise this as a possible problem of partitioning by trader type. However, the effect of such errors in the assignment of trades on the order imbalance ratios we use should be minimal. A high frequency trader that initiates a buy(sell) trade will tend to offset that trade before the end of the trading day causing their end of day position to be neutral irrespective of how large the within day order flow imbalance of the high frequency trader was(see for example, Menkveld (2011) in particular Figure 5 or Broggard et al (2014)). This means that when we calculate the order flow imbalance for the day the effect of small trades from high frequency traders will be minimal.

From the constructed order imbalance ratio of each firm and each investor group we extract the influence of common-factors due to market-wide effects by regressing against the order imbalance ratio of firm *i* on day *t* the market excess return of that day. This allows us to extract the effect of market-wide news which has the potential to introduce a common factor which could mimic the effects of sentiment. The residuals from this regression contain the isolated sentiment component and form the basis of our measure of sentiment for firm *i* used in our later analysis. We call order flow imbalance/sentiment obtained across all trades total-OFIL, our retail constructed order flow imbalance and sentiment we call retail-OFIL, stealth and large trade constructed measures we call stealth-OFIL, and large-OFIL respectively. A value above unity for this ratio implies an excess of buying pressure and therefore optimistic sentiment while a value below unity implies an excess of selling pressure and thus pessimistic sentiment. The partitioning of trades from different investors allows us to determine whether the sentiment of institutional investors is different from each other and different to that of retail investors.

3.3 Stock Characteristics and Risk Measures

We employ a range of stock characteristics in our analysis which Baker and Wurgler (2007) noted as being possible features of stocks that are hard to value and more costly to arbitrage or by Barberis (2005) as possible investor habitats. Firm size (SIZE) is the market value of the firm (share price x number of shares issued) and has been identified as a characteristic of hard to value stocks due to a paucity of information, see for example Arbel and Strebel (1982). Smaller stocks are also more expensive to undertake arbitrage with because trading costs tend to be higher. Moreover, size was identified as a possible investor habitat by

Barberis et al (2005). Illiquidity also makes firms more difficult to value and more expensive to arbitrage. To capture illiquidity we utilise the Amihud (2002) illiquidity ratio (ILLIQ) which is calculated as the absolute return to volume ratio scaled by the market illiquidity ratio and captures the price impact of one unit of volume[15]. D'Avolio (2002) finds that approximately 10% of stocks are never short sold so are not arbitraged. These stocks tend to be small and illiquid stocks. The book-to market ratio (B/M) is included because a high book to market ratio can be an indicator of distress which makes these stocks more difficult to value and more expensive to arbitrage.

Evidence of higher costs of arbitrage for these firms is provided by Geczy et al (2002) who finds that the cost of short selling stocks with low book to market was five times higher than the costs of short selling stocks with high book to market. Moreover, distressed stocks is a preferred habitat for some investors but a habitat other investors avoid. We include price (PRICE) as the logarithmic value of the firms closing share price as low share prices can be a symptom of distress making firms difficult to value and arbitrage. Age (AGE) is the number of years a firm has been listed on the exchange. As noted by Baker and Wurgler (2007) younger firms have a shorter history of performance so they may be difficult to value because there is a higher level of uncertainty associated with them. The correlation with the market (CM) is the daily correlation in stock returns between firm *i* and the market. Institutional investors tend to have a preference for firms highly correlated with the market such as S&P500 index constituents. While D'Avolio (2002) shows that S&P500 index constituents can be used as collateral for short selling activity indicating they are easy arbitrage. A high level of correlation with the market also makes stocks easier to arbitrage and value as a greater component of their valuation can be referenced from other sources i.e. the market return.

Intangibles (INT) represent the proportion of intangible assets the firm reports as a ratio to overall assets and represents the component of firms assets that are most difficult to value. Therefore firms with higher levels of

[15] To be precise this measure is also scaled by 10^6 as suggested by Amihud (2002) to correct for the very small size of the Amihud ratio.

intangibles become more difficult to value as overall firm value contains a greater amount of uncertainty. Earnings per share (EPS) is the total earnings of a firm divided by the number of shares issued. The value of firms with low or negative earnings have more uncertain future values especially if this is combined with other characteristics that make firms hard to value. Earnings growth(EG), which is the return on equity (ROE) multiplied by one minus the dividend pay-out ratio, may also be associated with firms that are hard to value. Firms with high earnings growth are more difficult to value as the range of future growth rates are likely to be more dispersed. We also utilise the proportion of share ownership (OWN) held by institutional investors which we obtain from Thomson Reuters 13-F filings This is the total institutional ownership as a percentage of shares outstanding. Ownership is important as it identifies whether a firm is held more aggressively by retail or institutional investors which identifies whether the stock is a habitat of retail or institutional investors. Furthermore D'Avolio (2002) shows that firms with high institutional ownership are cheaper to arbitrage.

The market return R_m is the return to the equal weighted market return provided by CRSP. The risk free rate is the return to a US treasury bill (R_f). The SMB, HML and UMD factors are obtained from Ken French's website and refer respectively to the return on a portfolio of small firms minus the return to a portfolio of large firms, the return to high book-to-market firms minus the return to low book to market firms, the return of the highest performing portfolio minus the return of the lowest performing portfolio measured over the previous six months. Macroeconomic variables that we use are all obtained from Compustat, these are described later.

4. Summary Statistics

In this section we examine the sentiment levels of different types of investors across the market as a whole and for firms organised into groups on the basis of firm characteristics. We also provide some preliminary evidence of how changes to order flow imbalance, an initial measure of shock, influences co-movement. The role of this section is to motivate the analysis we undertake using the STCC GARCH model later.

4.1 Sentiment Levels

We begin by reporting the average daily order-imbalance level (OFIL) across individual stocks as this forms the basis of our measurement of sentiment. Since we have already extracted common factors due to market news, if sentiment is neutral, on average OFIL will be close to unity (over time and across stocks firm specific news is diversified away). However, if optimistic sentiment exists on average OFIL will be >1, but if average sentiment is pessimistic OFIL will be <1[16]. We calculate the correlation coefficient between our different investor sentiment measures and the Baker-Wurgler (2006) sentiment index which has become the primary sentiment metric (see for example, Antoniou, Doukas and Subrahmanyam (2013), Stambaugh, Yu and Yuan (2012a, 2012b)). We find that Total-OFIL, the measure of overall sentiment and the Baker-Wurgler index has a correlation value of 0.294(p value 0.00). Retail investor sentiment has a correlation of 0.255(p value 0.00), stealth investor sentiment has a correlation of 0.188(p value 0.00) and large investor sentiment a correlation of 0.087 (p value is 0.02). We would not necessarily expect the correlation between different sentiment measures to be very high. DeVault et al (2016) for example, find that the correlation between the Baker-Wurgler index and the measure of consumer confidence, an alternative measure of sentiment, is on average about 20%. Moreover, changes in sentiment, as measured by an index may take some time to lead to actual changes in trading behavior which would lead to a lack of synchronicity between order flow changes and changes in sentiment when recorded through a sentiment index.

In Figure 1 Panel A to D we trace out the average level of sentiment (OFIL) for each investor group each day during the sample period to show how sentiment levels of each investor group fluctuates over time. Panel A shows that average total sentiment has no trend movements until about 2000, increased steadily until about

[16] Order-flow imbalance levels will be influenced by positive and negative news but on average should reflect the stochastic nature of information arrival. Although market-wide news should also arrive randomly so that the effect of good news on order flow one day is cancelled by bad news another day it is possible that some residual market-wide news effects exist.

2005 then declined slightly until 2007. During 2007 there was a steep decline in sentiment which was not reversed by the end of the sample period. Similar trends are observable for retail, stealth and large OFIL except that the decline in sentiment after 2007 is most pronounced for large-OFIL and least pronounced in retail-OFIL. The decline in sentiment during 2007, most evident for stealth and institutional trading reflects the start of the financial crisis. During 2007 the banking sector faced a "credit crunch" as subprime mortgages began to default and securities linked to their value began to decline in value. As banks began to realise these losses interbank lending, an important source of liquidity for the financial system, declined substantially. This led to a less favourable outlook and therefore a reduction in sentiment.

Panel A of Table 1 provides average daily sentiment levels for the full sample of stocks and for broad sub-samples based on whether or not firms are S&P500 index constituents. This partition is motivated by Harford and Kaul (2005) who showed that indexing leads to higher correlations between the order flow of one firm and another and by Barberis et al (2005) who showed that indexing is associated with sentiment[17]. We also partition according to whether OFIL on that day was >1 or <1. This allows us to examine separately firms characterized by optimistic and pessimistic sentiment. The average value of total-OFIL across all stocks is slightly positive (1.053) which is consistent with positive sentiment. The average retail-OFIL is 1.039, the stealth-OFIL is 1.047 and the large-OFIL is 1.133. This suggests that large traders have on average the greatest positive sentiment and retail investors the least.

For S&P500 index constituents the total average imbalance is 1.158, 1.028 for retail investors, 1.077 for stealth traders and 1.204 for large traders. This shows that large institutional traders have the most positive sentiment but retail investors have the least. For non S&P500 index constituents average OFIL values are

[17] Harford and Kaul (2005) use principal components analysis on order flow and returns to show that a principal component which reflects indexing is almost perfectly correlated with the returns and order flow for of S&P500 index constituents. Some investors who switch into equity from less risky investments because sentiment has increased will invest heavily in stocks which are constituents of an index or may even invest in an index product. This happens because some investors select products that they are more familiar with and avoid investments that are less well reported in the financial press.

generally smaller and closer to unity, although even amongst these stocks the sentiment displayed by large institutional traders appears elevated.

When we split the sample according to whether the OFIL ratio of firms is >1 or <1 on day *t* we find that total-OFIL is 1.585 when the ratio >1 but 0.636 when the ratio is <1. This partitioning also reveals large differences between OFIL averages across the different investor groups. Large-OFIL displays the greatest imbalance in favour of buying pressure[18] when the OFIL ratio is >1 (large-OFIL is 1.815) and also displays the largest imbalance in favour of selling pressure when the ratio is <1(large-OFIL is 0.588). Stealth traders display the smallest average imbalances (1.427 when OFIL >1 and 0.702<1). Retail investors have an average OFIL of 1.509 when the ratio is >1 but a ratio of 0.665 when retail-OFIL is >1.

Although our results show that large institutional traders exhibit more positive sentiment than retail traders this does not necessarily imply that retail investors are less prone to sentiment. If retail investors have positive sentiment for some stocks but negative sentiment for others there may still be high levels of sentiment present which does not manifest itself in average retail-OFIL values because positive sentiment in one stock is being offset by pessimistic sentiment in another.

To determine whether sentiment levels may be dependent on the characteristic of a security Panel B provides OFIL information for groups of stocks which are sorted by a range of stock characteristics linked to sentiment. Panel B shows there are large differences in the average values of OFIL both across the trader groups and across the different stock categories. For example, the average total-OFIL is 0.9038 for the small firm quintile but 1.1224 for the large firm quintile suggesting positive sentiment for large firms and negative sentiment for small firms. Positive sentiment is also associated with firms that have low illiquidity, low book to market, high prices, older ages, higher institutional ownership, higher intangibles, a lower EPS, and higher

[18] Although large institutional traders will trade the largest dollar volumes each day their order flow imbalance over time and across different stocks should on average equate. Larger dollar volume trades should not necessarily be associated with larger dollar imbalances over time and across stocks.

earnings growth. While pessimistic sentiment is displayed by firms with high illiquidity, low prices, young ages, low institutional ownership. The disparity in total-OFIL or sentiment levels across stocks with different characteristics suggests that these characteristics have an important influence over the overall sentiment displayed.

We also find that the sentiment level displayed by each investor group is influenced by firm characteristics. Retail traders display higher levels of positive sentiment for larger firms, low book-to-market stocks, high price stocks, older stocks, firms with higher levels of institutional ownership, lower EPS and higher earnings growth. Pessimistic sentiment is displayed by retail investors for small stocks, low priced stocks and young stocks. Stealth traders display positive sentiment for larger stocks, low illiquidity stocks, low book-to market stocks, high price stocks, stocks with high institutional ownership, low EPS values and high earnings growth. However, negative sentiment is displayed by stealth traders when stocks have a low price. Large traders display the greatest amount of positive sentiment as average OFIL levels in each group are positive but the most positive sentiment is displayed within the groups containing small stocks, liquid stocks, low book-to-market and stocks with high prices, older stocks, stocks with high institutional ownership, high intangibles, low EPS and high earnings growth. This suggests that the firm characteristics influence the sentiment of each type of investor and that average retail-OFIL values are on average lower because positive sentiment in some stocks is being offset by negative sentiment in other stocks.

4.2 Sentiment Correlation among Investors

We next measure the correlation between the change in the OFIL (we refer to the change in OFIL as D-OFIL) of stock *i* and stock *j*. Since we have controlled for market-wide news effects and because good and bad news arrives randomly the buying and selling activity of one stock and another should be uncorrelated. However, if variations in OFIL are due to sentiment shocks then changes in buying and selling activity will be influenced by a factor common to all stocks in the sample which raises the correlation in D-OFIL amongst stocks.

As shown by Dorn et al (2008) the trading activity of retail investors does not appear to be random which suggests that sentiment must also be influencing the buying and selling activity of retail investors in different stocks. Our analysis broadens this debate by examining whether the co-ordination of trading behavior across different stocks is also influenced by the investor group.

Panel A of Table 2 presents average pairwise correlations across all stocks and for sub-samples and shows that the average pairwise correlation between the total D-OFIL of stocks is 3.8% which does not suggest that total D-OFIL across stocks is correlated. Average total D-OFIL pairwise correlations are also low for the S&P500 constituents (1.6%) and non S&P500 constituents (4.9%). We do not find that indexing per se leads to higher correlations in the total D-OFIL of one firm and another if firms are already in the S&P500 index[19]. Average pairwise total D-OFIL correlations of firms is 5.2% if the total-OFIL is >1. Only when total-OFIL is <1 does the average correlation between the total D-OFIL of firms appear to be elevated (18.5%). This indicates greater correlation in trading activity across stocks when total-OFIL is <1 and sentiment is pessimistic.

The average pairwise retail D-OFIL correlation across all stocks is 9.5% and therefore the highest of the three investor groups. This indicates retail trading activity across different stocks is more coordinated than is evident for both types of institutional traders since the average pairwise correlation in D-OFIL across stocks is only 5.6% for stealth traders and 5.3% for large institutional traders. The average retail D-OFIL pairwise correlations are also higher than for institutional investors when stocks are partitioned according to whether stocks are S&P500 constituents or non S&P500 constituents and when they are partitioned according to

[19] The prior removal of the market factor is likely to account for this.

whether investor OFIL is >1 or <1. It is also noticeable that although retail D-OFIL correlations are always larger, the D-OFIL correlations of large institutional traders are also elevated if stocks are S&P500 constituents (13.6%) and elevated for stealth traders if stocks are non S&P500 constituents (7.1%).

Panel B of Table 2 presents average pairwise D-OFIL correlations calculated between stock *i* and *j* within groups partitioned by firm characteristics. Average pairwise total D-OFIL correlations are positive but vary according to the firm characteristics. The highest average pairwise correlations are associated with firms in the youngest age group (14.47%) and the group containing the smallest firms (8.09%). This suggests that there is greater co-ordination in the trading activity of firms within some specific groups than there is across all firms in general showing that the degree of co-ordination in trading activity is influenced by firm characteristics.

Firm characteristics also influence the size of correlations for the different investor groups. Moreover, in some cases there is a relationship between the size of the correlations and the size of the firm characteristics. For example, average retail D-OFIL pairwise correlations are elevated for all firm size groups but rise almost monotonically as firms get larger (average correlations are 8.73% for small firms but 20.09% for large firms). The influence of firm size on the magnitude of average correlations is also apparent for stealth and large institutional investors. For both groups their average D-OFIL correlations are higher for small firms than for large firms. D-OFIL correlations amongst the small firm group are 15.39% for stealth traders and 41.40% for large institutional traders.

For retail investors the D-OFIL correlations are elevated and seem associated with the size of the characteristic when firms have higher levels of institutional ownership, when firms have larger intangible values and lower EPS values. For stealth traders illiquidity, higher book-to market ratios, high price, young age, lower intangibles, higher EPS values and lower earnings growth raise the D-OFIL correlations. For large traders high illiquidity, high book-to-market, high price, a young age, high correlation with the market, raise D-OFIL correlations.

We next undertake some multivariate analysis. From the stocks in our sample we form portfolios containing randomly assigned securities. Each portfolio contains twenty stocks which will be enough to diversify the effects of firm specific information but also allow for cross-sectional variation in OFIL measures. We then calculate the correlation between firm *i's* D-OFIL and the market D-OFIL for each stock within a portfolio. Using these values we create the average of these correlations for portfolio *p*. We regress against each portfolio correlation the average firm characteristic of each portfolio. This allows us to examine whether characteristics contribute to portfolio correlation. The results are presented in Panel C of Table 2 and show that some firm characteristics influence correlations. The results for Total-D-OFIL indicate that the correlations are positively influenced by firm size, intangibles and return correlation with the market but negatively influenced by illiquidity, age and ownership. The correlations from retail investors are influenced positively by their return correlation with the market and firm age and negatively related to illiquidity. The correlations for stealth investors are positively related to the extent returns are correlated with the market, ownership and the average order imbalance. The correlations calculated for large investors are positively related to size and the average order imbalance and negatively related to the return correlation of a firm with the market. We do not find that S&P membership influences correlation for any investor group. Overall the results show that firm characteristics influence correlation but not in a homogeneous way across investors.

4.3 Multifactor Time-Series Models

Kumar and Lee (2006) regress the returns of portfolios in excess of the risk free return (portfolios are formed on the basis of firm size to form five size based groups) against their order flow imbalance measure and risk factors which control for market-wide effects that can introduce common factors. A positive relationship between sentiment and returns would suggest that more positive sentiment leads to higher portfolio returns and therefore higher co-movement. Kumar and Lee (2006) found that the sentiment of small investors was associated with the portfolio returns of small firms but there was no relationship between sentiment and the returns of large firms. We extend their analysis by estimating equation (1) which allows us to determine whether the sentiment of different types of traders influence returns.

$$R_{pt}\text{-}R_{ft} = \alpha + \beta_1(R_{mt}\text{-}R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \beta_5 OFIL_{pt} + \xi_{pt} \qquad (1)$$

Where, R_{pt} is the return to the randomly created portfolio *p*. In the analysis we use we do not create portfolios formed by firm size as in Kumar and Lee (2006) because this favors finding a relationship between small firms and retail investor sentiment as retail investors are more heavily concentrated in small firms. To ensure our portfolios are able to reflect the sentiment of each of our different investor groups we form portfolios by randomly combining stocks to form equal-weighted portfolios. R_{ft} is the return to the risk free rate. Risk measures are as previously defined. With the inclusion of the Fama-French systematic factors and the use of portfolio returns the effect of both firm specific and additional market-wide news effects are controlled for. In the Kumar and Lee (2006) analysis the effect of institutional sentiment on co-movement is not considered. To show how the sentiment of different investors influences returns we estimate different versions of equation (1) so that each version includes the OFIL of the different investors. Each regression contains either Total, Retail, Stealth or Large OFIL and one version contains Retail, Stealth and Large OFIL in the same regression. Across the different regressions we can determine how the isolated sentiment of each investor group influences co-movement. ξ_{pt} is the residual return. We use monthly interval analysis to compare our results more clearly with those of Kumar and Lee (2006).

The results of this estimation are provided in Panel A of Table 3 and show that Total-OFIL is positively related to excess stock returns showing that this measure of sentiment influences returns and therefore co-movement. However, as well as total market sentiment having an influence over portfolio excess returns we also find that both retail stealth and large institutional OFIL are positively related to portfolio returns when these measures of sentiment are included in the regression separately. This suggests that sentiment levels of these investors appear to raise co-movement. The regression which includes Retail, Stealth and Large OFIL shows that only Retail and Large OFIL are positively correlated with returns. Stealth OFIL is no longer significant suggesting that the influence of sentiment associated with Stealth investors if fully captured by Retail and Large investor sentiment as we do not find that the OFIL of stealth traders has any independent influence on portfolio returns.

4.4 *Correlation, Order Flow Imbalance and Excess Market Return*

We next calculate the correlation between the daily change in the order flow imbalance level D-OFIL of portfolio *p* and the market excess return R_m-R_f. We use the change in order flow imbalance (D-OFIL) as a measure of sentiment shock associated with portfolio *p*. The correlation between sentiment shocks and R_m-R_f allow us to examine how shocks affect co-movement on average. Kumar and Lee (2006) show that changes to their order imbalance measure is negatively associated with market returns (Table III Panel B, p2466) but do not discuss this result in any depth. This is an important result as it suggests that shocks or surprises to sentiment reduce co-movement so requires further examination.

We extend this analysis by calculating the average pairwise correlation between the excess market return R_m-R_f and the D-OFIL of portfolio *p* calculated for each of the investor groups on day *t*. Average pairwise correlations are presented in Table 3 Panel B, and shows that there is on average a negative correlation between total D-OFIL and the market excess return. We also find that there is a negative association between retail D-OFIL and the market excess return suggesting that changes in the OFIL of retail investors reduces co-movement. The D-OFIL of stealth and large institutional investors is not significantly correlated with changes in market excess returns. This is consistent with the changes to OFIL of these investors not impacting on co-movement changes.

4.5 Non-Stationarity of Order Flow Levels

An important reason for studying the impact of order-flow shocks is that order flow imbalances are likely to contain a non-stationary component. Schmeling (2007) found that sentiment indexes are non-stationary and whilst we are measuring sentiment differently to Schmeling (2007) who used an index of consumer confidence, if order-flow metrics reflect sentiment they too are likely to be non-stationary. Moreover, Ülkü and Weber (2013) undertake Dickey-Fuller non stationarity tests on aggregate order flow of different types of institutional investors and find that the level of each series is non-stationary but changes are stationary. Since our order flow imbalance measures are constructed from order flow information we would expect them to be non-stationary also.

Prior to estimating the smooth transition model we apply Ng-Perron non-stationarity tests on the OFIL and D-OFIL series of each of the stocks and for the portfolios we described in Section 4.3. As shown in Table 3 Panel C we find that for the sample stocks only 3.4% of total-OFIL measures appear to be stationary but 100% of total D-OFIL measures are stationary. Across the portfolios, 5.08% of the total-OFIL series are stationary and 100% of the total D-OFIL series are stationary[20]. A similar pattern is evident for the retail, stealth and large-OFIL and D-OFIL series.

In this section, through a variety of tests we have shown that both retail and institutional sentiment can influence co-movement. The tests we have used however can be considered flawed as none of the tests we have examined tests directly whether order flow changes or shocks directly influence co-movement. Instead they indirectly suggest that changes in order flow imbalance and therefore sentiment are connected to co-movement and that changes in order flow imbalance lead to a reduction in co-movement. The analysis of this section motivates us to examine in the next section the STCC GARCH model which allows a direct

[20] In addition to the Ng-Perron tests we also undertake stationarity tests using Augmented Dickey Fuller, Phillips-Perron and Elliott-Rothenberg-Stock tests and find that in general order flow imbalance in levels is non-stationary while first differences are stationary.

relationship between order flow changes and return co-movement to be identified if it exists. Moreover, we have shown that when examining the issue of how sentiment influences stock returns shocks to sentiment should be examined rather than levels as levels will tend to be non-stationary. This suggests that the analysis of Kumar and Lee (2006) can not only be extended to institutional investors but can be improved upon in a number of ways.

5. Smooth Transition Analysis

Motivated by the desire to provide direct evidence that shocks to sentiment influence co-movement we now introduce the STCC GARCH model we estimate which is the primary focus of our analysis[21]. A specific advantage of the smooth transition model is that it only identifies a change in order flow imbalance as a shock if the change in the order flow imbalance reaches an endogenously determined threshold. An advantage of measuring shocks in this way is that minor variations in order flow imbalance will be ignored making shocks less noisy.

The bi-variate STCC GARCH model proposed by Berben and Jansen (2005) and Silvennoinen and Teräsvirta (2005), enables us to capture the intertemporal time-varying behavior of stock return correlations[22]. In the smooth transition model we estimate $\{Y_t\}$ contains the vectors $\begin{bmatrix} y_{pt} \\ y_{mt} \end{bmatrix}$. Where y_{pt} is the excess return of portfolio p on day t, $(R_{pt}-R_{ft})$ and y_{mt} is the excess return of the market on day t, $(R_{mt}-R_{ft})$. We utilize portfolio returns rather than stock returns to ensure that we eliminate the effects of firm specific news. We eliminate

[21] We utilise order flow shocks rather than return shocks for a variety of reasons. First, Hasbrouck and Seppi (2001) in empirical work and the theoretical models we discussed in Section 2 show that causality runs from trading activity to prices. Second, returns contain frictions such as the bid-ask spread and errors in the analysis and interpretation of information, see for example Black (1986) so do not fully reflect order-flow or fundamentals. Moreover, the use of returns would not allow us to isolate the impact of sentiment which is our central focus. We recognise as shown by Ülkü and Weber (2013) recently that there may also be some bi-directional causality since for example high buying pressure will raise returns which will encourage more trading activity. We attempt to mitigate these effects by eliminating serial correlation in both the return series and the order flow series but cannot guarantee these effects are eliminated completely.

[22] Similar models (that capture possible regime switches) have been used in the context of stock market integration (Kearney and Poti, 2006; Silvennoinen and Teräsvirta, 2009; Savva, 2011), business cycles synchronization (Savva et al, 2010) and various macroeconomic relationships (Kapetanios and Tzavalis, 2010; Koop and Potter, 2007; Li, Philippopoulos and Tzavalis, 2000; among others).

any additional effects of market-wide information by including in the specification the market-wide risk factors HML, SMB and UMD[23] and a vector of N macroeconomic innovations ($MACRO_{Nt}$). These include the innovations associated with unexpected inflation, the monthly growth in industrial production, changes in the term structure, change in the difference between the yields of Moody's BAA rated and AAA rated corporate bonds, changes in the monthly unemployment rate, and innovations in the average hourly earnings. The *n* λ coefficients capture the effects of changes in these macroeconomic variables on portfolio returns.

We also control for the effects of macroeconomic news in the market return series y_{mt}. To control for potential autocorrelation, lagged values of the portfolio/market returns are also included.[24] The bi-variate specification we estimate for each portfolio is provided by equation (2a) and (2b) below.

$$y_{pt} = \alpha_p + \sum_1^k \phi_{pk} y_{p,t-k} + \beta_p (R_{mt} - R_{ft}) + \omega_p HML_t + \delta_p SMB_i + \theta_p UMD_t + \sum_1^N \lambda_{pn} MACRO_{Nt} + \varepsilon_{pt} \quad (2a)$$

$$y_{mt} = \alpha_m + \sum_1^k \phi_{mk} y_{m,t-k} + \sum_1^N \lambda_{mN} MACRO_{Nt} + \varepsilon_{mt} \quad (2b)$$

In equation (2a) and (2b) excess returns are conditional on all the information available up to time *t-1* i.e. ($Y_t=E[Y_t|\Psi_{t-1}]$). In order to capture any temporal effects in the error volatilities and correlations, the error process of each return series is represented by E_t which contains the vectors $\begin{bmatrix} \varepsilon_{pt} \\ \varepsilon_{mt} \end{bmatrix}$. These errors are assumed to have a zero mean and time varying structure as shown below.

$$E_t \mid \Psi_{t-1} \sim N(0, H_t) \quad (3)$$

$$H_t = E[E_t E_t^{'} \mid \Psi_{t-1}], \quad (4)$$

$$h_{pt} = w_p + \alpha_p \varepsilon_{p,t-1}^2 + \beta_p h_{p,t-1}, \quad (5)$$

$$h_{mt} = w_m + \alpha_m \varepsilon_{m,t-1}^2 + \beta_m h_{m,t-1}, \quad (6)$$

[23] Recall we already extract (Rmt-Rft)
[24] The appropriate lag order is defined using the Swartz Information Criterion (SIC).

In this model the conditional variances of returns for both portfolio p (h_{pt}) and the market return (h_{mt}) are assumed to follow a GARCH(1,1) specification described by equation (5) and (6) respectively. The use of this specification is motivated by the empirical literature that has shown it adequately captures persistence of stock return second moments.[25]

The contemporaneous conditional covariance between portfolio p and the market portfolio m and therefore the conditional correlation measure we use in our analysis is described below

.

$$h_{pmt} = \rho_{pmt}(h_{pt}, h_{mt})^{1/2} \tag{7}$$

$$\rho_{pm,t} = \rho_{pm,0}(1-G_t(s_t;g,c)) + \rho_{pm,1}G_t(s_t;g,c) \tag{8}$$

The smooth transition model allows the conditional correlations to change smoothly between two constant states as a function of the transition variable s_t as shown in (8) where the conditional correlation in the pre-shock period is ($\rho_{pm,0}$) and the post shock correlation is ($\rho_{pm,1}$). G_t is the transition function whose values are bounded by 0 and 1. To capture temporal changes in $\rho_{pm,t}$ we follow Berben and Jansen (2005) and Silvennoinen and Teräsvirta (2005) by letting the logistic function

$$G(s_t;g,c) = 1+e(-g(s_t-c))^{-1}, \qquad g>1 \tag{9}$$

where s_t is the order flow imbalance transition variable which we will call $OFIL_{st}$ to distinguish it from OFIL. The $OFIL_{st}$ are the residuals from a regression in which the percentage change in OFIL between day t and day $t-1$ is regressed against a constant and the macroeconomic factors $\sum_1^N MACRO$. To remove any serial

[25] The sizes of α and β, determine the short and long run dynamics of the resulting volatility series, respectively. Large β coefficients indicate that shocks to conditional variance take a long time to die out, implying persistent volatility. Large α parameters indicate that volatility reacts quite intensively to new information. Consequently, if α is large (and significant) and β is small, this means the volatility process is characterized by spikes.

dependence we also include the lagged values of $OFIL_{st}$[26]. The residuals from this regression will contain the effects of sentiment on order flow imbalance but will not contain the effects of firm specific, market-wide news (contained in Rm-Rf and already extracted as OFIL is used in the regression to obtain $OFIL_{st}$) or any additional market news stemming from macroeconomic factors[27].

Parameters g and c determine the smoothness and location, respectively, of the transition between the two correlation regimes. The starting values of g and c are determined by a grid search and are estimated in one step by maximizing the likelihood function. The likelihood function we estimate is shown below.

$$l_t(\theta) = -\frac{1}{2} ln|H_t| - \frac{1}{2} E_t' H_t^{-1} E_t \qquad (10)$$

where θ is the vector of all the parameters to be estimated.

$$\mathrm{L}(\theta) = \sum_{t=1}^{n} l_t(\theta) \qquad (11)$$

To allow for potential non-normality of $E_t|\Psi_{t-1}$, robust "sandwich" standard errors (Bollerslev and Wooldridge, 1992) are used for the estimated coefficients.

The resulting Smooth Transition Conditional Correlation (STCC) GARCH model is able to capture a wide variety of patterns of change where $\rho_{pm,0}$ and $\rho_{pm,1}$ represent the pre and post shock conditional correlations. Differing $\rho_{pm,0}$ and $\rho_{pm,1}$ imply that the conditional correlations increase ($\rho_{pm,0} < \rho_{pm,1}$) or decrease ($\rho_{pm,0} > \rho_{p,m,1}$) in the post shock period, with the pace of change determined by the slope parameter g. This change is abrupt for large g, and becomes a step function as $g \to \infty$, with a more gradual change represented by smaller values of this parameter.

[26] Again appropriate lag order is determined by SIC.
[27] We include the Fama-French market-wide factors so the effect of macroeconomic news should be captured by the market return. However, since the effect of different macroeconomic factors may have a diverse impact we also include the macroeconomic factors which will allow us to capture the impact of macroeconomic news even when the responses between factors vary.

The parameter c defines the location of the transition and represents the order flow *threshold* used in this study to identify shocks. For each portfolio, when the percentage change in daily portfolio order flow is less (greater) than the corresponding threshold (parameter c), the conditional correlation between the return of portfolio p and the market return is closer to the state defined by $\rho_{pm,0}$ ($\rho_{pm,1}$) on that day. Once a shock is detected the conditional correlation in the pre-shock and post-shock regime is calculated. We estimate the model (equations 2 -11) for all portfolios in bivariate combinations with the market. In subsequent tables we report the averages of the estimated parameters $\rho_{pm,0}$ and $\rho_{pm,1}$ for the portfolio samples that are described. We use four different measures of $OFIL_{st}$ in our estimation. We use shocks to total-$OFIL_{st}$, shocks to retail $OFIL_{st}$ and shocks to stealth and institutional $OFIL_{st}$ to examine whether shocks to the sentiment of different investors directly influences co-movement.

Prior to estimating the smooth transition model we apply the Lagrange Multiplier (LM) test to each bi-variate equation to test the time varying model specified against a constant conditional correlation model as suggested by Berben and Jansen (2005) and Silvennoinen and Teräsvirta (2005) and find that the null hypothesis is almost always rejected which suggests that a time varying model is appropriate[28] .

6. Smooth Transition Results

6.1 All Firms

In Table 4 Panel A we present the average of the pre and post shock conditional correlations obtained from the smooth transition model prior and subsequent to a shock to $OFIL_{st}$. These are the averages of the conditional correlations $\rho_{pm,t}$ calculated for the pre ($\rho_{pm,0}$) and post shock period ($\rho_{pm,1}$). In addition to these conditional correlations we also report the average change in conditional correlation after a shock to $OFIL_{st}$ (average difference between $\rho_{pm,0}$ and $\rho_{pm,1}$ across all portfolios) along with a t-test which examines whether this average change is significant. We also conduct a further t-test which examines whether the post shock change in correlation is different for positive and negative shocks to $OFIL_{st}$ (the + in Table 4). We also report

[28] More specifically, the hypothesis is rejected for 92.3% of the total sample, 92.2% for retail, 92.2% for medium and 92.4% for large.

the percentage change in conditional correlation, each average change represents, to gauge the potential economic significance of the conditional correlation changes we have identified.

For example, prior to a positive shock to total-$OFIL_{st}$ the average conditional correlation between portfolio *p* and the market portfolio is 0.1860 and after a positive shock to total-$OFIL_{st}$ average conditional correlations fall to 0.1693. Prior to a negative shock to total-$OFIL_{st}$ pre shock conditional correlations are 0.2094 and post shock conditional correlations are 0.1782. This represents an average change of -0.0167(-8.98%) in conditional correlation after positive shocks to total-$OFIL_{st}$ and an average change in conditional correlation of -0.0312(-14.90%) after negative shocks to total-$OFIL_{st}$. The + in Table 4 indicates that the change in correlation associated with positive and negative shocks to total-$OFIL_{st}$ is statistically significant from each other using a *t-test* at a 5% level of significance. Moreover, 63.2% of firms experience a fall in conditional correlation after a positive shock to total $OFIL_{st}$ and 64% following a negative shock to total $OFIL_{st}$.

The effect of total-$OFIL_{st}$ shocks reveal an asymmetry in the size of the average conditional correlation changes associated with positive and negative shocks because negative shocks lead to larger changes in average conditional correlation than positive shocks. The usual explanation for such an asymmetry is that there has been a rise in conditional volatility of the negative shock sample relative to the positive shock sample. There are two possible competing explanations for this. The leverage explanation for the asymmetry suggested by Black (1976) or Christie (1982) suggests that a fall in the value of the portfolio increases financial leverage, which makes the portfolio riskier raising its volatility. An alternative view for such an asymmetry suggested by Campbell and Hentschel (1992) is that the asymmetry is due to time variation in expected returns associated with the pricing of volatility. The causality of these two theories is therefore different. The leverage hypothesis suggests that return shocks lead to changes in conditional volatility, whereas the time-varying risk premium argument suggests that return shocks are caused by changes in conditional volatility. An analysis of the average pre and post shock volatilities suggest that average post shock volatility after a negative shock to total $OFIL_{st}$ rises by 15%, about 8% for retail and stealth $OFIL_{st}$ and by about 4% after negative shocks to large $OFIL_{st}$. In contrast volatility changes after positive shocks are almost neutral and never more than +4% or -4% for trader group.

Conditional correlation reductions following a shock to retail-$OFIL_{st}$ as well as stealth-$OFIL_{st}$ and large-$OFIL_{st}$ are also evident. The largest average change in conditional correlation follows a shock to retail-$OFIL_{st}$ as positive shocks on average lead to a 10.09% reduction in conditional correlation while a negative shock on average leads to a 14.74% fall in conditional correlation. Positive (negative) shocks to stealth-$OFIL_{st}$ on average lead to a 8.07% (9.64%) decline in conditional correlations. This indicates that post-shock average correlation changes are smaller than for retail investors but also that the asymmetry of positive and negative shocks on average conditional correlation is also smaller following stealth-$OFIL_{st}$ shocks than following retail $OFIL_{st}$ shocks. Shocks to large-$OFIL_{st}$ lead to the smallest post shock falls in average conditional correlation which are 7.01% following a positive shock to large-$OFIL_{st}$ and 5.87% following a negative shock to large-$OFIL_{st}$. Moreover, in contrast to what we found for retail and stealth traders positive shocks to large-$OFIL_{st}$ lead to slightly larger average changes in conditional correlation than do negative shocks to large-$OFIL_{st}$ which causes the asymmetry to be reversed. A possible explanation for why shocks to large $OFIL_{st}$ have a smaller impact on changes to conditional correlations than shocks to retail $OFIL_{st}$ may stem from large institutional traders having very positive sentiment on average which might dilute the effects of a sentiment change.

There is already a large literature that examines the impact of return shocks on correlation, see for example, Forbes and Rigbon (2002), Ang and Chen (2002), Hong, Tu and Zhou (2007). The results from these studies show that positive shocks to return lead to a rise in co-movement and negative shocks to return lead to even larger increases in co-movement. Our results are not comparable to these studies for a number of reasons. We do not examine the effect of return shocks on correlation but examine the effects of order flow imbalance, our measure of sentiment. Moreover, we extract from order flow imbalance the effects of the market return, and factors reflecting size, book-to market and momentum as well as macroeconomic information. This provides us with a measure of order flow imbalance that is free of the effects of information originating from market-wide factors other than sentiment. In contrast, in these previous studies the effect of information shocks, in particular market shocks, is allowed to influence returns in the pre and post shock period. It is therefore the response of investors to information shocks that alter returns leading to an increase in correlation.

To examine the role of market wide information further we estimate the smooth transition model using a measure of market sentiment without the market adjustments. This allows us to capture the effects of information trading on the order-flow imbalance. We calculate market sentiment as the average of the D-OFIL values across all stocks. We do not extract any market factors from this series. When we estimate the smooth transition model we find that the pre shock correlation is 0.2067 rising to 0.5100 following a shock. This suggests that it is the arrival of new market information that leads to a rise in post-shock correlations that is noted in previous studies that examine how return shocks influence correlation.

6.2 Influence of Existing Sentiment

Panel B of Table 4 shows that the magnitude of the post shock change in conditional correlation and the direction of the asymmetry is sensitive to whether the pre shock $OFIL_{st}$ value is >1 or <1. Partitioning the firms in this way suggests that the pre-shock level of sentiment, influences the size of changes to conditional correlation in the post shock period. When the pre-shock $OFIL_{st}$ ratio is >1 and sentiment is optimistic positive shocks to $OFIL_{st}$ generally lead to larger changes in conditional correlation than negative shocks but when the pre-shock $OFIL_{st}$ ratio is <1 and sentiment is pessimistic negative shocks lead to larger changes in conditional correlation than positive shocks to $OFIL_{st}$. This suggests that shocks that re-enforce existing sentiment lead to larger post-shock changes in conditional correlations than shocks that are not re-enforcing.

When the pre shock total-$OFIL_{st}$ ratio is >1 positive shocks to total-$OFIL_{st}$ lead to a 14.94% fall in average conditional correlation but average conditional correlation decreases by only 9.79% after a negative shock to retail-$OFIL_{st}$. When the pre-shock total-$OFIL_{st}$ is <1 positive shocks to total-$OFIL_{st}$ lead to a 9.82% fall in average conditional correlation but following negative shocks to total-$OFIL_{st}$ there is a 23.34% reduction in conditional correlation. This shows that when sentiment is optimistic positive shocks to total-$OFIL_{st}$ have a larger impact on average conditional correlation reduction than negative shocks to total-$OFIL_{st}$ but when sentiment is pessimistic negative shocks to total-$OFIL_{st}$ lead to larger reductions in average conditional correlation than positive shocks to total-$OFIL_{st}$.

We also find that the pre-shock level of sentiment also influences the direction of the asymmetry for the different investor groups. When the pre-shock retail-$OFIL_{st}$ is >1 positive shocks to retail-$OFIL_{st}$ lead to a 26.57% reduction in conditional correlation but when shocks are negative there is a 15.36% reduction in conditional correlation. When sentiment is pessimistic and the pre-shock retail-$OFIL_{st}$ is <1 positive shocks to retail-$OFIL_{st}$ lead to a 25.94% reduction in conditional correlation but a 28.85% reduction following a negative shock to retail $OFIL_{st}$.

When the pre-shock stealth-$OFIL_{st}$ ratio is >1 the average changes in conditional correlation following a shock to stealth-$OFIL_{st}$ shows a large asymmetry in which positive shocks to stealth-$OFIL_{st}$ lead to a 13.26% reduction in correlation but negative shocks to stealth-$OFIL_{st}$ only lead to a 3.89% reduction in average conditional correlation. However, when stealth-$OFIL_{st}$ is <1 the asymmetry is reversed as positive shocks lead to a 13.41% reduction in average conditional correlation but negative shocks to stealth $OFIL_{st}$ lead to a 19.15% reduction in average conditional correlation. When large-$OFIL_{st}$ is >1 positive shocks to large-$OFIL_{st}$ lead to slightly larger changes in conditional correlation than negative shocks. However, when large-$OFIL_{st}$ is <1 the asymmetry is not reversed as positive shocks to large-$OFIL_{st}$ lead to a 7.40% fall in conditional correlation and negative shocks to large-$OFIL_{st}$ lead to a 3.16% fall in conditional correlation. This suggests that only for retail and stealth investors are shocks to sentiment always re-enforcing.

6.3 S&P500 Index Constituents

Table 5 provides the conditional correlation results from the smooth transition model separately for portfolios which are constructed from S&P500 index constituents and non S&P500 index constituents respectively. We make this segregation because the S&P500 index is an important habitat for some institutional investors and membership has been shown to be associated with elevated levels of sentiment, see for example Barberis et al (2008). Moreover, preliminary results presented in Table 1 and 2 suggested that there was more positive sentiment and greater co-ordination of trading activity for S&P500 index constituents than for non S&P500 index constituents. Panel A provides the smooth transition conditional correlation results associated with shocks to total-$OFIL_{st}$ while Panel B provides the results for shocks to retail, stealth and large institutional $OFIL_{st}$.

The conditional correlations prior to a shock to total-$OFIL_{st}$ indicate that portfolios drawn from S&P500 index constituents have higher conditional correlations than portfolios drawn from non S&P500 index constituents. This suggests that S&P500 index constituents have higher levels of co-movement than non S&P500 index constituents in general and is consistent with the findings of Harford and Kaul (2005) who compare return unconditional correlations of S&P500 index constituents and non S&P500 index constituents[29]. Higher levels of co-movement for portfolios drawn from S&P500 index constituents is also evident for the models estimated using retail, stealth and large institutional $OFIL_{st.}$

Positive shocks to total-$OFIL_{st}$, for portfolios drawn from S&P500 index constituents, leads to a 20.07% fall in average conditional correlation while negative shocks lead to a 24.07% fall in conditional correlation. In

[29] The frequency used by Harford and Kaul (2005) is 15minute returns which show that the average pairwise unconditional correlation between securities within the S&P500 index during 1996 is 0.117 but 0.017 for non S&P500 constituents.

contrast, positive shocks to total $OFIL_{st}$ for portfolios drawn from non-S&P500 index constituents lead to a 19.37% fall in conditional correlation on average but a 30.30% average fall in conditional correlation if the shock to total-$OFIL_{st}$ is a negative one. This shows that positive shocks to total-$OFIL_{st}$ have a similar impact on S&P500 and non S&P500 portfolios but negative shocks to total -$OFIL_{st}$ have a larger impact on portfolios comprising of non S&P500 index constituents. Moreover, we also find that the asymmetry in which positive shocks to total-$OFIL_{st}$ lead to a smaller change in correlation than negative shocks is apparent for portfolios comprising of S&P500 index constituents and for portfolios comprising of non S&P500 index constituents.

In panel B of Table 5 we present the conditional pre and post shock correlations associated with portfolios drawn from S&P500 and non S&P500 index constituents for retail, stealth and large institutional traders. We find that shocks to retail-$OFIL_{st}$ lead to larger falls in average conditional correlation when portfolios comprise of S&P500 index constituents than when portfolios consist only of non S&P500 index constituents. The fall in average conditional correlation for portfolios comprising of S&P500 index constituents following a positive shock to retail-$OFIL_{st}$ is 22.47% but a fall of 24.68% arises following negative shocks to retail-$OFIL_{st}$. Changes in conditional correlation following shocks to retail-$OFIL_{st}$ when portfolios comprise of non S&P500 index constituents are smaller. Following positive shocks to retail-$OFIL_{st}$ there is actually a small rise rather than a fall in average conditional correlation. For portfolios drawn from both S&P500 and non S&P500 index constituents we find that shocks to retail-$OFIL_{st}$ lead to larger reductions in conditional correlation if the shock is negative confirming the asymmetry we identified earlier.

For portfolios comprising of S&P500 index constituents positive and negative shocks to stealth $OFIL_{st}$ lead to approximately a 10% and a 8% fall in post shock conditional correlations respectively. While shocks to stealth-$OFIL_{st}$ for non S&P500 constituents lead to a fall in conditional correlation of approximately 10% following positive or negative shocks to stealth-$OFIL_{st}$. These changes are therefore of a similar magnitude for S&P500 and for non S&P500 constituents. Positive shocks to large-$OFIL_{st}$ lead to a fall of 8.90% in conditional correlation if portfolios comprise of S&P500 index constituents and a 15.03% fall in conditional correlation if the shock to large-$OFIL_{st}$ is negative. Shocks to non S&P500 constituents have a smaller impact as positive shocks to large-OFIL lead to a 3.65% fall in average conditional correlation while negative shocks to large-$OFIL_{st}$ have no impact on average conditional correlation. This shows that changes to conditional correlation following shocks are generally larger for S&P500 index constituents than for non S&P500 index constituents.

6.4 Overall Analysis of Smooth Transition Results.

This section has shown that positive and negative shocks to order flow imbalance which capture shocks to sentiment lead to economically large reductions in the conditional correlation between portfolio returns and the market return in the post shock period. This indicates that sentiment changes that increase sentiment and shocks that reduce sentiment both lead to reductions in co-movement. In general we find that negative shocks lead to larger reductions in co-movement than positive shocks.

Another key discovery of the STCC GARCH model is that shocks to the sentiment of retail investors leads to larger reductions in co-movement than do shocks to the sentiment of institutional stealth investors or large institutional traders. This is consistent with our findings in Table 3 and those of Kumar and Lee (2006) that showed that changes to retail sentiment were negatively correlated with the market return. With the use of the shocks obtained from the STCC GARCH model we have shown that shocks to the sentiment of stealth and large institutional traders also lead to reduced co-movement. This was not apparent when we calculated the average pairwise correlations between the portfolio D-OFIL of traders and the market return. This difference probably stems from the advantage of the STCC GARCH model which only uses shocks that are above the threshold causing these shocks to be less noisy than changes.

Having shown that changes in sentiment lead to a reduction in co-movement we next examine the firm characteristics that are associated with sentiment shocks. We undertake this analysis because we have also shown that the level of sentiment is related to firm characteristics. Moreover, being able to identify the characteristics of firms that are associated with sentiment changes will allow investors to build portfolios that will be most able to take advantage of the conditional correlation reductions we have identified.

7. Firm Characteristics and Sentiment Changes

In this section we explore the influence that firm characteristics play in influencing the change in sentiment. We do this because different firms are likely to be influenced to differing degrees by sentiment because of the difficulties and costs associated with arbitrage, prevalence of investor habitats and diverse investors being influenced by sentiment differently and in a non-synchronous way.

Because we wish to identify firm level characteristics we return to firm level analysis of OFIL. We estimate a panel regression with fixed-effects in which the firm level D-OFIL is regressed against the firm characteristics we have examined. D-OFIL is the change in order flow imbalance and is used here as a proxy for order flow shocks. Our aim with this analysis is to gauge whether firm characteristics are likely to influence changes in the level of investor sentiment for a firm[30]. We estimate the regression separately for total D-OFIL, retail D-OFIL, stealth D-OFIL and large D-OFIL to establish how the characteristics of individual firms influences changes to the sentiment of specific investor groups. Since we have shown that positive and negative shocks influence conditional correlation differently we estimate the model separately for positive and negative changes to D-OFIL[31]. The model we estimate is described by equation (12).

$$\text{D-OFIL}_{itm}=\beta_1 + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \beta_5 SIZE_{it} + \beta_6 ILLIQ_{it} + \beta_7 PRICE_{it} + \beta_8 B/M_{it} + \beta_9 EPS_{it} + \beta_{10} EG_{it} +$$

$$\beta_{11} AGE_{it} + \beta_{13} CM_{it} + \beta_{14} OWN_{it} + \beta_{15} INT_{it} + \beta_{16} SP_{it} + \varepsilon_{it} \qquad (12)$$

[30] We are unable to undertake this analysis on the portfolio level D-OFIL as at a portfolio level firm characteristics diversify away.
[31] We partition firms according to whether the change in the order flow imbalance is positive or negative. We then estimate the regression separately for positive and negative samples.

where D-OFIL$_{itm}$ is the change in the order flow imbalance of investor group *m* (total, retail, stealth or institutional) in month *t* for firm *i*. β_1 is a constant. SMB, HML and UMD are Fama-French risk factors, since OFIL has already extracted the influence of (R_{mt}-R_{ft}) we do not include this again. We include the Fama-French factors as additional controls for market wide information as they are important determinants of expected returns and may capture some changes to order flow imbalance that are associated with market-wide information that we have not already captured. They therefore control for changes in order flow that are unrelated to sentiment. Size, ILLIQ, PRICE, B/M, EPS, EG, AGE, CM, OWN, INT are all characteristics which we have shown previously may influence the sentiment of investors. Size and ILLIQ are features that make some stocks more difficult to arbitrage. The low price of some stocks, EPS, earnings growth, age, level of intangibles are all characteristics that can make some stocks more difficult to value. Size, book-to-market, investor ownership, and correlation with the market are all characteristics that may be associated with the habitat of specific investors. We also include a dummy variable which has a value of unity if a firm is a member of the S&P500 index but has a value of zero otherwise(SP). We include this variable to determine whether membership of the S&P500 index influences sentiment after we have incorporated and therefore controlled for firm characteristics.

Results are contained in Table 6. These show that for each investor group positive and negative shocks are always influenced by the age of the firm in terms of its exchange listing and the EPS of the firm. These appear to be the only characteristics that exert an independent effect on sentiment changes across each investor group. However, we also find that specific characteristics are associated with specific investor groups. This diversity may reflect differences in the habitats of the different investors and the relative importance of different firm characteristics for the determination of sentiment and co-movement.

In addition to AGE and EPS positive and negative changes to total D-OFIL are influenced by INT, OWN, and price. While positive shocks only are influenced by earnings growth and B/M while negative changes are influenced by illiquidity. For retail investors there are greater differences between which firm characteristics influence positive and negative shocks. No additional characteristics to age and EPS influences both positive and negative shocks. OWN, EG and PRICE influence positive changes to D-OFIL only and Illiquidity influences negative shocks only. For stealth investors no other firm characteristics influence changes in D-OFIL other than AGE and EPS. In the case of large D-OFIL as well as the influence of AGE and EPS illiquidity influences positive and negative changes in D-OFIL while earnings growth and intangibles influence positive D-OFIL changes only.

8. Robustness Tests

8.1 Sub-Period Analysis

We next estimate the smooth transition model over different sub-periods. The first covers the period 1993-2000 and the second 2001-2011. These have been selected to cover the period prior to decimalization and subsequent to decimalization. Separating these two periods is important as the post decimalization period is associated with a reduction in trading costs and raised trading activity, see for example Bessembinder (2003) and also coincides with a rise in trading activity from high frequency traders.

The results from the estimation of the smooth transition model over these two periods is contained in Table 7 and show that our key finding that sentiment shocks (shocks to $OFIL_{st}$) lead to reduced co-movement is maintained across both sub-periods. However, it is also obvious that shocks to sentiment have slightly larger effects on co-movement in the period 1993-2000. During the period prior to substantial high frequency trading positive shocks to total-$OFIL_{st}$ lead to nearly a 15% fall in average conditional correlation while negative shocks lead to just over a 16% fall in average conditional correlation. In the period of high frequency trading positive shocks to total-$OFIL_{st}$ lead to a 6% fall in average conditional correlation while negative shocks to total-$OFIL_{st}$ lead to approximately a 12% fall in average conditional correlation. However, in both periods shocks to total-$OFIL_{st}$ lead to lower average conditional correlations in the post shock period and the discovery that positive shocks leads to smaller changes in average conditional correlations than negative shocks is robust across the two periods.

The slightly reduced effect of order flow imbalance shocks on co-movement is also evident in the results for retail, stealth and institutional traders. Retail investors exhibit the greatest changes to co-movement in response to their $OFIL_{st}$ shocks but in the period of high frequency trading the changes in co-movement are smaller. Even though the change in co-movement is smaller for retail investors in the later period on average positive and negative shocks to retail-$OFIL_{st}$ still lead to changes in co-movement of nearly 10% in response to positive retail-$OFIL_{st}$ shocks and over 11% in response to negative $OFIL_{st}$ shocks. Stealth investors experience the greatest differences across the two periods as the effect of positive shocks to stealth $OFIL_{st}$ leads to an average reduction in conditional correlation of -11.44% in the first period but a -5.57% fall in the second. While negative shocks to stealth-$OFIL_{st}$ lead to an average fall in conditional correlation of 14.81% in the first period and a fall of 5.96% in the second period. The effect of positive shocks to large-$OFIL_{st}$ in the earlier period is about -7% and this is almost unchanged in the later period. In the earlier period negative shocks to large-$OFIL_{st}$ lead to a 12% reduction in co-movement while in the later period negative shocks to large $OFIL_{st}$ have no statistical effect on co-movement.

8.2. Different Return Horizons

We next estimate the smooth transition model using different return intervals to establish the robustness of our results using shocks to order flow captured over one day. We do this by estimating the smooth transition model using two-day and five-day returns and measuring the shock using the D-OFIL value over two and five days respectively. As before the smooth transition model analyses positive and negative shocks to order flow sentiment separately. We report the results using Total D-OFIL, Retail D-OFIL, Stealth D-OFIL and Large D-OFIL in Table 8. A finding that the reductions in co-movement weaken using longer return intervals would suggest that the effects of shocks to sentiment contain a transitory component while larger declines in co-movement following shocks would suggest the effect of the shock is amplified over longer horizons.

The results associated with shocks to Total-D-OFIL suggest that at a two day return horizon the magnitude of the fall in correlation following a shock to order flow sentiment is substantial and comparable to that of one day returns. Positive shocks at a five day interval lead to a slight rise in correlation. However, negative shocks lead to a larger fall in correlation at a five day horizon. For retail sentiment we find broadly similar results at two and five day intervals (positive shocks lead to slightly larger correlation reductions). For stealth investors the effect of shocks to their sentiment tend to diminish slightly as return horizons increase. For shocks to large trader sentiment positive and negative shocks lead to slightly larger decreases in correlation at a two-day and five-day return interval. This shows that except for five day positive shocks to Total sentiment and negative shocks to Stealth sentiment the results at different intervals are broadly similar. This suggests that negative shocks to sentiment across all investors tends to be longer lasting.

9. Conclusion

In this paper we studied the impact that shocks to sentiment, measured through shocks to order flow imbalance, have on co-movement. We are motivated to undertake the analysis because the paper by Kumar and Lee (2006) which documents a link between the order flow levels of retail investors and the portfolio returns of small investors raised a number of unanswered questions.

The replication of some of the Kumar and Lee analysis shows that the sentiment of institutional investors as well as the sentiment of retail investors, is an important determinant of stock returns. This suggests that both retail and institutional investors display sentiment that may influence co-movement. A problem with the analysis of Kumar and Lee (2006) is that they do not directly model the effect of sentiment or sentiment changes on co-movement. This motivates us to estimate the STCC GARCH model which provides direct measures of how correlations between a portfolio and the market change in response to an order flow shock.

Our results show that positive and negative shocks to total order flow reduce co-movement. Positive shocks to the order flow imbalance levels of retail traders has a larger impact on co-movement changes than is the case for shocks to the order flow imbalance of stealth and large traders. We discover an asymmetry in how positive and negative shocks to sentiment influence co-movement. Positive shocks have a smaller influence on co-movement than negative shocks, an asymmetry that is evident for total shocks, retail shocks and stealth shocks but is reversed for shocks associated with large institutional traders.

We examine the firm characteristics which have been associated with sentiment and find that a range of these are correlated with shocks to order flow imbalance. This suggests that order flow shocks are linked to firm characteristics which suggests that sentiment changes and their relationship to co-movement changes is influenced by firm level features.

References

Amihud, Yakov 2002, Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

Ang, Andrew, and Joe Chen. 2002, Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics*, 63(3): 443-494.

Antoniou, C, Doukas, J.A and A. Subrahmanyam, 2015, Investor sentiment, beta and the cost of equity capital, Management Science, 347-367.

Arbel, A. Strebel, P. J. 1982, The neglected and small firms effects. Financial Review, 17, 201—218.

Bagnoli, M. Clement, C., Crawley,M., and Watts S. 2014, The relative profitability of analyst stock recommendations. What role does investor sentiment play? Unpublished working paper University of Purdue.

Baker, M. and Wurgler J. 2007, Investor sentiment in the stock market, Journal of Economic Perspectives, American Economic Association, 21 (2) 129-152.

Barber, B.M. and Odean T., 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, Journal of Finance 55 773-806.

Barber, B. and Odean T. 2008, All that glitters: The effect of attention and news on the buying behaviour of individual and institutional investors, The Review of Financial Studies, 21, 785-818.

Barber, B.M., Odean T., and Zhu, N. 2009, Do retail trades move markets?, The Review of Financial Studies, 22 151-186.

Barberis, N. Shleifer, A. Wurgler, J. 2005, Co-movement. Journal of Financial Economics 75, 283-317.

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

Barclay, M.J.,Warner, J.B. 1993, Stealth trading and volatility: which trades move prices? Journal of Financial Economics 34, 281–306.

Berben, R-P. Jansen, W.J. 2005, Co-movement in international equity markets: A sectorial view. Journal of International Money and Finance 24, 832-857.

Bessembinder, H. 2003, Trade execution costs and market quality after decimalization, Journal of Financial and Quantitative Analysis,

Black, F. 1986, Noise, Journal of Finance, VI 529-543.

Blume, M., and MacKinlay C. and Terker B. 1989, Order imbalances and stock price movements on October 19 and 20 1987. Journal of Finance 44, 827-848.

Bollerslev, T. 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. Review of Economics and Statistics 72, 498-505.

Bollerslev, T. Wooldridge, J.M. 1992 Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, Econometric Reviews 1, 143—173.

Boyer, M. Chapelle, A. Szafarz, A. 2008, No contagion, only globalisation and flight to quality. Centre Emile Bernheim Working paper No 08/018.

Broggard, J, Hendershott T. and Riordan R. 2014, High-Frequency Trading and Price Discovery. Review of Financial Studies.

Brown, Gregory W. and Cliff, M. T. 2005, Investor Sentiment and Asset Valuation. Journal of Business, 78(2): 405–40.

Campbell, John, and Hentschel, L. 1992, No news is good news. An asymmetric model of changing volatility in stock returns. Journal of Financial Economics, 31, 281-318.

Campbell, John, Y., Ramadorai T. and Schwartz, A. 2009, Caught on tape: institutional trading stock returns and earnings announcements, Journal of Financial Economics, 92 66-91.

Chakravarty, S. 2001, Stealth-trading: which traders' trades move stock prices? Journal of Financial Economics 61, 289–307.

Chan, W. S. 2003, Stock price reaction to news and no-news: Drift and reversal after headlines. Journal of Financial Economics 70, 223-260.

Chelley-Steeley, P.L. Lambertides, N. Savva, C.S 2013, Illiquidity shocks and the co-movement between stocks; New evidence using smooth transition, Journal of Empirical Finance 23 p1-15.

Chordia, T., Roll, R. Subrahmanyam, A. 2002, Market liquidity and trading activity. Journal of Finance 56, 501-530. 111-130.

Chordia, T. Subrahmanyam, A. 2004 Order imbalance and individual stock returns, Journal of Financial Economics 72, 485-518.

Chordia, T., and Hu, J. Subrahmanyam A. and Tong Q. 2015, Order Flow and Equity Costs of Capital, Unpublished working paper. SSRN. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2428041

Christie, A. 1982, The stochastic behaviour of common stock variances: value,leverage and interest rate effects. Journal of Financial Economics,10 407-432.

Cohen, R., Gompers, P. Vuolteennaho, T. 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. Journal of Financial Economics. 66, 409-462.

Conrad, J.S. Johnson. K.M. Wahol, S. 2001, Institutional trading and soft dollars. Journal of Financial Economics 50 397-416.

Congressional report 2014, High frequency trading background, concerns, and regulatory developments. Prepared by G. Shorter, R. Miller.

Cushing, D. and Madhavan A. 2000, Stock returns and trading at the close. Journal of Financial Markets, 3 45-67.

D'Avolio, G. 2002, The market for borrowing, Journal of Financial Economics, 66 271-306.

DeVault, L.A. Sias, R.W. and Starks L.T. 2016, Who are the sentiment traders? Evidence from the cross-section of stock returns. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2539858

Dorn, D., Huberman Gur, Sengmueller P. 2008, Correlated trading and returns, Journal of Finance 58 885-920.

Durnev, A. Morck, R. Yeung B. 2004, Value enhancing capital budgeting and firm specific stock return variation. Journal of Finance 59, 65-105.

Easley, D. O'Hara, M. 1987, Price, Trade Size and Information in Securities Markets, Journal of Financial Economics 19, 69-90.

Eitrheim, O. and Terasvirta, T. 1996, Testing the adequacy of smooth transition autoregressive models. Journal of Econometrics 74, 59-76.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Forbes, K.J., Rigobon, R., 2002. No contagion, only independence: measuring stock market co-movements. Journal of Finance 57, 2223-2261.

Franses, P. H., van Dijk, D., 2000. Non-linear time series models in empirical finance. Cambridge University Press.

Geczy, C. Musto, D. Reed, A. 2002, Stocks are special too: an analysis of the equity lending market. Journal of Financial Economics 66, 241–269.

Gemmill, G., 1996, Transparency and liquidity, A study of block trades on the LSE under different publication rules. Journal of Finance 51 1765-1790.

Gompers, P. and Metrick, A. 2001, Institutional investors and equity prices. Quarterly Journal of Economics, 116, 229-260.

Graham, John R. 1999, Herding among investment newsletters: theory and evidence. Journal of F'inance, 54(1), pp. 231-68.

Griffin, J.M. Harris, J.H. and Topaloglu, S. 2003 The dynamics of institutional and individual trading. Journal of Finance 58 2285-2320.

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

Hvidkjaer, S. 2006 A trade based analysis of momentum, Review of Financial Studies, 19 457-491.

Hvidkjaer, S. 2008 Small Trades and the Cross-Section of Stock Returns, Review of Financial Studies 21:1123-1151.

Harford, J. Kaul A. 2005 Correlated order flow pervasiveness, sources, and pricing effects, Journal of Financial and Quantitative Analysis 40, 29-55.

Hasbrouck, J. 1991 Measuring the information content of stock trades. Journal of Finance 46, 179-207.

Hasbrouck, J. Seppi, D. 2001. Common factors in prices, order flows and liquidity. Journal of Financial Economics 59, 383-411.

Hong, H, T. Lim Stein J.C. 2000 Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55, 265-295.

Hong, Yongmiao, Jun Tu, and Guofu Zhou. 2007. Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation. *Review of Financial Studies*, 20(5): 1547-1581.

Kaniel, R.S. Saar, G. and Titman, S. 2008 Individual investor trading and stock returns, Journal of Finance 63 273-310.

Keim, D. and Madhavan A. 1996. The upstairs market for large block transactions: Analysis and measurement of price effects. Review of Financial Studies 9 1-36.

Kelley, E.K. and Tetlock, P.C. 2013, How wise are crowds? Insights from retail orders and stock returns. Journal of Finance, 68 1229-1265.

Kapetanios, G. Tzavalis, E. 2010, Modeling structural breaks in economic relationships using large shocks. Journal of Economic Dynamics & Control 34, 417-436.

Kearney, C. and Poti, V. 2006, Correlation dynamics in European equity markets. Research in International Business and Finance 20, 305-321.

Klemkosky, R.C. 1977, The impact and efficiency of institutional net trading imbalances, Journal of Finance, 32, 79-86.

Koop, G. and Potter, S. 2007, Estimation and forecasting in models with multiple breaks. Review of Economic Studies 74, 763–789.

Kothari S., Lewellyn, J. and Warner, J. 2006, Stock returns, aggregate earnings surprises, and behavioral finance. Journal of Financial Economics 79, 537–568.

Kraus, A., and Stoll, H.R. 1972. Price impacts of stock trading on the New York Stock Exchange. Journal of Finance 569-588.

Kumar, A., and Lee, C.M.C. 2006 Retail investor sentiment and return co-movements. Journal of Finance 61 2451-2486.

Kyle, A. 1985 Continuous auctions and insider trading, Econometrica, 53 1315-1335.

Lakonishok, J. Shleifer, A. and Vishny R.W. 1992 The impact of institutional trading on stock prices, Journal of Financial Economics 32 23-34.

Lee, C.M.C. and Radhakrishna B. 2000 Inferring investor behaviour: Evidence from TORQ data, journal of Financial Markets, v3 2 183-204.

Lee, C.M.C. and Ready M..J. 1991 Inferring trade direction from intraday data , Journal of Finance 46, 733-746.

Lee, Charles, Shleifer A. and Thaler R.H. 1991 Investor sentiment and the closed-end fund puzzle. Journal of Finance, 46(1): 75–109.

Li, C.A., Philippopoulos, A., Tzavalis, E., 2000 Inflation and exchange rate regimes in Mexico, Review of Development Economics 4, 87-100.

Menkveld, A. 2011 Foresight driver review, Electronic Trading and Market Structure", available from So
cial Science Research Network, SSRN abstract = 1986892

Mikkelson, W. and Partch M. M. 1985 Stock price effects and costs of secondary distributions. Journal of Financial Economics 165-194.

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

Odders-White E. and Ready, M. 2008, The probability and magnitude of information events.
Journal of Financial Economics, 87 227-248.

Pirinsky, C. Wang Q. 2006 Does corporate headquarters matter for stock returns? Journal of Finance 61(4) 1991-2015.

Roll, R., 1988 R2, Journal of Finance, 43, p541-567.

Savva C.S. 2011 Modelling interbank relations during the international financial crisis. Economics Bulletin 31, 916-924.

Savva, C.S. Neanidis, K. Osborn, D.R. 2010 Business cycle synchronization of the Euro area with the new and candidate member countries. International Journal of Finance and Economics 15, 288-306.

Scharfstein, David S. and Stein, Jeremy C, 1990 Herd behavior and investment. American Economic Review, June 1990, 80(3), pp. 465-79.

Schmeling M. 2007 Institutional and individual sentiment: Smart money and noise trader risk?. International Journal of Forecasting 23(1):127-145.

Shleifer, Andrei, and Vishny, R. 1997. The limits of arbitrage. Journal of Finance, 52 35–55.

Silvennoinen A., Teräsvirta, T., 2005. Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations, SSE/EFI Working Paper Series in Economics and

Silvennoinen, A., Teräsvirta, T., 2009. Modelling multivariate autoregressive conditional heteroskedasticity with double smooth transition conditional correlation GARCH model. Journal of Financial Econometrics 7, 373-411.

Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288-302.

Ülkü, N. and Weber, E. 2013, Identifying the interaction between stock market returns and trading flows of investor types; Look into the day using daily data, Journal of Banking and Finance, 37980 2733-2749.

Vuolteenaho, T. 2002, What drives firm-level stock returns? Journal of Finance 57, 233-264.

Wermers, R. 1999, Mutual fund herding and the impact on stock prices, Journal of Finance, 54 581-622.

Wurgler, J.2000, Financial markets and the allocation of capital. Journal of Financial Economics 58, 187-214.

Table 1: Order Flow Imbalances

Panel A and B of this table presents the average daily order flow imbalance level (OFIL) associated with sample stocks. OFIL for security *i* on day *t* is calculated as the ratio of the aggregate dollar value of buyer initiated trades to the aggregate dollar value of seller initiated trades less the market return. The values reported in each table are averaged across days and securities. Total refers to the total-OFIL which is calculated using all trades, Retail refers to retail-OFIL and is calculated using only buyer and seller initiated trades classified as small, stealth captures stealth-OFIL and is calculated from medium sized trades, Large refers to the large-OFIL calculated from only the largest trades. All firms provides averages across all firms, S&P500 and non S&P500 partitions firms according to whether or not firms are S&P500 index constituents. Order imbalance >1 and <1 are samples which are segregated according to whether the order imbalance ratio on day *t* is <1 or >1. In Panel B firms are sorted into one of five groups based on the magnitude of firm characteristics. Group 1 contains firms with the smallest characteristic value and group 5 contains firms with the largest characteristic value. The firm characteristics we use are SIZE(market value), ILLIQ (the Amihud illiquidity ratio), B/M(book-to-market ratio), PRICE(closing price), AGE(years listed on the exchange), CM(correlation between the daily return of stock *i* and the market), OWN(proportion of institutional investor ownership), INT(proportion of intangible asset values to total asset value), EPS (earnings per share) and EG(earnings growth).

Panel A:Average Order Flow Imbalance Levels(OFIL).

	All Firms	S&P500	Non SP500	order imbalance >1	order imbalance <1
Total	1.053	1.158	1.045	1.585	0.636
Retail	1.039	1.028	1.040	1.509	0.665
Stealth	1.047	1.077	1.046	1.427	0.702
Large	1.133	1.204	1.128	1.815	0.588

Panel B:Average Order Flow Imbalance Level(OFIL) by Firm Characteristic.

Total	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.9038	1.114	1.0622	0.8905	0.9511	1.0088	0.9794	1.0353	1.0961	1.0558
2	0.9356	1.089	1.0791	0.987 4	0.9760	1.0578	0.9824	1.0490	1.0933	1.0589
3	1.0203	1.0476	1.0648	1.0491	0.9984	1.0657	1.0403	1.0659	1.0762	1.0678
4	1.0872	0.9863	1.0426	1.0825	1.0305	1.0561	1.0804	1.0608	1.0461	1.0858
5 High	1.1224	0.9544	0.9936	1.1186	1.0480	1.0215	1.1091	1.0605	0.9925	1.0955

Retail	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.9524	1.0156	1.0492	0.9368	0.9520	1.0250	1.0107	1.0480	1.0723	1.0297
2	0.9966	1.051	1.0586	1.0134	0.9750	1.0552	1.0094	1.0354	1.0621	1.0379
3	1.0554	1.0569	1.0479	1.0511	0.9952	1.0458	1.0214	1.0444	1.0470	1.0546
4	1.0646	1.0331	1.032	1.0626	1.0264	1.0252	1.0460	1.0406	1.0367	1.0686
5 High	1.0454	1.015	1.0022	1.0722	1.0412	1.0065	1.0664	1.0415	1.0357	1.0794

Stealth	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.9809	1.0558	1.0601	0.9427	0.9986	1.0259	1.0206	1.0439	1.0768	1.0331
2	0.9639	1.0683	1.0628	1.0053	1.0110	1.0539	0.9964	1.0385	1.0703	1.0368
3	1.0252	1.0536	1.0466	1.0369	1.0216	1.0471	1.0119	1.0512	1.0515	1.0536
4	1.0658	1.0117	1.0304	1.0545	1.0397	1.0377	1.0518	1.0526	1.0234	1.0722
5 High	1.0708	0.992	1.017	1.078	1.0464	1.0269	1.0872	1.0574	1.0037	1.0776

Large	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG

1 Low	1.4909	1.171	1.1523	1.1248	1.0762	1.1248	1.1496	1.1285	1.1543	1.1416
2	1.0241	1.1472	1.1417	1.0822	1.0900	1.1198	1.1054	1.1357	1.1502	1.1334
3	1.051	1.1075	1.1326	1.1034	1.1024	1.1299	1.1176	1.1440	1.1307	1.1287
4	1.1385	1.0531	1.1274	1.138	1.1204	1.1231	1.1260	1.1321	1.1115	1.1424
5 High	1.1692	1.0494	1.1115	1.1634	1.1300	1.1156	1.1530	1.1363	1.0819	1.1604

Table 2: Correlation of Order Flow Activity.

This table presents the average pairwise unconditional correlation between the change in order-flow imbalance level (D-OFIL) of stock *i* and stock *j*. Total refers to the correlation between the D-OFIL of firm *i* and *j* when OFIL has been calculated using all trades. Retail presents average pairwise correlation results between the retail D-OFIL of stock *i* and *j* (small trade classification). Stealth reflects the average pairwise correlation in D-OFIL of stocks when OFIL is based on the medium size trade classification. Large refers to the correlations between the D-OFIL of firms when OFIL is calculated using only large trades. In Panel A All Firms refers to a sample in which stocks *i* and *j* are drawn from all firms available, S&P500 and non S&P500 samples partition S&P500 index and non-index constituents. OFIL >1 and <1 are samples which are segregated according to whether trader OFIL on day *t is* > or *<1*. In Panel B firms are sorted into one of five groups based on the magnitude of firm characteristics. These are SIZE(market value), ILLIQ (the Amihud illiquidity ratio), B/M(book-to-market ratio), PRICE (closing price), AGE (years listed on the exchange), CM(correlation between the daily return of stock *i* and the market), OWN(proportion of institutional investor ownership), INT(proportion of intangible assets), EPS (earnings to price ratio) and EG(earnings growth). Panel C provides the results from regressing portfolio i's correlation with the market against the average of the firm characteristics within a portfolio.

Panel A:Average Pairwise Correlation Between D-OFIL of Firm *i* and *j*.

	All Firms	S&P500	Non S&P500	OFIL >1	OFIL <1
Total	0.038	0.016	0.049	0.052	0.185
Retail	0.095	0.234	0.087	0.102	0.178
Stealth	0.056	0.033	0.071	0.052	0.161
Large	0.053	0.136	0.049	0.056	0.100

Panel B: Correlation Between D-OFIL of Firm *i* and *j* by Firm Characteristic.

Total	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.0809	0.0597	0.0543	0.0547	0.1447	0.0371	0.0670	0.0521	0.0395	0.0557
2	0.0591	0.0385	0.0376	0.0374	0.0616	0.0433	0.0661	0.0309	0.0446	0.0559
3	0.0658	0.0506	0.0552	0.0559	0.0568	0.0552	0.0562	0.0478	0.0343	0.0355
4	0.0690	0.0426	0.0599	0.0512	0.0400	0.0569	0.0516	0.0485	0.0627	0.0495
5 High	0.0234	0.0703	0.0740	0.0763	0.0539	0.0486	0.0434	0.0487	0.0656	0.0419

Retail	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.0873	0.1259	0.1148	0.1143	0.1513	0.0849	0.0763	0.0840	0.1561	0.0972
2	0.0795	0.1154	0.1074	0.1069	0.0885	0.1627	0.1053	0.0852	0.1477	0.0903
3	0.1037	0.1189	0.1085	0.1106	0.0928	0.1503	0.1139	0.1099	0.1171	0.1249
4	0.1626	0.0781	0.0735	0.0740	0.0905	0.1238	0.1285	0.1115	0.1019	0.1218
5 High	0.2009	0.1000	0.1064	0.1089	0.1346	0.0778	0.1731	0.1144	0.0709	0.1078

Stealth	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.1539	0.0628	0.0538	0.0545	0.2015	0.0569	0.0693	0.0820	0.0527	0.0905
2	0.0848	0.0535	0.0521	0.0521	0.1145	0.0741	0.0792	0.0459	0.0620	0.0954
3	0.1060	0.0565	0.0682	0.0645	0.1102	0.0526	0.0864	0.0560	0.0546	0.0522
4	0.1270	0.1159	0.0877	0.0888	0.0494	0.0831	0.0532	0.0521	0.0796	0.0493
5 High	0.0408	0.1210	0.1296	0.1373	0.0446	0.0580	0.0551	0.0547	0.0958	0.0481

Large	SIZE	ILLIQ	B/M	PRICE	AGE	CM	OWN	INT	EPS	EG
1 Low	0.4140	0.0665	0.0721	0.0704	0.2439	0.0869	0.0986	0.0717	0.0884	0.0856
2	0.2346	0.0694	0.0535	0.0521	0.0951	0.0807	0.0852	0.0442	0.0915	0.0573

3	0.1089	0.0697	0.0767	0.0723	0.0809	0.0692	0.0706	0.0773	0.0752	0.0730
4	0.0679	0.1027	0.0879	0.0874	0.0681	0.0907	0.0717	0.0792	0.0629	0.0870
5 High	0.1006	0.1386	0.1374	0.1433	0.0898	0.1233	0.0918	0.0858	0.0823	0.0818

Panel C: Multivariate Correlation Regressions

	TOTAL		RETAIL		STEALTH		LARGE	
	Coef	t-test	Coef	t-test	Coef	t-test	Coef	t-test
Constant	0.0810	2.99	0.0464	0.67	0.2412	6.90	-0.2469	-2.18
SIZE	0.0045	4.56	0.0012	0.46	-0.0004	-0.30	0.0144	2.56
ILLIQ	-0.0008	-2.88	-0.0038	-2.42	-0.0006	-0.45	0.0542	1.25
PRICE	0.0019	0.89	0.0015	1.23	0.0740	0.53	0.0382	1.04
B/M	0.0004	0.11	-0.0098	-1.20	-0.0041	-1.15	0.0082	0.51
EPS	0.0009	0.63	-0.0001	-0.03	0.0008	0.41	0.0000	-1.81
EG	0.0083	0.81	-0.0452	-1.39	-0.0249	-1.56	-0.0400	-0.74
AGE	-0.0252	-6.08	0.0406	3.85	-0.0415	-9.42	0.0127	0.77
OFIL	0.0043	1.73	-0.0322	-1.68	-0.0439	-4.73	0.0606	2.02
CM	0.0447	3.18	0.2926	7.56	0.1012	5.23	-0.3376	-3.31
OWN	-0.0065	-1.96	-0.0039	-0.46	-0.0089	-2.12	0.0174	1.43
INT	0.0130	2.42	0.0145	1.03	0.0148	2.22	0.0171	0.81
S&P	0.0057	0.51	0.0408	1.22	0.0196	1.45	0.0606	1.46
	0.23		0.21		0.30		0.07	

Table 3: Order Flow Imbalance and Returns

In Panel A of this table we present the coefficients from the estimation of the following regression equation. In this model R_{pt} are the returns to portfolio *p* in excess of the risk free rate (R_f), α and β_n are coefficients.

$R_{pt}-R_{ft}=\alpha+ \beta_1(R_{mt}-R_{ft})+\beta_2 SMB_t+ \beta_3 HML_t +\beta_4 UMD_t+ \beta_5 OFIL_{pt} +\xi_{pt}$

SMB is the return to the Fama-French small firm factor, HML is the return to the Fama-French book-to market factor and UMD is the return to the momentum factor. $OFIL_{pt}$ is the market return adjusted order flow imbalance level of portfolio p. The OFIL we use in the regressions are the Total, Retail, Stealth and Large OFIL values, we include each one in turn as well as estimating a version which includes retail, stealth and large OFIL. The *p* values are shown in parenthesis, *** indicates significance at a 1% level, ** at a 5% level and * at a 10% level. Panel B presents the average pairwise contemporaneous correlation coefficient between the change in order flow imbalance level (D-OFIL) on day *t* for firm *i* and the excess market return. In the column headed Total, Retail, Stealth and Large correlations are calculated from the D-OFIL of the different investor groups. Panel C presents the proportion of Ng-Perron non-stationarity tests that indicate the series is stationary. All stocks refers to the proportion of test results found to be stationary when the test is applied to individual stock level OFIL values. All portfolios refers to the proportion of OFIL values found to be stationary when the series being tested are portfolio order flow imbalance levels. Total, Retail, Stealth and Large in Panel C indicates which of the order flow imbalance levels are being tested.

Panel A: Relationship between Portfolio Returns and Investor OFIL

	Total	Retail	Stealth	Large	Full
Constant	>0.000	>0.000	>0.000	>-0.000	>0.000
	(0.34)	(-0.00)***	(0.58)	(-0.00)***	(-0.00)***
RMF	0.009	0.009	0.009	0.009	0.008
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
SMB	0.005	0.005	0.005	0.005	0.005
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HML	-0.004	-0.004	-0.004	-0.004	-0.004
	(-0.00)***	(-0.00)***	(-0.00)***	(-0.00)***	(-0.00)***
UMD	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Total-OFIL	0.017				
	(0.00.)***				
Retail-OFIL		0.107			0.080
		(0.00)***			(0.00.)***
Stealth-OFIL			0.030		0.013
			(0.06)*		(0.25)
Large-OFIL				0.091	0.070
				(0.00)***	(0.00)***
$\bar{R}$	0.945	0.945	0.945	0.940	0.940

Panel B: Average Pairwise Correlation between Investor D-OFIL and Market Excess Return

	Total	Retail	Stealth	Large
Cor	-0.030	-0.110	-0.073	0.058
p-value	0.04**	0.01***	0.280	0.126

Panel C: Results of Stationarity Tests on Stocks and Portfolios.

	All Stocks		All Portfolios	
	OFIL	D-OFIL	OFIL	D-OFIL
Total	3.40	100	5.08	100
Retail	3.75	100	4.88	100
Stealth	3.48	100	3.31	100
Large	2.37	100	2.70	100

Table 4: Smooth Transition Conditional Correlation Results

This table shows the average pairwise pre and post shock portfolio conditional correlation coefficient obtained from the STCC GARCH model. These are reported separately for positive shocks to $OFIL_{st}$ (+ shock) and negative shocks to $OFIL_{st}$(- shock). Total, Retail, Stealth and Large presents model results separately for shocks to total, retail, stealth and large $OFIL_{st}$. Pre shock cor is the average pairwise conditional correlation between portfolio p and the market prior to a shock to $OFIL_{st.}$ Post shock cor is the average pairwise conditional correlation between portfolio p and the market portfolio after a shock to $OFIL_{st}$. Δ cor is the average change in the post shock conditional correlation between the post shock and pre shock period. A + indicates that the average difference between the average conditional correlation change following positive shocks to $OFIL_{st}$ is statistically different to the average change in conditional correlation following negative shocks to $OFIL_{st}$ at a 5% level using a t-test. The t-test is a test of whether the change in conditional correlation following a shock to $OFIL_{st}$ is significant at a 5% level. % Δ in cor is the percentage change in average conditional correlation that takes place after a shock to $OFIL_{st}$. The % Δ in cor<0 captures the proportion of negative changes in conditional correlation as a percentage of all changes following a shock. In Panel A all firms presents the results for the full sample of firms while Panel B partitions the sample according to whether the trader OFIL in the pre-shock period was >1 or <1.

Panel A. All Firms

	Total		Retail		Stealth		Large	
	+ shock	- shock	+shock	-shock	+shock	-shock	+shock	-shock
Pre shock cor	0.1860	0.2094	0.3378	0.2843	0.3085	0.3040	0.3953	0.3358
Post shock cor	0.1693	0.1782	0.3037	0.2424	0.2836	0.2747	0.3676	0.3161
Δ cor	-0.0167	-0.0312+	-0.0341	-0.0419+	-0.0249	-0.0293+	-0.0277	-0.0197+
t-test	-22.55	-26.50	-26.25	-33.72	-22.99	-28.87	-19.05	-6.96
% Δ in cor	-8.98%	-14.90%	-10.09%	-14.74%	-8.07%	-9.64%	-7.01%	-5.87%
% Δ Cor <0	63.2%	64.0%	62.5%	73.5%	51.07%	66.9%	62.3%	49.7%

Panel B. Firms with pre-shock $OFIL_{st}$ > 1

	All Trades		Retail		Stealth		Large	
	+ shock	- shock	+shock	-shock	+shock	-shock	+shock	-shock
Pre shock cor	0.328	0.337	0.286	0.280	0.279	0.334	0.343	0.389
Post shock cor	0.279	0.304	0.210	0.237	0.242	0.321	0.316	0.366
Δ cor	-0.049	-0.033+	-0.076	-0.043+	-0.037	-0.013+	-0.027	-0.023+
t-test	-33.98	-17.61	-43.96	-24.23	-19.26	-13.39	-11.89	-12.73
% Δ in cor	-14.94%	-9.79%	-26.57%	-15.36%	-13.26%	-3.89%	-7.87%	-5.91%
% Δ in cor<0	54.7%	49.1%	69.8%	55.5%	70.6%	63.3%	45.6%	57.1%

Firms with pre-shock $OFIL_{st}$ <1

	All Trades		Retail		Stealth		Large	
	+ shock	- shock	+shock	-shock	+shock	-shock	+shock	-shock
Pre shock cor	0.275	0.287	0.293	0.312	0.261	0.235	0.392	0.412
Post shock cor	0.248	0.220	0.217	0.222	0.226	0.190	0.363	0.399
Δ cor	-0.027	-0.067+	-0.076	-0.090+	-0.035	-0.045+	-0.029	-0.013+
t-test	-5.04	-13.13	-2.08	-45.17	-7.76	-13.30	-7.74	-2.80
% Δ in cor	-9.82%	-23.34%	-25.94%	-28.85%	-13.41%	-19.15%	-7.40%	-3.16%
% Δ in cor<0	47.2%	71.3%	64.5%	66.3%	43.4%	58.5%	33.4%	42.3%

Table 5: Smooth Transition Results S&P500 Partitions

This table presents the average STCC GARCH pairwise pre and post shock portfolio conditional correlations for portfolios constructed from S&P500 and non S&P500 index constituents. These are reported separately for positive (+ shock) and negative shocks (- shock). Total, Retail, Stealth and Large presents model results separately for shocks to total, retail, stealth and large $OFIL_{st}$. Pre shock cor is the average pairwise conditional correlation between portfolio p and the market portfolio prior to shocks to $OFIL_{st}$. Post shock cor is the average pairwise conditional correlation after shocks to $OFIL_{st}$. Δ cor is the average change in the post shock conditional correlation between the post shock and pre shock period. The t-test row presents the t-value associated with the test of whether the change in conditional correlation following a shock is significant at a 5% level. A + indicates that the average difference between the conditional correlation change following a positive shock is statistically different to the average change in conditional correlation following a negative shock at a 5% level using a t-test. % Δ cor is the percentage change in average conditional correlation that takes place after an order flow shock. The % Δ in cor<0 captures the proportion of negative changes in conditional correlation as a percentage of all changes following a shock. In Panel A all firms presents the results for the full sample while Panel B and Panel C partitions the sample according to whether the order imbalance ratio of the portfolio in the pre-shock period was >1 or <1. Panel A presents results for portfolios constructed from S&P500 index constituents and from non S&P500 index constituents using STCC GARCH model shocks to total order flow. Panel B presents results for shocks to retail, stealth and large institutional order flow.

Panel A	S&P500 Index Portfolios		Non S&P500 Index Portfolios	
	+Total	-Total	+Total	-Total
Pre shock cor	0.413	0.442	0.191	0.198
Post shock cor	0.326	0.335	0.154	0.138
Δ cor	-0.086	-0.107+	-0.037	-0.060+
t-test	-9.69	-23.44	-77.97	-14.82
% Δ in cor	-20.07%	-24.07%	-19.37%	-30.30%
% Δ in cor <0	63.84%	65.82%	59.58%	68.59%

Panel B

S&P500 Index Portfolios	+Retail	-Retail	+Stealth	-Stealth	+Large	-Large
Pre shock cor	0.485	0.466	0.427	0.402	0.5125	0.5135
Post shock cor	0.376	0.351	0.383	0.370	0.4669	0.4363
Δ cor	-0.109	-0.115+	-0.044	-0.032+	-0.0456	-0.0772+
t-test	-37.80	-30.31	-17.03	-11.85	-40.34	-57.03
% Δ in cor	-22.47%	-24.68%	-10.30%	-7.96%	-8.90%	-15.03%
% Δ in cor <0	55.5%	69.8%	62.4%	59.6%	54.3%	55.7%

Non S&P500 Index Portfolios	+Retail	-Retail	+Stealth	-Stealth	+Large	-Large
Pre shock cor	0.277	0.260	0.267	0.264	0.384	0.414
Post shock cor	0.292	0.227	0.240	0.238	0.370	0.416
Δ cor	0.015	-0.033+	-0.027	-0.026+	-0.014	0.002+
t-test	4.63	-24.12	-33.29	-35.99	-8.96	1.40
% Δ in cor	5.56%	-12.69%	-10.11%	-9.85%	-3.65%	0.52%
% Δ in cor <0	48.7%	62.5%	59.5%	58.3%	49.6%	42.4%

Table 6: Firm Characteristics and Sentiment Shocks.

This table reports the coefficient values of the following regression in which the changes in order flow imbalance D-OFIL are regressed against a range of firm characteristics as shown below.

$$\text{D-OFIL}_{itm} = \beta_1 + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \beta_5 Size_{it} + \beta_6 ILLIQ_{it} + \beta_7 Price_{it} + \beta_8 B/M_{it} + \beta_9 EPS_{it} + \beta_{10} EG_{it} + \beta_{11} AGE_{it} + \beta_{12} CM_{it} + \beta_{13} OWN_{it} + \beta_{14} INT_{it} + \beta_{15} S\&P_{it} + \varepsilon_{it}$$

where D-OFIL*itm* is the change in the order flow imbalance (OFIL) of investor group m (total, retail, stealth or institutional) in month *t* for firm *i*. α is a constant. RMF is the market return in excess of the risk free rate, SMB is the small firm factor, HML is the book-to-market factor and UMD is the momentum factor. Size, ILLIQ, PRICE, B/M, EPS, EG, AGE, CM, OWN, INT are all firm characteristics which may be associated with D-OFIL. These characteristics are Size(market value), ILLIQ (the Amihud illiquidity ratio), B/M(book-to-market ratio), Price(closing price), AGE(years listed on the exchange), CM(correlation between the daily return of stock *i* and the market), OWN(proportion of institutional ownership), INT(proportion of intangible assets to all assets), EPS (earnings to price ratio) and EG(earnings growth). We also include a dummy variable which has a value of unity if a firm is a S&P500 index constituent but has a value of zero otherwise (S&P).

Variable	All		Retail		Stealth		Large	
	+	-	+	-	+	-	+	-
Con	19.3747***	-19.4674***	7.4425***	-8.1655***	10.0398***	-9.8108***	8.1107***	-4.5417***
SMB	0.0025	-0.0045***	0.0000	0.0015	0.0023	-0.0027	-0.0014	-0.0013
HML	0.0019	0.0025***	0.0007	0.0007	-0.0012	-0.0031	-0.0005	-0.0002
UMD	-0.0020	0.0011	-0.0035***	0.0016	0.0024	0.0000	0.0001	0.0018
Size	0.0275	-0.0726	0.0602	0.0907	0.0120	-0.1149	-0.0656	-0.0874
ILLIQ	-0.0031	0.0454***	-0.0024	-0.0242***	-0.0202	0.0147	0.7785***	-3.8720***
Price	-0.3250***	0.6076***	-0.1852**	-0.0130	0.0262	0.2501	0.0105	0.0436
B/M	-0.5618***	-0.0028	0.1641	0.1250	-0.2673	0.4208	-0.2045	0.3141
EPS	0.5324***	-0.4572***	0.1675***	-0.2413***	0.3277***	-0.3919***	0.0001***	0.0000***
EG	-0.5317***	-0.1089	-0.2749***	-0.0353	-1.2891	-2.0342	-0.3203***	-0.0649
AGE	-2.1112***	2.4852***	-0.8319***	0.5830***	-1.4683***	1.6673***	-1.0918***	0.4669**
CM	0.0512	-0.7967	-0.6154	0.8505	0.0599	-0.3302	0.1745	-1.3808
OWN	-0.7016***	0.7972***	-0.4034***	0.3127	-0.2647	0.3167	-0.2728	0.1508
INT	-1.1041***	0.8714**	-0.3012	0.4193	-0.1037	0.0713	-0.7122***	0.5074
SP dummy	-0.0279	0.0958	0.6058	-0.2103	-0.7352	0.5320	-0.2110	-0.4314
Adj R^2	0.4130	0.4503	0.2676	0.3458	0.1212	0.3508	0.2888	0.3744

Table 7: Sub-period Analysis of Smooth Transition Results

This table shows the average pairwise pre and post shock portfolio conditional correlation coefficient obtained from the STCC GARCH model for three separate sub-periods. These are reported separately for positive (+ shock) and negative shocks (- shock). Total, Retail, Stealth and Large presents model results separately for shocks to total, retail, stealth and large institutional order flow endogenously determined by the model. Pre shock cor is the average pairwise conditional correlation prior to a shock, post shock cor is the average pairwise conditional correlation after a shock. Δ cor is the average change in the post shock conditional correlation between the post shock and pre shock period. A + indicates that the average difference between the conditional correlation change following a positive shock is statistically different to the change in conditional correlation following a negative shock at a 5% level using a t-test. The t-test row presents the t-value associated with the test of whether the change in conditional correlation following a shock is significant at a 5% level. % Δ cor is the percentage change in average conditional correlation that takes place after an order flow shock. The % Δ in cor<0 captures the proportion of negative changes in conditional correlation as a percentage of the total number of post shock changes.

	Total				Retail			
	1993-2000		2001-2011		1993-2000		2001-2011	
	+	-	+	-	+	-	+	-
Pre shock cor	0.2318	0.2206	0.2418	0.2199	0.349	0.3076	0.3290	0.2665
Post shock cor	0.1974	0.184	0.2269	0.1928	0.3104	0.2514	0.2983	0.2355
Δ cor	-0.0344	-0.0366+	-0.0149	-0.0271+	-0.0386	-0.0562+	-0.0307	-0.0310
t-test	-32.48	-47.92	-11.92	-32.09	-29.28	-44.13	-14.83	-16.50
% Δ in cor	-14.84%	-16.59%	-6.16%	-12.32%	-11.06%	-18.27%	-9.33%	-11.63%

	Stealth				Large			
	1993-2000		2001-2011		1993-2000		2001-2011	
	+	-	+	-	+	-	+	-
Pre shock cor	0.3104	0.3	0.3071	0.307	0.3819	0.3407	0.4092	0.332
Post shock cor	0.2749	0.2556	0.29	0.2887	0.3551	0.2995	0.3804	0.3289
Δ cor	-0.0355	-0.0444+	-0.0171	-0.0183+	-0.0268	-0.0412+	-0.0288	-0.0031+
t-test	-21.35	-32.40	-12.10	-13.23	-11.97	-10.81	-16.40	-0.76
% Δ in cor	-11.44%	-14.80%	-5.57%	-5.96%	-7.02%	-12.09%	-7.04%	-0.93%

Table 8: Analysis Using Different Return Frequencies

This table shows the average pairwise pre and post shock portfolio conditional correlation coefficient obtained from the STCC GARCH model using three different return intervals. Two-day presents results using two-day return intervals, five day uses five day return intervals. These are reported separately for positive (+ shock) and negative shocks (- shock). Total, Retail, Stealth and Large presents model results separately for shocks to total, retail, stealth and large institutional order flow endogenously determined by the model. Pre shock cor is the average pairwise conditional correlation prior to a shock, post shock cor is the average pairwise conditional correlation after a shock. Δ cor is the average change in the post shock conditional correlation between the post shock and pre shock period. A + indicates that the average difference between the conditional correlation change following a positive shock is statistically different to the change in conditional correlation following a negative shock at a 5% level using a t-test. The t-test row presents the t-value associated with the test of whether the change in conditional correlation following a shock is significant at a 5% level. % Δ cor is the percentage change in average conditional correlation that takes place after an order flow shock.

	Total		Retail		Stealth		Large	
Two-day	+shock	-shock	+shock	-shock	+shock	-shock	+shock	-shock
Pre shock cor	0.2114	0.2426	0.3363	0.3088	0.3058	0.2870	0.3523	0.3523
Post shock cor	0.1874	0.2192	0.3081	0.2633	0.2935	0.2678	0.357	0.2991
Δ cor	-0.0240	-0.0234+	-0.0283	-0.0456+	-0.0124	-0.0193+	-0.0708	-0.0531+
t-test	30.55	30.35	23.14	33.57	9.42	18.38	41.75	19.58
% Δ in cor	-11.36%	-9.66%	-8.41%	-14.75%	-4.04%	-6.71%	-16.56%	-15.08%

Five-day	+shock	-shock	+shock	-shock	+shock	-shock	+shock	-shock
Pre shock cor	0.2086	0.2222	0.3119	0.2855	0.3223	0.2942	0.3882	0.3487
Post shock cor	0.2119	0.1954	0.2411	0.2530	0.3142	0.2943	0.3518	0.3256
Δ cor	0.0034	-0.0269+	-0.0708	-0.0325+	-0.0081	0.0001+	-0.0364	-0.0231+
t-test	-3.72	33.78	48.13	30.80	5.70	-0.05	16.87	9.67
% Δ in cor	1.62%	-12.08%	-22.69%	-11.40%	-2.51%	0.02%	-9.38%	-6.63%

Fig 1A: Order Flow Imbalance Over Time - Total

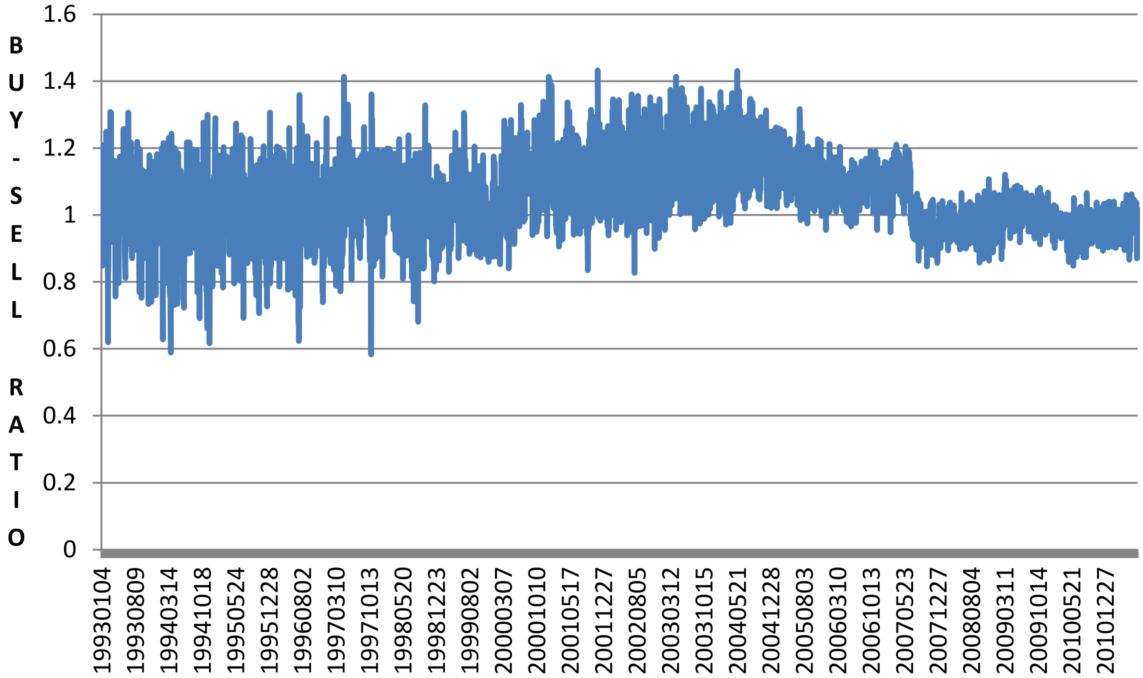
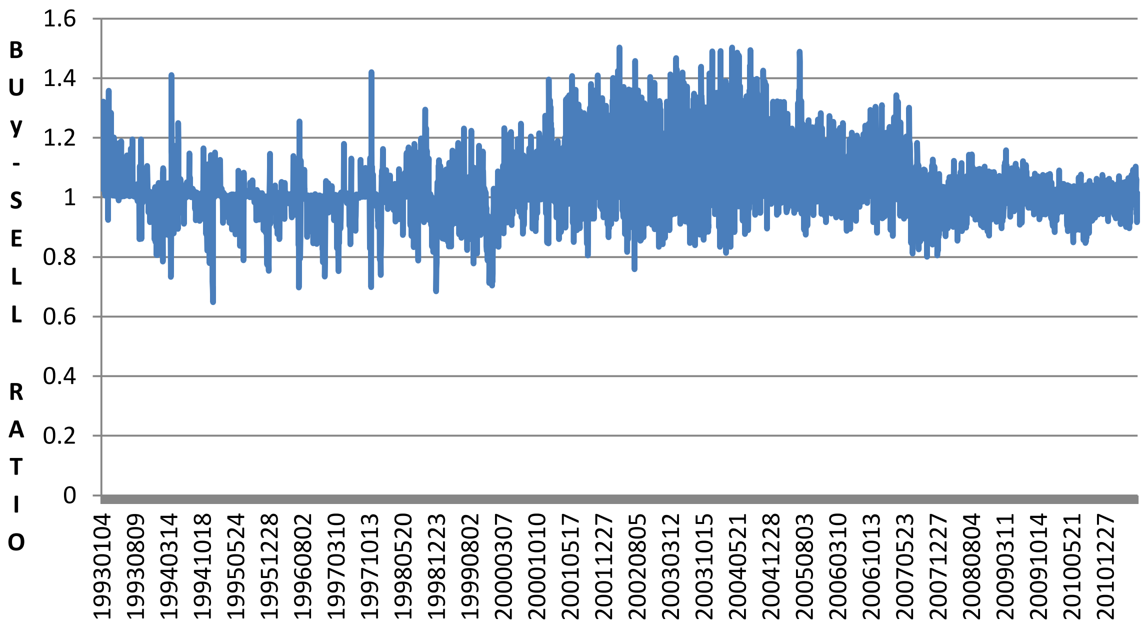


Fig 1B: Order Flow Imbalance Over Time - Retail



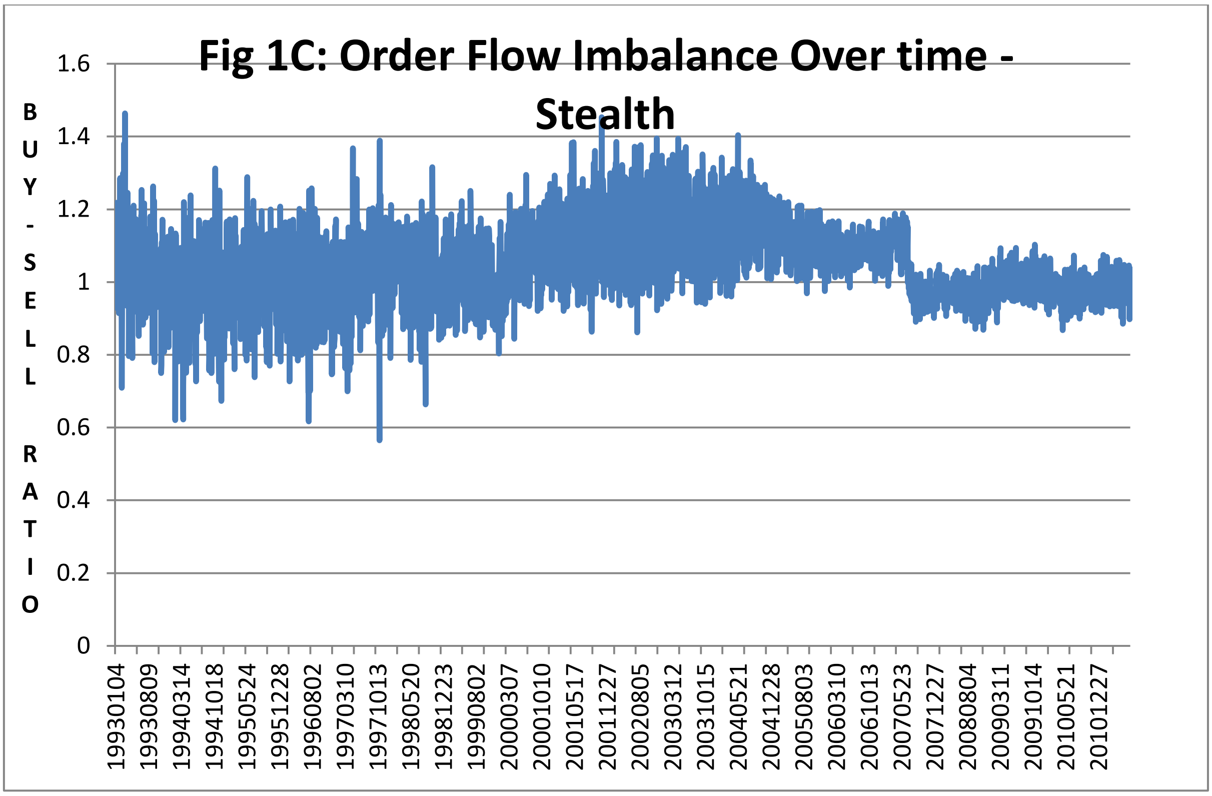


Fig 1C: Order Flow Imbalance Over time - Stealth
BUY-SELL RATIO
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
19930104
19930809
19940314
19941018
19950524
19951228
19960802
19970310
19971013
19980520
19981223
19990802
20000307
20001010
20010517
20011227
20020805
20030312
20031015
20040521
20041228
20050803
20060310
20061013
20070523
20071227
20080804
20090311
20091014
20100521
20101227

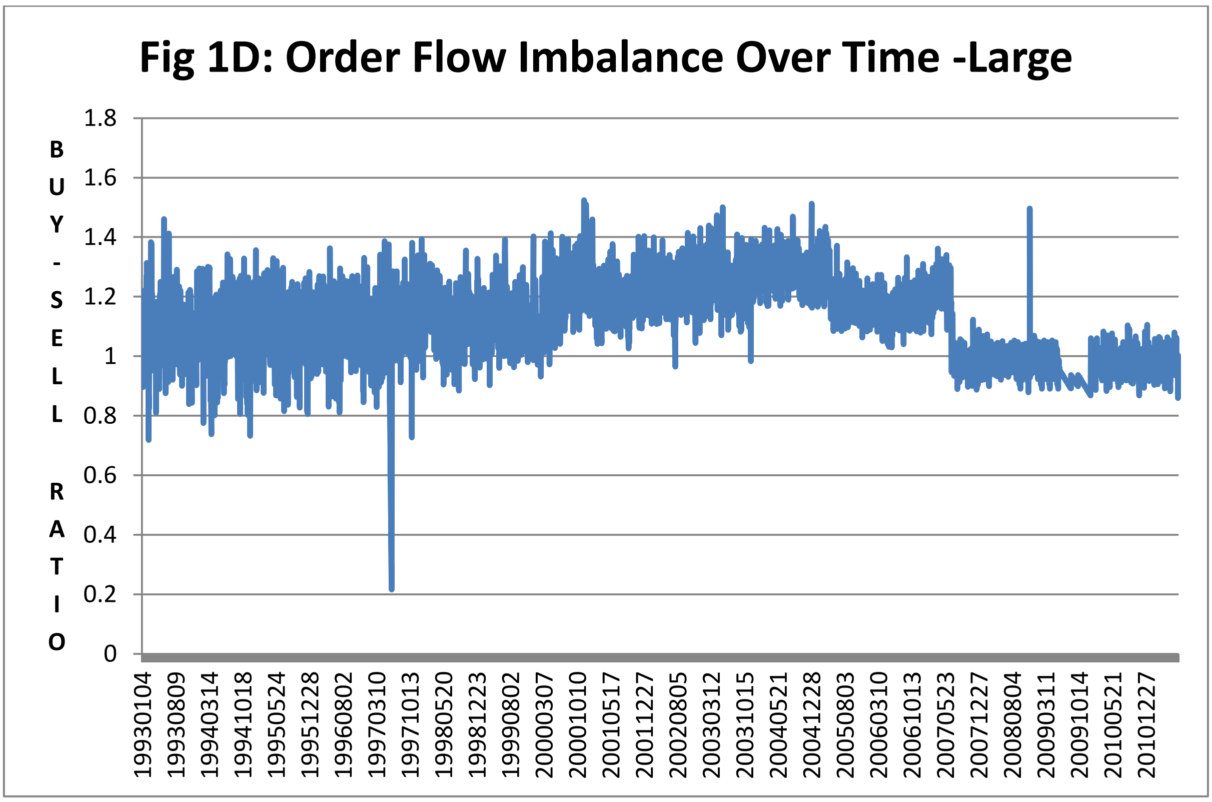


Fig 1D: Order Flow Imbalance Over Time -Large
BUY-SELL RATIO
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
19930104
19930809
19940314
19941018
19950524
19951228
19960802
19970310
19971013
19980520
19981223
19990802
20000307
20001010
20010517
20011227
20020805
20030312
20031015
20040521
20041228
20050803
20060310
20061013
20070523
20071227
20080804
20090311
20091014
20100521
20101227