The Shareholder Base Hypothesis of Stock Return Volatility: Empirical Evidence

Håkan Jankensgård, Anders Vilhelmsson*

Abstract

We use Swedish ownership data to explore whether a large and diversified shareholder base leads to lower volatility by improving the information content of stock prices. We find that volatility increases in the number of shareholders both with respect to the number of relatively large shareholders and the fraction of shares held by investors with stakes below 0.1%. Volatility is also positively related to the number of institutional owners, though negatively related to the number of under-diversified institutional owners. Foreign investors have no impact. Our results suggest that a large shareholder base does not lower volatility.

Key words: Volatility; ownership; shareholder base; portfolio concentration

JEL code: G1, G30, G32

** Håkan Jankensgård is Associate Professor in the Department of Business Administration, Lund University, Sweden. Anders Vilhelmsson is Associate Professor in the Department of Economics, Lund University, Sweden.*

The authors wish to thank an anonymous referee, Marc Lipson (Editor), Adri de Ridder, Milda Norkute, Joakim Westerlund, Jens Forssbaeck, Martin Holmén, Tom Aabo, Martin Strieborny, Ettore Croci, Björn Hagströmer and Frederik Lundtofte for useful suggestions. We also thank seminar participants at the Knut Wicksell Centre for Financial Studies, Lund University; School of Business Economics and Law, Gothenburg University; and School of Business and Social Sciences, Aarhus University. We are grateful to Sven-Ivan Sundqvist at SIS Ägarservice for supplying the ownership data. Jankensgård gratefully acknowledges the financial support of the Jan Wallander and Tom Hedelius foundation and the Tore Browaldh foundation. Vilhelmsson gratefully acknowledges the financial support of the Marianne and Marcus Wallenberg Foundation.

There is great variation in the structure of corporate ownership. Whereas some firms have a small number of domestic owners, others have a more diversified structure with institutional investors, private equity, business spheres, and foreign owners as a significant part of the ownership. Some firms have a large fraction of small investors, whereas others are dominated by a smaller number of relatively large ownership blocks. The questions we address in this article are whether stock return volatility is determined by what kind of investors the firm has and how many they are, which is hereafter referred to as the shareholder base. Previous literature has noted the central importance of stock return volatility to financial theory, as well as for practitioners in the investor community (e.g. Campbell, Lettau, Malkiel, and Zu, 2001; Zhang, 2010). We contribute to this literature by examining shareholder base determinants of volatility.

Our inquiry is motivated by an intriguing conjecture in the literature which holds that volatility decreases in the size of the firm's shareholder base. According to Wang (2007), the shareholder base-broadening effect occurs because each individual has only partial information about the firm. As the number of investors grows the accuracy of the information available about the stock increases, which in turn lowers the variance of stock returns. This follows from an extension of the Merton (1987) analysis of investor recognition, according to which a small shareholder base leads to higher expected returns because risks are insufficiently shared among investors. While the importance of the shareholder base is hinted at in several papers it has not yet, to our knowledge, been comprehensively investigated empirically (Merton, 1987; West, 1988; Wang, 2007; Rubin and Smith, 2009; Li, Nguyen, Pham, and Wei, 2011). This study fills this gap.

Stock return volatility may also decrease if the firm diversifies its shareholder base by attracting specific groups of shareholders who are better informed compared to individual domestic investors. The literature has identified two categories of shareholders that are likely to improve the information content of a firm's stock price: institutional and foreign investors. According to Rubin and Smith (2009), institutional investors are more financially sophisticated and therefore contribute to more efficient gathering and processing of information. Using a sample of US firms they find that institutional ownership leads to lower volatility for firms that do not pay a dividend because the information environment tends to be weaker for this category of firms. Li et al (2014) show that trading by foreign investors has a stabilizing influence on volatility in the context of emerging markets following financial liberalisation.

In this study we investigate whether the effects from increasing and diversifying the shareholder base persist in an economy that steers a middle course between the Anglo-American governance model and emerging markets where concentrated ownership is predominant. Sweden has long been a developed, open economy with liquid financial markets that mixes 1) a traditional governance model involving strong controlling owners and various mechanisms for corporate control (e.g. differential voting rights) with 2) a large influx of institutional and foreign owners, whose influence has grown substantially in recent decades. This variation makes for a useful setting for testing the relationship between the size and heterogeneity of the shareholder base, on the one hand, and stock return volatility, on the other.

We furthermore benefit from access to databases on corporate ownership that offer several advantages for identifying the shareholder base. SIS Ägarservice collects detailed ownership data on publicly listed firms in Sweden, and provides the actual ownership lists of these firms. This allows us to characterize firms' shareholder base in terms of the number of investors, their type and nationality, cash flow and voting rights, association with a business sphere, and so on for a broad cross-section of firms, from the very smallest listed firms to the large-cap multinationals. Furthermore, access to data from the VIRSO files (Visby Research in Stock Ownership) allows us to identify the size of the shareholder base on various dimensions, such as the number of individual, institutional and foreign investors, as well as other categories of owners such as trusts, operating companies, and churches. As will be explained in detail later in this article, several features of these databases are unique, and provide a better identification of the shareholder base than looking at aggregate ownership stakes. We construct a sample of ownership data spanning the years 2000 to 2013, yielding around 1900 firm-year observations. We use a rich set of control variables from theories related to earnings uncertainty, information asymmetry, leverage, and portfolio concentration.

The data does not bear out the hypothesis that a large shareholder base is conducive to lower volatility. In fact, the evidence is overwhelmingly to the contrary. We first show that volatility increases in the total number of shareholders. This effect persists when we break down the shareholder base into different groups based on the relative size of their ownership stakes, which captures differing incentives to produce information about the stock. Volatility increases in the number of relatively large shareholders (stakes>0.1%) as well as in the size of the 'micro-float' (that is, the fraction of shares held by small investors with stakes below 0.1%). These results are generally significant at the 1%-level and robust to various econometric approaches. We implement an instrumental-variable approach (2SLS) to deal with the potential problem that small investors self-select into high volatility stocks, which could also explain the association. The relationship remains statistically significant after we account for selection.

The evidence also fails to support the hypothesis that a more diversified ownership structure leads to lower volatility. Taken together, the evidence instead suggests that having a large number of institutional investors causes volatility to increase, whereas the number of foreign investors has no effect. There is a robust and highly significant relation between the number of institutional owners and volatility, consistent with previous research showing a positive relationship using US data (e.g. Campbell et al, 2001; Xu and Malkiel, 2003; Dennis and Strickland, 2009). The effect of institutional trading on volatility remains an unresolved issue, however, and our study complements previous studies because we use the number of institutions rather than the aggregate ownership stake. We argue that a count measure better captures the preconditions for the trading and herding-behavior that has been ascribed to institutional investors (Sias, 2004). A high degree of institutional ownership, on the other hand, could be explained by the presence of a limited number of large block holdings. If such block holdings are also more long-term in nature, as seems likely, such a measure will be a relatively poor proxy for institutional trading.

We interpret our findings to suggest a trading channel-explanation. Zhang (2010) argues that sources of stock return volatility can be divided into either fundamental factors related to uncertainty about future dividends, or factors related to trading patterns. We find, using a simultaneous equations-approach (3SLS), that various proxies for the size of the shareholder base are indeed positively related to trading volume (significant at the 1%-level), and that the effect on volatility from the shareholder base is accounted for by this channel. A plausible interpretation is that the size of the shareholder base, in particular the micro-float, captures the presence of noise trading.[1] Noise traders are affected by beliefs and sentiments that are not fully justified by fundamental news (Shleifer and Summers, 1990). Such traders do not, unlike so-called arbitrageurs, form rational expectations about future returns given the available information. While the presence of noise traders increases the value of information to rational investors, creating incentives to produce more information, prices do not necessarily become more efficient. An important reason for this is related to practical limits to arbitrage (Black, 1986; Shleifer and Summers, 1990). Our data instead supports the view that the traits typically attributed to noise traders (such as over-reaction to news and trend-following) instead serve to amplify stock return volatility. This interpretation is consistent with studies investigating the trading patterns of individual retail traders (e.g. Foucault, Sraer, and Thesmar, 2011).

We also contribute to a recent strand in the literature that seeks to explore heterogeneity among investors for explaining the relationship between institutional ownership and volatility (Rubin and Smith, 2009; Chichernea, Petkevich, and Zykaj, 2015; Ben-David, Franzoni, and Moussawi, 2014). Chichernea et al (2015) show that it is predominantly investors with a short investment horizon (high churn-rate) who are responsible for the positive association between institutional ownership and volatility. We find that it matters whether institutional investors have large or small ownership stakes. Whereas a high overall number increases volatility, the relation between the number of *large*

---

[1] Black (1986) defines noise as an arbitrary component in expectations that leads to an arbitrary price change consistent with expectations. He also predicted a relation between noise trading and volatility: "Anything that changes the amount or character of noise trading will change the volatility of price".

institutional ownership stakes and volatility tends to be *negative.* It follows that the positive relation reported initially is driven by the presence of a high number of relatively small institutional investors. We believe this is consistent with a story about investment horizons: large institutional holdings are more likely to be long-term investments and have a dampening effect on volatility, whereas having many small institutional investors is conducive to more trading. We also show that the reduction in volatility is even stronger when the largest institutional investors are over-exposed to the firm in question, using measures of portfolio concentration similar to those used in Ekholm and Maury (2014). We address concerns about endogeneity by showing that the effect of having a large number of institutional investors on volatility is in fact lower for high values of the firm's microfloat, which captures the presence of a multitude of small and uninformed traders. This speaks against the notion that institutional investors, in order to profit from their superior information, select into high-volatility stocks with a larger number of noise traders.

We conclude that a large and diverse shareholder base leads to a more volatile stock price. This may seem puzzling in light of well-documented attempts by corporate managers to boost the firm's shareholder base and the liquidity of its stock. The strand of research dealing with the "neglected stocks-hypothesis" has generally found a positive value-effect from boosting the visibility of the stock, which often includes measures aimed at increasing the size of the shareholder base (see Chichernea, Ferguson, and Kassa, 2015, for a summary of this literature). According to Merton's (1987) equilibrium model, neglected stocks (that is, those with a small shareholder base) are priced below their full information equilibrium price when investors follow only a subset of firms and are insufficiently diversified. Our findings should be viewed against the backdrop of Merton's intuition and the subsequent literature on neglected stocks. An increase in volatility resulting from a larger fraction of noise trading may be a price worth paying to avoid the detrimental value effects from being stuck in a neglected corner of the market.

The paper proceeds as follows. Section I reviews prior literature and presents the hypotheses to be tested in this article. Section II introduces the sample, variables, and empirical methodology. Section III contains the empirical analysis of the determinants of stock return volatility. Section IV concludes the paper.

I. Prior literature and hypotheses

A. The shareholder base and stock return volatility

According to Merton (1987) there is a relation between investor recognition (the shareholder base) and expected returns. A small shareholder base implies that only a fraction of the stock market is informed about the stock. This increases expected returns since risky assets are valued below their full information equilibrium price when there is insufficient sharing of risks. Extending on Merton's analysis, Wang (2007) demonstrates that increasing the size of the shareholder base reduces volatility. Wang argues that volatility decreases because the accuracy of the price signal improves, i.e. the degree to which the stock price reflects underlying fundamentals. On an intuitive level, when more information is reflected in the stock price, fewer events will count as news requiring a reassessment of the price. When the collection and processing of information is efficient events are, to a larger extent, anticipated by the market.

A higher information content of the stock price can come about in different ways. According to Holmström and Tirole (1993) increases in the shareholder base (liquidity trading in their model) increases the marginal value of information production for speculators. The resulting flow of information into the market increases the amount of information that is reflected in the stock price. Rubin and Smith (2009) make a similar argument for institutional investors. According to their institutional sophistication-hypothesis, institutional investors, by way of their higher financial sophistication, improve the information content of market prices which in turn lowers the volatility of stock returns.

The above outlined arguments suggest that a larger and more diversified shareholder base is associated with lower volatility. We will refer to this as **the shareholder base-hypothesis** of stock return volatility.

B. Other theories of stock return volatility

This section takes its point of departure in the two broad theories of stock return volatility referred to by Zhang (2010) as “fundamentals” and “the trading channel”. The trading channel refers to the empirically observed connection between trading volume and volatility. A higher trading volume implies a higher level of volatility, reflecting that prices tend to move when trades occur. Whereas ownership per se does not affect volatility, the trading behavior of different categories of owners is likely to do so. For example, institutional investors as a group have been associated with herding behavior (Sias, 2004).

In the terminology of Zhang (2010) “fundamentals” refer to variables that impact uncertainty about future free cash flows. The literature has investigated several proxies that capture the extent of uncertainty about future free cash flows, henceforth referred to as **earnings uncertainty**.[2] These include the cross-sectional variation in earnings (Pástor and Veronesi, 2003; Wei

[2] Earnings can be viewed as a rough proxy for free cash flow to the extent that the depreciation charge included in earnings approximates the investment outlay that would be subtracted from operating cash flow to arrive at free cash flow.

and Zhang, 2006); growth options (Cao, Simin, and Zhao, 2008); technology intensity (Schwert, 2002); and firm focus (Dennis and Strickland, 2009). Taken together, the findings in these studies support the notion that fundamental uncertainty is positively related to stock return volatility.

Uncertainty about fundamentals may be exacerbated by lack of information about the firm. When **information asymmetries** are high investors have more difficulty predicting future states, leading to larger and more frequent revisions and hence higher volatility. The firm's information environment is likely to be impacted by factors such as the firm's age (Pástor and Veronesi, 2003); the disclosure requirements of the exchange on which it is listed (ibid); size (Singhvi and Desai, 1971); and dividends (Baskin, 1989).[3]

Volatility is furthermore determined by **leverage**. The volatility of equity returns will be amplified by leverage because the firm's equity is a levered position in the firm's assets (Black, 1976). Previous papers have documented an association between financial leverage and stock return volatility (e.g. Dennis and Strickland, 2009; Rubin and Smith, 2009).

Ownership matters also because different owners have different preferences for corporate governance (Li et al, 2011) and risk-taking (Faccio, Marchica, and Mura, 2011). It is reasonable to expect that a preference for low-risk policies is systematically related to the degree of diversification on part of the largest owners, or what Ekholm and Maury (2014) have labelled 'portfolio concentration'. According to this argument, then, the more of an investor's wealth that is concentrated to a given firm, the larger are her incentives to monitor managers and limit risk-taking in this firm. We will refer to this as the **portfolio concentration-**hypothesis of stock return volatility**.**

[3] The informational effect of dividends obtains because investors who rely on them for input in a valuation are less susceptible to irrational pricing. That is, dividends provide an "anchor" or reference point that partially replaces the need to collect information on other value indicators in the process of pricing a firm's stock.

II. Sample, model specification, and variables

A. Sample and data

The sample used in this study is the intersection of three databases: SIS Ägarservice (ownership data), the VIRSO files (shareholder base data) and Datastream (financial data). The sample covers the period between 2000 and 2013. Merging these databases yields around 1900 firm-year observations. The industry composition is as follows: basic materials 6%; industrials 32%; consumer goods 11%; health care 13%; consumer services 13%; telecommunications 2%; technology 13%; and oil & gas 3%; financials 7%. There are no utility firms in the sample. Our data covers the main exchange in Sweden (Nasdaq OMX, which in turn consists of the Large-cap, Mid-cap, and Small-cap lists). In addition, it covers firm listed on the Nordic Growth Market (NGM). Firms listed on the two exchanges for risk capital, First North and Aktietorget, are not included. Our sample represents more than 99% of the total market capitalization of Swedish listed firms.

We obtain the data on the shareholder base from the VIRSO files.[4] These files draw on raw data from Euroclear Sweden (formerly VPC). The advantages of using data based on Euroclear Sweden are stated in Bodnaruk and Österberg (2009). Euroclear Sweden bases its calculation of the number of shareholders on an electronic register in which all outstanding shares, and their owner, are automatically recorded. The Compustat estimate of the number of shareholders of record, however, is based on information in 10-K filings, where firms are required to approximate the

[4] VIRSO is short for Visby Research on Stock Ownership. It is a project aimed at compiling and disseminating data on ownership of Swedish securities to enhance academic research. It is carried out in collaboration between Uppsala University and Visby Högskola.

number of holders of each class of common equity. Furthermore, the Compustat data does not identify individuals holding stock through a brokerage house, employee stock option plan, or trust. As a result, multiple individuals will appear as a single shareholder of record when they own their stock through the same intermediary. The data from Euroclear Sweden does not have this bias.

The VIRSO files furthermore match the raw information on security ownership in Euroclear Sweden with various official registers to identify the nature of the owner. Each share is connected to an individual or legal entity through a Social Security number or organization number (which is required to register a company with the Swedish Companies Registration Office, "Bolagsverket"). The organizational number will differ depending on whether the unit is an operating company; non-profit association; church; foundation; and so on. Alternatively, the basis for identification of legal entities is the classification used by Statistics Sweden (SCB, the official administrative agency of statistics in Sweden). In addition, the VIRSO files identifies mutual funds through a particular register of firms that are under the supervision of Finansinspektionen, the authority entrusted with the task of monitoring financial firms and the stability of the financial system in Sweden.

We also utilize ownership data collected by SIS Ägarservice, a Stockholm-based firm specialized in providing and analyzing ownership information. They provide ownership lists for public firms in Sweden that show the number of A, B and preference shares held by a particular individual or legal entity; the associated cash flow and voting rights; the owner's nationality; a flag indicating whether the owner is classified as an insider[5]; and the 'weight quota'. The weight quota is defined by SIS Ägarservice as the focal firm's share of the owner's Swedish portfolio less the firm's share of the

[5] This classification is based on Finansinspektionen, the central authority for overseeing financial markets in Sweden. A person is considered an insider if he or she can be regarded to have a relationship with the company such that it is likely to provide inside information about its affairs. Typically, but not limited to, this would be an executive officer, director of the board, or an owner with more than 10% of the votes.

total Swedish stock index. A positive weight is interpreted as the owner being over-exposed in this firm relative to the exposure one would obtain from investing in a broad index. A negative weight is correspondingly interpreted as being under-exposed to the firm in comparison with the index. A value of zero means that the owner has an exposure to the focal firm that is equal to that obtained from investing in the index. Typically an ownership list identifies the 50-200 largest shareholders.[6] Ownership stakes smaller than 0.1% are not published for integrity reasons. They are rather summed up in a residual figure, which we refer to as 'the micro-float'.

Another powerful feature of SIS Ägarservice is that they allow the user to view the ownership list organized according to spheres of influence. Activating this option means that individual owners who can be assumed to act in a synchronized way, based on e.g. kinship, are listed as a group. This reveals where the control of the firm ultimately lies ('ultimate ownership'). As an example consider the ownership of Clas Ohlson, a large-cap retailer, at year-end 2013. The raw ownership list would suggest that control lies with the single largest owner (at 17.3% of votes). However, the sphere-view reveals that the members of the Haid-family (none of whom was the single largest owner) together controlled 37% of the votes and had ultimate control.

SIS Ägarservice also traces out and provides detailed information on the holdings of the 20 most influential spheres in Sweden. These spheres are typically built around the holdings of successful industrialists. One example is the Kinnevik-sphere, which originated in the various business ventures of the entrepreneur Jan Stenbeck. A salient feature about several of these spheres is that they use a complex web of controlling stakes and cross-holdings to achieve effective control, though cash flow rights are typically much smaller than control rights (see Collin, 1998).

---

[6] The number is capped at 200, so even if there are more separate owners with stakes exceeding 0.1% only 200 would show in the list. The maximum of 200 occurs in 8% of cases.

B. Empirical approach

Our main empirical model is a straightforward OLS estimation of the determinants of volatility with unbalanced panel data. As a complement we also estimate a dynamic panel data model that includes lagged volatility as an explanatory variable.[7] Also the dynamic model is estimated with OLS since GMM methods such as Arellano and Bond (1991) do not work when the explanatory variables show high time-series persistence. The general model is outlined in Equation 1:

$$\begin{aligned} Volatility_{j,t} = {} & \alpha_k + f_l + d_t + \beta_1 Earnings\ uncertainty_{j,t-1} \\ & + \beta_2 Information\ asymmetry_{j,t-1} + \beta_3 Leverage_{j,t-1} \\ & + \beta_4 Portfolio\ concentration_{j,t-1} + \beta_5 Shareholder\ base_{j,t-1} \\ & + v_{j,t} \end{aligned} \qquad (1)$$

where α_k is industry fixed effects, f_l is listing fixed effects, d_t is period fixed effects, and $v_{j,t}$ is an error term. Earnings uncertainty$_{j,t-1}$, information asymmetry$_{j,t-1}$, leverage$_{j,t-1}$, portfolio concentration$_{j,t-1}$, and shareholder base$_{,t-1}$ are arrays of variables related to the five theoretical dimensions reviewed in

[7] It is important to note that excluding an autoregressive term from a regression (a regular time series regression or as in our case a static panel data regression) is not an omitted variables problem. For the non-panel case most standard textbooks in statistics point out that autocorrelation in the residuals still leaves OLS estimates unbiased and consistent (but inferences will be incorrect). The reader is referred to Greene for details (2002, p.265). This also holds true for static panel data models (Baltagi, 2008, p.92). Excluding the AR(1) term leaves residuals autocorrelated within firms (over time), but correct inference is still obtained by clustering standard errors at the firm level (which we do in all specifications). Another potential disadvantage with a dynamic specification is that when the dynamic structure is incorrectly specified, coefficient estimates will be biased (Lee, 2012). When T is small, as in our case, correctly specifying the dynamics is difficult. However, we report the results including an auto-lag since doing so is common in the empirical literature on volatility.

section I.B. The subscript *t* indexes time, *k* industries, *l* stock exchange listings and *j* indexes firms. Based on recommendations in Petersen (2009) we use standard errors clustered at the firm level. Since the model always includes period fixed effects we need only to account for clustering of standard errors within firms (over time). In all estimations we lag the independent variables one period since volatility is estimated based on daily returns during the year, whereas the independents are measured at year-end.

It is important to note that the analysis in Merton (1987) is essentially an incomplete information equilibrium model. The theoretical concept of shareholder base described by Merton should capture the *fraction* of investors in the market who are informed about a certain security. This means that one has to account for time-variability in stock market participation (especially in the context of panel data analysis).[8] One could deal with this either by scaling the number of shareholders by the total number of investors in the economy (or in our dataset) and repeat the panel data analysis, or estimate Fama-McBeth regressions, where scaling will not present that big of an issue (since the total number of investors is the same at a given point in time). We choose the latter approach and, in addition to the panel data analysis, report results from Fama-McBeth regressions with Newey West corrections.

C. Variables

1. Volatility

We construct our measure of volatility by using the theory of realized volatility originally proposed in Andersen and Bollerslev (1998). They show that the true unobservable volatility can be very well approximated by using squared return data of a higher frequency than the variance being

[8] We thank an anonymous referee for suggesting this point.

constructed. In practice market microstructure noise means that the sample frequency cannot be arbitrarily high since this introduces bias (see for example Aït-Sahalia, Mykland, and Zhang, 2005, as well as Hansen and Lunde, 2006). In practice 5 min to 15 min returns are often used to construct daily volatility, and yearly volatility is often constructed from daily data. We compute volatility by first squaring daily stock returns and summing the squared terms over all trading days in the year. Our estimate of the annualized total volatility, *Volatility*, is defined as the square root of this sum.[9]

As an alternative measure we also calculate idiosyncratic volatility by applying the same procedure to the standard errors from an estimation of the market model in which daily stock returns are regressed on the Swedish market index. Similar to other authors in the literature we find that the correlation is very high (0.98) and it does not matter to our results which of the volatility measures we use. We prefer to use total volatility because it also captures the extent to which our independent variables have a systematic component. We do not pursue other approaches to calculate idiosyncratic volatility, such as deriving it from an estimate of the Fama and French 3-factor model (FF3, Fama and French, 1993). The factors in FF3 are instead included as independent variables in our regressions.

2. Shareholder base

To measure the overall size of the shareholder base we use the number of common shareholders of record, *Nrshareholders,* as obtained from the VIRSO files. This variable corresponds to the

---

[9] To make sure our daily returns are not contaminated by market microstructure noise we calculate so-called volatility signature plots for daily, bi-daily and weekly returns. Market microstructure noise will bias the estimate if the frequency is too high but, reassuringly, we find the same level of volatility for all three sample frequencies.

shareholder base measure used in Bodnaruk and Östberg (2009). We also consider two measures that break down the shareholder base into two categories of investors depending on how large the ownership stakes are. *Nrshlarge* is defined as the number of shareholders that appear on the SIS Ägarservice ownership list (covering those that have a stake larger than 0.1%). It thus captures the number of shareholders with relatively large ownership stakes. *Microfloat* is defined as the fraction of shares held by investors whose ownership stake is smaller than 0.1%. It thus captures the relative importance of investors with very minor stakes, considered as a group.[10] We believe the distinction between these two categories of investors is meaningful in light of the hypothesis presented in section I.B. Following the logic in Holmström and Tirole (1993), we argue that the presence of a large fraction of small and unsophisticated investors (the microfloat) increases the value of information for speculators, which increases the total information production. On the other hand, relatively large investors (in our case, those that have stakes larger than 0.1%) have stronger economic incentives to collect and process information about the stock. While the mechanism is different, both arguments imply that volatility should be a decreasing function of the number.

To measure the diversity of the shareholder base we look at institutional and foreign ownership.[11] We employ several measures of institutional ownership. We first calculate the number of institutional owners, *Nrinst*. This is obtained from the VIRSO files. We sum the number of mutual

[10] As previously noted the number 0.1% is given by privacy laws and is thus arbitrary. From our point of view a threshold in percentage terms introduces the concern that the microfloat in certain large cap firms has a very different meaning than in small cap firms. In the former group 0.1% could still signify relatively large stakes that are beyond the means of most private retail investors. We are able to document that the results presented in section III.B are robust across different size-tranches, suggesting that this is not unduly affecting our results.

[11] Our data allows us to quantify the impact of other groups of owners as well, such as trusts, churches, and other legal entities (essentially operating companies). We do not report these to preserve space. Results are available from the authors.

funds (including hedge funds), insurance companies, and trust funds.[12] We obtain the variable *Nrinstdef* by deflating it with the total number of shareholders. To create a measure similar to the one used by previous studies on stock return volatility we use *Instown*, defined as the sum of ownership stakes held by institutional investors. It is limited to the 15 largest owners of the firm. Typically the 15th largest ownership stake is below 1% and summing stakes below this size has a negligible impact on the aggregate number.[13]

Nrforeign is defined as the number of investors not domiciled in Sweden. A caveat is that, with few exceptions, this variable comprises only legal entities, since the names and personal numbers of non-Swedish individuals are not available in the raw data from Euroclear Sweden. Another caveat is that there is no distinction between foreign institutional owners and other kinds of entities such as operating companies. *Nrforeigndef* deflates this variable with the total number of shareholders. *Foreignown* is the sum of ownership stakes held by institutional investors. Similar to *Instown* it is limited to the stakes held by 15 largest owners of the firm.

3. Other variables

[12] Pension funds are included mainly in the insurance companies' category, but some are also recorded among trust funds.

[13] This limitation is imposed also because these owners had to be classified manually by the authors since they are obtained from SIS Ägardata, which does not indicate the owner's status as an institution.

Our proxies for portfolio concentration focus on the firm's largest owner, as they have both the means and incentives to influence corporate policy. *Sphere* takes the value one if the largest owner is on SIS Ägarservice's list of the 20 most influential spheres in Sweden. Spheres of influence tend to have more concentrated holdings compared to e.g. institutional investors. As a more explicit measure of the degree of over- or under-exposure we use the portfolio concentration measure computed by SIS Ägarservice. This variable is labeled *Weight.* It is similar to but not identical to the portfolio concentration measure used in Ekholm and Maury (2014). These authors use the firm's share in the owner's portfolio of equities without benchmarking against a broad index, which is the case with our data. A significant difference is that we use the weight quota of the largest owner, whereas Ekholm and Maury use the average portfolio concentration of all the firm's shareholders.[14]

We introduce five variables that capture earnings uncertainty. *Diversification* is the number of product segments for which the firm reports revenue (Datastream code: WC19500). This definition differs from the firm focus-variable in Dennis and Strickland (2009) who construct a Herfindahl-index based on similar segment information. *Geographical* is the number of geographical segments for which the firm reports revenue (WC19600). This variable measures the degree of diversification in terms of the firm's presence in various geographical markets. Being diversified geographically will tend to make cash flows more stable if international product markets are less than perfectly correlated. *Intangibles* is intangible assets divided by total assets (WC02649/WC02999). Intangible assets are inherently more associated with uncertainty.

[14] Our measure is not without caveats. It is only reported for legal entities. Individuals are not covered, and the value is by default zero for such investors. This problem is addressed by including, in untabulated regressions, the indicator variable *Family*, which captures non-legal entities (i.e. physical persons). It is defined to take the value 1 if the largest owner is an individual or group of individuals. Another caveat is that *Weight* is unavailable for foreign legal entities. We deal with this by checking if including a variable that takes the value 1 if the largest owner is an investor not domiciled in Sweden changes the outcome (*Foreinglarge*). Results are insensitive to including these additional controls.

Booktomarket is defined as the book value of equity divided by the market value of equity (WC03501/WC08001). This variable corresponds to the growth-option hypothesis in Cao et al (2008), according to which we would expect firms with more growth options to be more volatile. *Earnings* is defined as earnings divided by total assets (WC01706/WC02999). While ideally one would use the volatility of earnings, as in Pàstor and Veronesi (2003), this reduces to a single cross-sectional observation since the whole sample data of each firm is required for the calculation. Following Wei and Zhang (2006), we instead use actual earnings, the level of which tends to be correlated with earnings volatility. In addition to the proxies introduced above, we include industry dummies, because certain sectors are likely to have fundamentally higher uncertainty about future earnings potential (Schwert, 2002).

We employ several proxies for information asymmetries. *Bidask* targets information asymmetries between different kinds of investors. It is defined as the average daily difference between the bid and ask-price of the firm's B-shares divided by the price, i.e. the relative spread (Datastream codes: PB and PA). A large spread is typically assumed to indicate the presence of traders who possess superior information about the firm. *Dividend* is a dummy that takes the value one if the firm pays a common dividend in a given year (WC04551). *Size* is defined as the log of total assets (WC02999). Our model also contains listing fixed effects. We calculate dummies for the small cap, mid cap, and NGM lists (*Small*, *Mid*, and *Ngm*, respectively). These different listing options have different information disclosure requirements. Finally, we compute *Leverage* as the book value of debt divided by total assets (WC03255/WC02999).

III. Empirical analysis

A. Descriptive statistics

Table I reports the descriptive statistics for the variables introduced in section II.C. All financial ratios are winsorized at the 97.5 and 2.5th percentiles to mitigate concerns about the impact of outliers. *Microfloat* exhibits useful variation, ranging from a maximum of 92% of all outstanding shares to below 1%. On average, 29.4% of a firm's total outstanding shares consists of very small stakes (<0.1%). It can also be seen that the variable *Weight* has much smaller negative values than positive ones. This is no surprise because it targets the largest owner in the firm, who are very rarely under-exposed to the firm in question (in comparison with the market index).

Table II shows the descriptive statistics for volatility on a yearly basis between 2000 and 2013. It is clear from Table II that volatility displays a large variation over time. As expected, the volatility peaks in the years in which the economy went through turmoil (the stock market crash around 2001 and the financial crisis in 2008).

Table III reports the average yearly cross-sectional Pearson correlations between a subset of variables used in this study. Volatility exhibits clear tendency to co-vary with several of the independent variables with the sign of the correlation largely consistent with the empirical predictions. Size (-), dividend (-), and earnings (-) stand out as highly correlated with volatility. *Nrinstdef* and *Instown* exhibit a negative correlation with volatility, though this is quite possibly due to a positive association with size. Several of the independents are correlated but generally the coefficient does not exceed 0.6, which suggests that multicollinearity is not a problem in this sample.

[INSERT TABLE I ABOUT HERE]

[INSERT TABLE II ABOUT HERE]

[INSERT TABLE III ABOUT HERE]

B. Baseline regressions

Table IV reports the results from the application of Eq. 1 to our data. The model is estimated three times. Model 1 adds the log value of *Nrshareholders* to the baseline of control variables. Model 2 replaces *Nrshareholders* with our two measures of the size of the shareholder base (*Microfloat* and *Nrshlarge*). Model 3 goes one step further and adds our main measures of the heterogeneity of the shareholder base (*Nrinstdef* and *Nrforeigndef*). Model 4 adds lagged volatility to the model as an additional control.

[INSERT TABLE IV ABOUT HERE]

In Model 1 we note that volatility increases in the number of investors (*Nrshareholders*). Similarly, the relation between *Microfloat* and *Nrshlarge*, on the one hand, and volatility on the other, is positive (Model 2). All measures of the shareholder base are highly significant. We interpret these results as evidence against the conjecture that a large shareholder base has a dampening effect on volatility. We return in sections III.C and III.E to a more detailed discussion of these findings.

Model 3 provides evidence that a larger number of institutional owners is associated with higher volatility (significant at the 5%-level).[15] These results are in line with previous studies which, using US data, generally find a positive relation (Campbell et al, 2001; Xu and Malkiel, 2003; Dennis and Strickland, 2009). However, it is worth noting that the variable typically used in the literature is the sum of institutional ownership stakes,[16] whereas we use a count measure that basically weighs each institution equally. We argue that a count measure better approximates institutional trading because a measure based on ownership stakes might be unduly affected by the inclusion of one or more large block holdings. While one should be careful not equate the size of ownership stake with trading frequency, it is quite plausible that the larger the stake, the more strategic and long-term is also the investment. With our measure, on the other hand, a high value signals the presence of a multitude of relatively small institutional holdings, which are likely to be more prone to trade and exhibit the herding behavior that previous literature has ascribed them (Sias, 2004). When we use *Instown*, which is similar to the institutional ownership-variable used in the literature, we find no significant relation. We return in section III.F to the distinction between small and large institutional holdings.

Model 4 shows that the statistical significance of the shareholder base-proxies continue to hold if lagged volatility is included in the model.

In Table IV many of the independent variables are highly significant. Most signs are consistent with expectations. For example, large firms; firms that pay dividends; have low leverage; are well diversified; have a high bid-ask spread; and have a high earnings realizations are associated

[15] The large coefficient is a result of the fact that we deflate with the total number of shareholders, which is a much greater number. In untabulated regressions, we use the log of *Nrinst* as an alternative. The relation is again positive and significant (this time at the 1% level).

[16] We also try this variable in Eq 1, but it is not significant in explaining volatility.

with lower volatility. Interestingly, both forms of corporate diversification (geographical and product segments) are negatively related to volatility. On the whole, variables associated with fundamental uncertainty and information asymmetries do well in terms of explaining volatility, whereas proxies for portfolio concentration (*Weight* and *Sphere*) do poorly. The listing dummies are generally significant with the expected signs. Notably the NGM listing has a large positive coefficient, indicating significant excess volatility for stocks traded on this list. We also note that the model explains variation in volatility well as indicated by the adjusted R^2 of 52%, which compares favorably with previous studies.[17]

One result deserves a brief elaboration. The coefficient on *Booktomarket* is positive, indicating that firms with more growth options have lower volatility.[18] This is at odds with the finding reported in Cao et al (2008), who report that growth indicators are positively related to idiosyncratic risk. We attribute this difference to the empirical design rather than using different samples. Whereas they regress the volatility of a value-weighted portfolio of stocks on a time trend and a proxy for growth option, we estimate a panel regression without such a trend but with a rich set of control variables suggested by theory. Similar to us, but contradicting Cao et al (2008), Rubin and Smith (2009) also find a positive and statistically significant coefficient for the book-to-market ratio in their two sub-samples (dividend-payers and non-dividend payers).

---

[17] For example, Rubin and Smith (2009) report R^2s around 20-30% in their full specifications (Tables 4 and 5), and Li et al (2014) report values between 22 and 37%.

[18] We can appreciate the economic intuition behind the positive sign on *Booktomarket* by recognizing another common interpretation of this variable in the literature, namely that it proxies for firm value, similar to how Tobin's Q is applied in many studies. Viewed this way, a low value for *Booktomarket* can be read as signaling a higher-valued firm with a successful and proven business model. Our result is then suggesting that this "business success" has a stabilizing effect that dominates the higher uncertainty from having a relatively larger share of value coming from future growth options.

As discussed previously, to account for time-variability in stock market participation we also estimate Fama-McBeth regressions with Newey-West corrections. The results are reported in Table V. As can be seen, our main conclusions are unaffected by the change in estimation method. Both *Nrshlarge* and *Microfloat* remain significant (at the 1 and 5%-levels, respectively), as does *Nrinstdef* (5% significance). Also when we use this econometric framework, *Nrforeigndef* fails to achieve significance at conventional levels. The similarity in results can be understood by the results in Pastor et al. (2016) who show that the coefficient estimates from Fama-McBeth are a special case of time fixed effects panel data estimation.

[INSERT TABLE V ABOUT HERE]

C. Endogeneity

Self-selection presents us with an endogeneity-concern regarding the interpretation of the relation between the number of investors and volatility. It is possible, for example, that small shareholders self-select into high-volatility shares if they are prevented from using leverage because of financial constraints and therefore seek higher-risk assets (Frazzini and Pedersen, 2014).[19] To address this issue we use the number of shares (*Nrshares*) as instrument in a 2SLS-approach. This variable has a strong positive correlation with *Microfloat*, and there is no a priori reason why the number of shares

[19] A similar self-selection could occur if smaller investors are simply more risk-seeking. There is, to our knowledge, little in the literature to suggest such a monotonic relationship between the size of the ownership stake and risk-seeking behaviour. Risk aversion is typically assumed to be a function of wealth and other state variables, or psychological factors such as whether the investor operates in the domain of losses (Kahneman and Tversky, 1979).

should be linked to volatility. In our first stage regression we regress *Microfloat* on the same independent variables as reported in Table IV plus the log of *Nrshares*. As expected, *Nrshares* is highly significant (p-value=0.001) with a positive sign in the first stage. In the second stage, we use the predicted values from this regression as our instrument. The coefficient remains significant at the 1% level. Compared to the baseline results in Table IV the effect on volatility is now stronger, suggesting that the causal effect is higher than indicated by the initially estimated coefficient.

The relationship between the number of institutional owners and volatility may also reflect selection into high-volatility stocks. Yan and Zhang (2009) argue that better informed institutions prefer stocks with high idiosyncratic volatility because it allows them to take advantage of mispricing. To investigate this issue we use *Microfloat*, which we consider a direct measure of the importance of small and less informed investors in the shareholder base. We interact *Nrinstdef* and *Microfloat* (results are not tabulated but available from the authors). We expect institutions to have a bigger impact when the microfloat is large, because this would indicate more mispricing due to noise trading. Fig. 1 shows how the effect of *Nrinstdef* on volatility changes for different values of *Microfloat*. It suggests that the relation between *Nrinstdef* and volatility is weaker for high values of *Microfloat*, which contradicts the selection-story.

[INSERT FIG. 1 ABOUT HERE]

D. Survivorship bias

An important aspect of our sample is that it contains data only for firms that are listed as per December 31, 2013. Firms that have been listed in the sample period but for some reason have made an exit prior to this date are not included. A potential critique against our analysis in section III.B is therefore that it suffers from survivorship bias. It is not an unreasonable assumption that those firms that make an exit are fundamentally more risky (and therefore more volatile). For example, a firm may delist because it defaults on its debt and becomes bankrupt. To mitigate concerns about survivorship bias, we carry out two additional regressions in which we change the starting year. The results are reported in Table VI. For ease of comparison, Model 1 contains the result from the estimation using the full sample (and is thus identical with Model 3 in Table IV). In Model 2 the starting year is changed from 2000 to 2005, and in Model 3 it is set to 2010. The more recent the starting year, the more balanced and representative the sample becomes since new and presumably more risky firms enter the sample as we move along towards 2013. According to Table VI, the results with regard to the ownership variables are robust to narrowing in the starting year of the estimation, so survivorship bias appears not to be a major threat to the validity of our results. The impact of *Microfloat* gets larger as the time period is narrowed, suggesting that including more of the smaller and more risky firms increases the importance of this variable.

[INSERT TABLE VI ABOUT HERE]

E. The trading channel explanation

The results in section III.B do not support the hypothesis that a large shareholder base leads to lower volatility. In fact, we find the opposite. A natural way to explain such a result is to argue that a large shareholder base is conducive to more trading, which previous literature has shown to be positively related to volatility (e.g. Rubin and Smith, 2009). Our interest at this point is to see to which extent our proxies for the shareholder base explain trading volume. Any such effect would point in the direction of a trading channel explanation. We also wish to understand if the original relationship between the shareholder base and volatility is accounted for by this effect on volume. These are testable propositions, for which purpose we collect data on annual trading volumes (Datastream code: VO).

To investigate the trading channel explanation, Table VII presents the results from a 3SLS-estimation in which volatility and trading volume are the dependent variables. Our exogenous variables in the volatility-equation are *Geographical, Diversification* and *Intangibles*. We believe these variables explain volatility strictly on a fundamental level (earnings uncertainty) as opposed to the trading channel, and they also meet the relevance-criterion, as indicated by their significance in our baseline regressions[20]. In the volatility-equation we exclude the shareholder base-variables, on the argument that they ought to have an impact on volatility only through their influence on trading volumes.

[INSERT TABLE VII ABOUT HERE]

The results in Table VII support the trading channel-explanation. We find that *Microfloat* and *Nrshlarge* are positively related to trading volumes, and both are significant at the 1%-level. What is more, when we reintroduce the shareholder base variables in the volatility-equation none of them maintains the significance they had in the baseline regressions.[21] That is, when we model volatility

[20] The three variables are jointly significant with p<0.01.

[21] These regressions are not tabulated.

and volume simultaneously, the shareholder base variables lose their ability to explain volatility, suggesting there is no direct effect. We conclude that the effect the shareholder base has on volatility operates through the trading channel (as captured by trading volume).

F. Institutional ownership – a closer look

According to the results in section III.B, having a large number of institutional owners leads to lower volatility. A recent development in the literature has been to explore heterogeneity among institutional investors. Chichernea et al (2015) argue that ignoring institutional heterogeneity can produce confounding results because institutions exhibit differences in trading behavior, which in turn is caused by differences in objectives, limits, and other characteristics of each institution. They show that it is predominantly investors with a short investment horizon that brings about the link between institutional investors and volatility since they have a higher trading intensity.

We contribute to the literature on institutional heterogeneity by looking at how large institutional ownership stakes differ from their smaller counterparts. Moreover, we add a measure that takes into account the degree to which the largest institutions are over-exposed to the company in question. Our results, reported below, support the argument in Chichernea et al (2015) that heterogeneity among institutions are important for understanding stock return volatility.

First, we split institutions into two groups, one containing only the number of large institutional owners, here defined as the number of institutions among the firm's 15 largest shareholders, *Nrinstlarge*. The other group, as represented by *Nrinst*, now captures the number of small stakes held by institutional owners. Since Nrinstlarge has a very low median number we use the log of these variables in this analysis rather than deflating with the total number of shareholders.

We argue that large ownership stakes are more likely to be strategic and long-term investments, thus capturing reduced trading, similar to institutions with a low churn rate (Yan and Zhang, 2009; Chichernea et al, 2015). Model 1 in Table VIII yields results consistent with this argument. Whereas *Nrinst*, as before, enters the models with a positive sign, the coefficient on *Nrinstlarge* is negative. Having many institutions among the *largest* owners therefore has a dampening effect on volatility. This, we believe, is consistent with the institutional preference hypothesis in Rubin and Smith (2009), according to which institutions are restricted by various rules and constraints that make them prefer "prudent stocks", i.e. low volatility.

Large institutional owners may have an even stronger preference for low-volatility stocks if the equity investment represents a large portion of their portfolio of equities (that is, a high portfolio concentration). According to Ekholm and Maury (2014) investors that have a high portfolio concentration (that is, are under-diversified) have stronger incentives to collect information and monitor the firm. Therefore, if the institution is over-exposed to the stock in question we expect volatility to be reduced even further. We construct a measure labeled *Instconc* that weighs each institutional investor's portfolio concentration (our variable *Weight*) with the size of the institutional ownership stake. In this case we focus on the three largest investors. To understand the computation, let us say that the first and second largest investors of the firm are institutions and the third is an individual. The institutions hold 15% and 8% of outstanding stocks, and have portfolio concentrations of 0.5 and 0.25, respectively, indicating over-exposure to the stock. Then our measure would be calculated as: 0.5*15%+0.25*8%+0= 0.1, which indicates that, considered as a group, the largest institutional investors are over-exposed to the firm. Model 2 in Table VIII reports the results the results for *Instconc*. Consistent with expectations, having institutions with a heavy portfolio concentration among the largest investors reduces volatility.

[INSERT TABLE VIII ABOUT HERE]

IV. Conclusions

A major conclusion of this study is that the shareholder base is important for understanding cross-sectional variation in stock return volatility. The main focus has been investigating an intriguing but so far under-researched conjecture which holds that volatility decreases in the size and diversity of the firm's shareholder base because these factors improve the information content of the stock price. We show that various proxies for the size and diversity of the shareholder base in fact increase volatility, contradicting the conjecture. This conclusion is robust to changes in the econometric approach, and holds after taking self-selection into account.

The data strongly supports a trading channel explanation for these findings. We are able to verify that proxies for the size of the shareholder base are positively related to trading volume, and that its effect on volatility goes through this channel. We have therefore identified some of the factors that appear to underlie the association between volume and volatility reported in the literature. We have proposed that the association between having a large shareholder base and high volatility may be indicative of the presence of noise trading. Such trading is often linked to behavioral biases like trend-following and over-confidence. However, we caution against interpreting this evidence as saying that a large and diversified shareholder base is inherently bad. Our results need to be seen in relation to the literature on neglected stocks: somewhat higher volatility could be a price well worth paying to avoid the detrimental value effects from "flying below the radar" of the broader investment community.

References

Aït-Sahalia, Y., P. Mykland, and L. Zhang, 2005, "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial studies* 182, 351-416

Andersen, T. and T. Bollerslev, 1998, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International economic review*, 885-905

Arellano, M., and S. Bond, 1991, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies* 58, 277-297

Baltagi, B.H., 2008, "Econometric Analysis of Panel Data," 4th edition, Wiley.

Baskin, J., 1989, "Dividend policy and the volatility of common stock," *Journal of Portfolio Management* 15, 19-25

Ben-David, I., F. Franzoni, and R. Moussawi, 2014,"Do ETFs Increase Volatility?" NBER Working Paper No. 2007

Black, F., 1976, "Studies of stock price volatility changes," Proceedings of the Journal of American Statistical Association

Black, F., 1986, "Noise," *Journal of Finance* 41, 529-543

Bodnaruk, A. and P. Östberg, 2009, "Does Investor Recognition Predict Returns?" *Journal of Financial Economics* 91, 208-226

Campbell, J.Y., M. Lettau, B. Malkiel, and Y. Xu, 2001, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *The Journal of Finance* 56, 1-43

Cao, C., T. Simin, and J. Zhao, 2008, "Can Growth Options Explain the Trend in Idiosyncratic Risk?" *The Review of Financial Studies* 216, 2599-2633

Chichernea, D., M. Ferguson, and H. Kassa, 2015, "Idiosyncratic Risk, Investor Base, and Returns," *Financial Management* 44, 267-293

Chichernea, D.C., A. Petkevich, and Zykaj, B. B., 2015, "Idiosyncratic Volatility, Institutional Ownership, and Investment Horizon," *European Financial Management* 21, 613-645

Collin, S.O., 1998, "Why are these Islands of Conscious Power Found in the Ocean of Ownership? Institutional and Governance Hypotheses Explaining the Existence of Business Groups in Sweden," *Journal of Management Studies* 356, 719-746

Dennis, P. and D. Strickland, 2009, "The Determinants of Idiosyncratic Volatility," Working Paper

Ekholm, A. and B. Maury, 2014, "Portfolio concentration and firm performance," *Journal of Financial and Quantitative Analysis* 49, 903-931

Faccio, M., M. Marchica, and R. Mura, 2011, "Large Shareholder Diversification and Corporate Risk-Taking," *The Review of Financial Studies* 2411, 3601-3641

Fama, E.F. and K.R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics* 331, 3-56

Foucault, T., D. Sraer, and D.J. Thesmar, 2011, "Individual Investors and Volatility," *The Journal of Finance* 66, 1369-1406

Frazzini, A. and L.H. Pedersen, 2014, "Betting against beta," Journal of Financial Economics 111, 1-25

Greene, W., 2002, "Econometric Analysis," 6th edition, Prentice Hall.

Hansen, P. and A. Lunde, 2006, "Realized variance and market microstructure noise," *Journal of Business & Economic Statistics* 242, 127-161

Holmström, B. and J. Tirole, 1993, "Market Liquidity and Performance Monitoring," *Journal of Political Economy* 101, 678-709

Kahneman, D. and A. Tversky, 1979, "Prospect theory: An analysis of decision under risk," *Econometrica* 46, 171-185

Lee, Y., 2012, "Bias in Dynamic Panel Models under Time Series Misspecification," Journal of Econometrics 169, 54-60.

Li, D., Q.N. Nguyen, P.K. Pham, and S.X. Wei, 2011, "Large Foreign Ownership and Firm-Level Stock Return Volatility in Emerging Markets," *Journal of Financial and Quantitative Analysis* 464, 1127–1155

Merton, R.C., 1987, "*A Simple Model of Capital Market Equilibrium with Incomplete Information*," *Journal of Finance* 42, 483-510

Pástor, L., R. Stambaugh, and L. Taylor, 2016, "Do Funds Make More When They Trade More?" *Journal of Finance*, forthcoming

Pástor, L. and P. Veronesi, 2003, "Stock Valuation and Learning about Profitability," *Journal of Finance* 585, 1749-1790

Petersen, M., 2009, "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies* 22, 435-480

Rubin, A. and D.R. Smith, 2009, "Institutional ownership, volatility and dividends," *Journal of Banking & Finance* 334, 627-639

Schwert, G.W., 2002, “Stock Volatility in the New Millennium: How Wacky Is NASDAQ?” *Journal of Monetary Economics* 49, 3-26

Sias, R.W., 2004, “Institutional Herding,” *Review Of Financial Studies*, 171, 165-206

Singhvi, S.S. and H.B. Desai, 1971, "An Empirical Analysis of the Quality of Corporate Financial Disclosure," *The Accounting Review* 461, 129-138

Shleifer, A. and L. Summers, 1990, “The noise trader approach to finance,” *Journal of Economic Perspectives* 4, 19-33

Wang, J., 2007, "Foreign equity trading and emerging market volatility: Evidence from Indonesia and Thailand," *Journal of Development Economics* 842, 798-811

Wei, S.X., and C. Zhang 2006. "Why Did Individual Stocks Become More Volatile?" *Journal of Business* 791, 259-292

West, K.D., 1988, “Dividend Innovations and Stock Price Volatility,” Econometrica 1, 37-61

Xu, Y., and B. Malkiel, 2003, “Investigating the Behavior of Idiosyncratic Volatility,” *The Journal of Business* 4, 613-645

Yan, X. and Z. Zhang, 2009, "Institutional Investors and Equity Returns: Are Short-term Institutions Better Informed?" *Review of Financial Studies* 22, 893-924

Zhang, C., 2010, "Reexamination of the Causes of Time-Varying Stock Return Volatilities." *Journal of Financial & Quantitative Analysis* 453, 663-684

Table I. Descriptive statistics 2000-2013

This table shows descriptive statistics for the full sample of firms from 2000-2013. *Size* is the natural log of total assets. *Ngm*, *Small*, and *Mid* are binary variables indicating which list the firm's share is trading on. *Bidask* is the average daily difference between the bid and ask-price of the firm's B-shares divided by the price (relative spread). *Dividend* is a binary variable that takes the value 1 if the firm pays a dividend, zero otherwise. *Leverage*is the firm's interest-bearing debt divided by total assets. *Geographical* is the number of geographical segments for which the firm reports sales. *Diversification* is the number of product segments for which the firm reports revenue. *Earnings* is net income divided by total assets. *Booktomarket* is book value of equity divided by market value of equity. *Intangibles* is the ratio of intangible assets to total assets. *Sphere* is a binary variable that takes the value 1 if the firm belongs to one of the 20 main spheres in Sweden, zero otherwise. *Weight* is a coefficient indicating the weight the firm has in the portfolio of the largest shareholder, in relation to the weight it has in a broad index. *Nrshlarge* is the number of shareholders with an ownership stake exceeding 0.1%. *Microfloat* is the sum of ownership stakes below 0.1% as a percentage of the total number of shares. *Nrshareholders* is the total number of shareholders (in thousands). *Instown* is the sum of shares held by institutional investors (among the 15 largest shareholders) divided by the total number of shares. *Nrinst* the number of institutional owners. *Nrinstdef* is *Nrinst* divided with the total number of shareholders. *Nrinstlarge* is the number of institutions among the 15 largest owners. *Nrforeigndef* is the sum of the fraction of shares held by foreign investors. *Leverage*, *Earnings, Booktomarket* and *Intangibles* are winsorized at the 2.5% and 97.5% percentiles. Stdev is the standard deviation and N is the number of firm-year observations.

	Mean	Median	Maximum	Minimum	Stdev	N
Size	14.299	14.046	22.579	4.691	2.388	3119
Ngm	0.033	0.000	1.000	0.000	0.180	3780
Small	0.407	0.000	1.000	0.000	0.491	3780
Mid	0.315	0.000	1.000	0.000	0.465	3780
Bidask	0.022	0.009	1.941	0.000	0.070	2745
Dividend	0.568	1.000	1.000	0.000	0.495	3181
Leverage	0.204	0.168	0.649	0.000	0.188	3106
Geographical	2.433	1.000	10.000	0.000	2.910	3780
Diversification	2.097	1.000	10.000	0.000	2.011	3780
Earnings	-0.005	0.041	0.264	-0.734	0.189	3117
Booktomarket	0.667	0.537	2.171	0.069	0.493	2904
Intangibles	0.200	0.134	0.701	0.000	0.202	3093
Sphere	0.250	0.000	1.000	0.000	0.433	2465

Weight	16.969	0.000	100.000	-0.300	32.393	2463
Nrshlarge	161.186	185.000	200.000	6.000	47.528	2464
Microfloat	29.429	27.000	92.300	0.200	14.809	2465
Nrshareholders (Th.)	24.912	5.822	1023.498	0.052	86.786	2550
Instown	22.531	21.800	84.300	0.000	14.642	2460
Nrinst	274.965	54.000	4324.000	0.000	555.457	2549
Nrinstdef	0.013	0.010	0.096	0.000	0.010	2549
Nrforeigndef	0.050	0.042	0.586	0.000	0.036	2550

Table II. Descriptive stats for volatility by-year

This table shows the descriptive statistics for the total realized volatility (*Volatility*) by year for the sample period 2000-2013. The headings q1, q2, and q3 are the first, second, and third quartiles, Stdev is the standard deviation and N is the number of firms each year.

	Mean	q1	q2	q3	Max	Min	Stdev	N
2000	63.487	36.972	52.731	81.983	209.821	20.703	34.060	139
2001	64.319	37.790	50.820	82.972	195.410	22.599	36.784	150
2002	66.862	38.249	56.086	82.578	211.955	23.164	37.446	161
2003	56.953	29.637	41.239	62.432	591.861	17.192	55.878	168
2004	40.783	23.619	32.220	48.057	171.502	11.781	26.772	170
2005	36.888	23.138	29.940	41.386	146.766	13.042	22.686	180
2006	39.101	27.913	34.299	42.754	132.189	16.659	17.420	189
2007	39.203	29.724	33.872	42.623	164.641	17.475	18.199	205
2008	57.107	45.978	53.230	64.377	197.348	24.407	19.106	222
2009	50.626	37.200	45.274	58.535	126.149	19.807	19.946	231

2010	40.254	29.553	33.638	43.526	136.594	17.091	19.892	231
2011	46.339	34.143	40.754	49.007	303.111	18.811	25.817	238
2012	42.087	27.428	35.408	45.831	321.653	15.502	28.587	247
2013	34.634	21.162	25.825	36.242	165.657	12.158	24.447	249

Table III. Correlation matrix for selected variables

This table presents the average Pearson correlation coefficients calculated each year between selected variables used in this study. The estimates are based on a common sample of observations between 2000 and 2013. For variable definitions, see Table I or section II.C.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Volatility	1.000													
2. Size	-0.479	1.000												
3. Bidask	0.368	-0.469	1.000											
4. Dividend	-0.538	0.479	-0.237	1.000										
5. Leverage	-0.128	0.426	-0.093	0.174	1.000									
6. Earnings	-0.507	0.353	-0.243	0.445	0.029	1.000								
7. Booktomarket	-0.053	0.184	0.109	-0.029	0.218	0.012	1.000							

8. Weight	-0.150	0.237	-0.105	0.142	0.079	0.087	0.034	1.000						
9. Microfloat	0.181	0.084	-0.191	-0.167	-0.079	-0.154	-0.069	-0.103	1.000					
10. Nrshareholders	0.073	0.243	-0.260	-0.086	-0.014	-0.096	-0.108	0.037	0.389	1.000				
11. Instown	-0.196	0.244	-0.196	0.171	0.007	0.129	-0.101	0.026	-0.107	0.201				
12. Nrinstdef	-0.351	0.542	-0.293	0.382	0.215	0.247	-0.058	0.169	-0.149	0.066	0.392	1.000		
13. Ln(Nrshareholders)	-0.150	0.699	-0.395	0.200	0.112	0.101	-0.035	0.124	0.500	0.510	0.154	0.124	1.000	
14. Nrforeigndef	-0.079	0.262	-0.158	-0.007	0.079	0.044	-0.047	0.054	-0.045	0.125	0.203	0.318	0.084	1.000

Table IV. Determinants of stock return volatility

This table presents panel data estimates of slope coefficients in multivariate regressions of stock return volatility on various explanatory variables. Model 4 is a dynamic panel data model that includes lagged volatility (*Volatility-1*) as an explanatory variable. The dependent variable is total stock return volatility (*Volatility*). All explanatory variables are lagged one year. All models contain period and industry fixed effects. Standard errors clustered at the firm level are reported in parenthesis. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. For variable definitions, see Table I or section II.C.

	Model 1	Model 2	Model 3	Model 4
Constant	72.062***	76.902***	76.938***	55.255***
	[5.792]	[8.714]	[6.290]	[5.803]

Volatility(-1)				0.389***
				[0.056]
Size	-2.694***	-1.651***	-1.777***	-1.112***
	[0.612]	[0.635]	[0.435]	[0.379]
Ngm	21.786***	21.616***	23.026***	17.498***
	[6.587]	[5.798]	[6.293]	[4.607]
Small	0.562	0.689	0.516	0.948
	[1.244]	[1.261]	[1.300]	[0.972]
Mid	2.745**	2.589**	2.598**	2.437**
	[1.299]	[1.305]	[1.259]	[0.978]
Leverage	11.462***	12.062***	10.128**	8.827***
	[4.099]	[4.484]	[4.035]	[3.168]
Bidask	92.466***	85.968***	101.337***	47.622***
	[21.791]	[27.613]	[22.198]	[18.015]
Dividend	-5.534***	-5.674***	-5.798***	-2.168**
	[1.383]	[1.403]	[1.352]	[1.013]
Geographical	-0.216	-0.345*	-0.227	-0.201
	[0.178]	[0.182]	[0.181]	[0.147]
Diversification	-0.519*	-0.535*	-0.522*	-0.257
	[0.307]	[0.294]	[0.298]	[0.245]
Earnings	-34.114***	-35.323***	-34.360***	-20.662***
	[5.765]	[5.377]	[5.863]	[5.160]
Booktomarket	3.846***	2.361*	3.743***	3.642***
	[1.351]	[1.387]	[1.436]	[1.168]
Intangibles	3.898	6.217	2.107	0.527
	[3.327]	[5.094]	[3.335]	[2.692]
Sphere	-1.308	0.048	-0.555	-0.790

	[1.141]	[1.250]	[1.129]	[0.856]
Weight	-0.017	-0.012	-0.016	-0.014
	[0.014]	[0.014]	[0.014]	[0.011]
Ln(Nrshareholders)	3.133***			
	[0.656]			
Microfloat		0.126***	0.128**	0.089**
		[0.044]	[0.049]	[0.040]
Nrshlarge		0.037***	0.035***	0.018*
		[0.011]	[0.012]	[0.010]
Nrinstdef			116.579**	138.583***
			[53.927]	[40.644]
Nrforeigndef			-3.316	-2.836
			[12.997]	[10.707]
Period fixed	Yes	Yes	Yes	Yes
Firm fixed	Yes	Yes	Yes	Yes
N	1865	2056	1865	1830
R^2	0.516	0.441	0.516	0.581

Table V. Determinants of stock return volatility, Fama MacBeth regressions

This table presents estimates of slope coefficients in multivariate regressions of stock return volatility on various explanatory variables. The dependent variable is total stock return volatility (*Volatility*). Estimation is done using Fama MacBeth regressions. The models contain industry dummies. Newey West corrected standard errors (using five lags) are reported in parenthesis. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. For variable definitions, see Table I or section II.C.

	Model 1	Model 2	Model 3	Model 4
Constant	59.139***	64.984***	52.759***	31.524***
	[3.432]	[8.416]	[5.058]	[5.066]

Volatility(-1)				0.339***
				[0.025]
Size	-2.763***	-1.945**	-1.432***	-0.771*
	[0.656]	[0.659]	[0.424]	[0.409]
Ngm	18.925**	14.901	22.331***	17.992***
	[8.347]	[9.408]	[3.251]	[3.131]
Small	0.127	-0.266	0.356	0.803
	[0.718]	[0.928]	[1.237]	[1.191]
Mid	1.968**	1.763**	2.203*	2.141*
	[0.753]	[0.797]	[1.285]	[1.237]
Leverage	8.971**	10.399***	10.254***	9.180***
	[3.315]	[3.014]	[2.841]	[2.735]
Bidask	57.554	89.985	101.709***	39.470**
	[49.372]	[55.517]	[16.245]	[16.371]
Dividend	-7.101***	-6.601***	-5.006***	-2.221*
	[1.314]	[1.376]	[1.175]	[1.166]
Geographical	-0.018	-0.126	-0.431**	-0.386**
	[0.106]	[0.124]	[0.186]	[0.178]
Diversification	-0.552*	-0.747**	-0.803***	-0.536**
	[0.278]	[0.263]	[0.280]	[0.270]
Earnings	-31.844***	-34.188***	-35.854***	-22.711***
	[4.088]	[4.964]	[2.822]	[2.872]
Booktomarket	0.921	-0.845	3.462***	2.341**
	[1.475]	[2.024]	[0.986]	[0.954]
Intangibles	10.131**	11.558	3.992	2.842
	[4.185]	[6.877]	[2.478]	[2.392]
Sphere	-1.146	0.484	-0.888	-1.218

	[0.798]	[1.405]	[1.131]	[1.083]
Weight	-0.018**	-0.008	-0.013	-0.009
	[0.007]	[0.009]	[0.014]	[0.013]
Ln(Nrshareholders)	2.711***			
	[0.627]			
Microfloat		0.094**	0.093**	0.048
		[0.042]	[0.037]	[0.036]
Nrshlarge		0.027**	0.038***	0.024**
		[0.012]	[0.012]	[0.011]
Nrinstdef			122.217**	100.579*
			[56.642]	[55.032]
Nrforeigndef			8.944	9.373
			[14.412]	[13.848]
Industry fixed	Yes	Yes	Yes	Yes
N	1865	2056	1865	1830
R^2	0.374	0.326	0.389	0.447

Table VI. Survivorship bias

This table presents estimates of slope coefficients in OLS multivariate regressions of stock return volatility on various explanatory variables. The dependent variable is total stock return volatility (*Volatility*). In contrast to Table IV, in these regressions we change the starting year to address the issue of survivorship bias. Model 1 has the year 2000 as starting year, whereas Models 2 and 3 have the years 2005 and 2010, respectively. As we shorten the time period the sample becomes less impacted by survivorship bias. All models contain period and industry fixed effects. Standard errors clustered at the firm level are reported in parenthesis. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. For variable definitions, see Table I or section II.C.

	Model 1	Model 2	Model 3

Constant	76.938***	51.093***	66.152***
	[6.290]	[6.194]	[7.211]
Size	-1.777***	-1.996***	-2.545***
	[0.435]	[0.461]	[0.576]
Ngm	23.026***	22.977***	38.587***
	[6.293]	[7.413]	[11.575]
Small	0.516	0.186	0.869
	[1.300]	[1.299]	[1.333]
Mid	2.598**	2.399*	2.949*
	[1.259]	[1.390]	[1.634]
Leverage	10.128**	9.671**	15.678***
	[4.035]	[4.239]	[5.373]
Bidask	101.337***	110.502**	59.771
	[22.198]	[44.209]	[50.769]
Dividend	-5.798***	-7.379***	-9.194***
	[1.352]	[1.419]	[2.043]
Geographical	-0.227	-0.083	0.151
	[0.181]	[0.201]	[0.244]
Diversification	-0.522*	-0.671**	-0.853**
	[0.298]	[0.279]	[0.349]
Earnings	-34.360***	-25.445***	-32.691***
	[5.863]	[5.979]	[8.736]
Booktomarket	3.743***	4.475***	4.978***
	[1.436]	[1.518]	[1.870]
Intangibles	2.107	0.221	-3.811
	[3.335]	[3.231]	[3.546]

Sphere	-0.555	-1.085	-1.315
	[1.129]	[1.226]	[1.378]
Weight	-0.016	-0.016	-0.006
	[0.014]	[0.013]	[0.015]
Microfloat	0.128**	0.150***	0.190***
	[0.049]	[0.048]	[0.054]
Nrshlarge	0.035***	0.028**	0.020
	[0.012]	[0.014]	[0.018]
Nrinstdef	116.579**	179.250***	102.035
	[53.927]	[62.955]	[86.777]
Nrforeigndef	-3.316	2.523	-1.425
	[12.997]	[14.052]	[20.968]
Period fixed	Yes	Yes	Yes
Industry Fixed	Yes	Yes	Yes
N	1865	1484	800
R^2	0.516	0.496	0.521

Table VII. Determinants of volatility and trading volume

This table presents a simultaneous estimation of volatility and trading volume by 3SLS. The dependent variables are total stock return volatility (*Volatility*, Equation 1) and the annualized (log of) trading volume (*Trading Volume*, Equation 2). The model contains period and industry fixed effects. Standard errors clustered at the firm level are reported in parenthesis. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. For variable definitions, see Table I or section II.C.

	Equation 1	Equation 2
	Volatility	Ln (Trading Volume)
Constant	81.933***	-5.690**
	[4.890]	[2.871]

Ln (Trading Volume)	4.013***	
	[0.730]	
Volatility		0.056
		[0.035]
Size	-4.426***	0.771***
	[0.709]	[0.084]
Ngm	25.343***	-1.985**
	[2.952]	[0.876]
Small	1.105	-0.167
	[1.154]	[0.112]
Mid	2.567**	-0.115
	[1.181]	[0.146]
Leverage	14.217***	-1.608***
	[2.856]	[0.463]
Bidask	111.315***	-8.095**
	[16.345]	[3.955]
Dividend	-2.625**	-0.443*
	[1.301]	[0.230]
Geographical	0.032	
	[0.132]	
Diversification	-0.575**	
	[0.231]	
Earnings	-31.351***	1.217
	[2.765]	[1.231]
Booktomarket	5.573***	-0.730***
	[1.059]	[0.166]
Intangibles	3.377**	

	[1.697]	
Sphere	0.735	-0.275***
	[1.092]	[0.106]
Weight	-0.017	0.002
	[0.013]	[0.001]
Microfloat		0.026***
		[0.006]
Nrshlarge		0.006***
		[0.002]
Nrinstdef		3.802
		[5.413]
Nrforeigndef		1.593*
		[0.904]
Period fixed	Yes	Yes
Industry Fixed	Yes	Yes
N	1860	1860

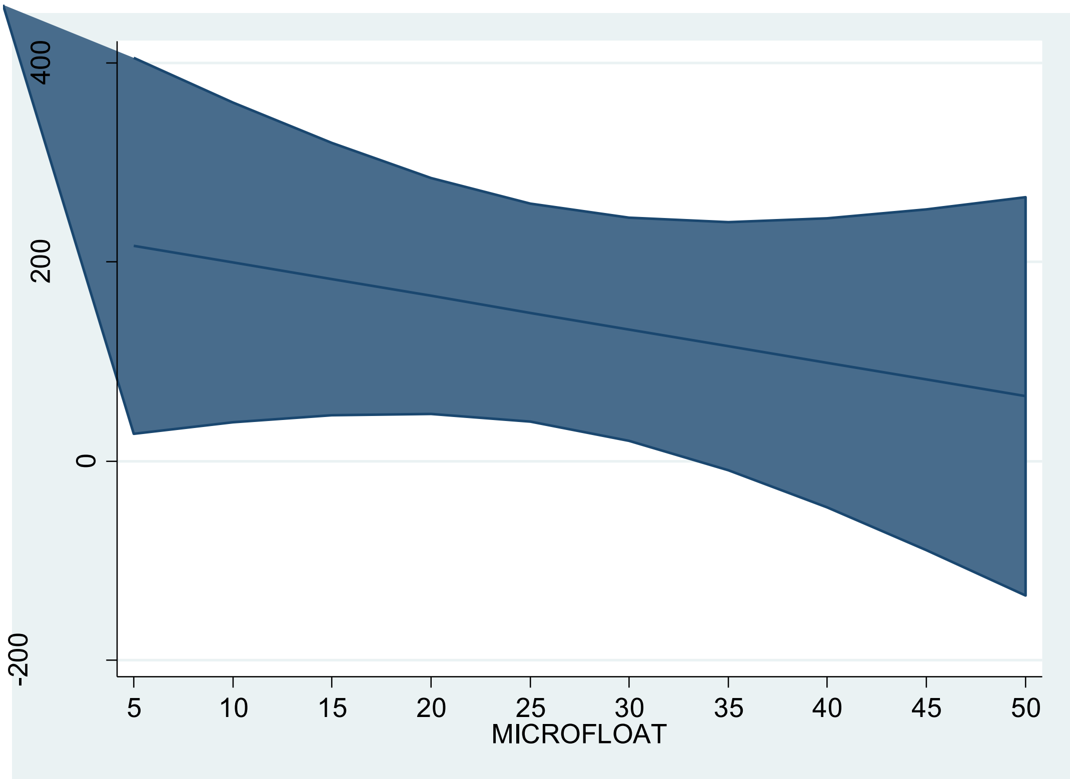

Fig. 1. This figure shows the average marginal effects of *Nrinstdef* for different values of *Microfloat* based on Model 3 in Table IV plus an interaction term between *Nrinstdef* and *Microfloat*. The shaded area is a 95% confidence interval calculated using the delta method.

Table VIII. Institutional ownership and volatility - extensions

This table presents estimates of slope coefficients in multivariate regressions of stock return volatility on various explanatory variables. The dependent variable is total stock return volatility (*Volatility*). We analyze the impact of the number of relatively large institutional owners (*Nrinstlarge*, Model 1) and the impact of a high portfolio concentration of the three largest owners (*Instconc*, Model 2). All models contain period and industry fixed effects. Standard errors clustered at the firm level are reported in parenthesis. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. For variable definitions, see Table I or section II.C.

	Model 1	Model 2
Constant	92.770***	73.860***
	[9.367]	[7.883]
Size	-3.250***	-1.258**

	[0.827]	[0.522]
Ngm	20.940***	24.513***
	[6.349]	[6.844]
Small	0.608	0.717
	[1.262]	[1.263]
Mid	2.865**	2.293*
	[1.302]	[1.271]
Leverage	12.035***	5.165
	[3.981]	[4.392]
Bidask	97.341***	132.776***
	[22.918]	[26.079]
Dividend	-5.475***	-6.325***
	[1.363]	[1.612]
Geographical	-0.161	-0.102
	[0.178]	[0.191]
Diversification	-0.504*	-0.453
	[0.295]	[0.280]
Earnings	-33.648***	-36.289***
	[5.848]	[6.827]
Booktomarket	4.471***	2.498
	[1.533]	[1.545]
Intangibles	2.546	2.041
	[3.219]	[3.727]
Sphere	-1.261	-0.528
	[1.155]	[1.229]
Weight	-0.014	-0.005
	[0.014]	[0.015]

Microfloat	0.069	0.131**
	[0.049]	[0.052]
Nrshlarge	0.029**	0.038***
	[0.013]	[0.014]
Ln (Nrinst)	3.191***	
	[1.046]	
Ln (Nrinstlarge)	-2.609**	-2.287
	[1.204]	[1.486]
Nrinstdef		151.247***
		[54.690]
Instkonc		-0.160**
		[0.070]
Nrforeigndef	9.937	2.385
	[13.810]	[14.603]
N	1865	1470
R^2	0.520	0.543