

Retail Investor Recognition and the Cross Section of Stock Returns

Eric C. Chang^{*} and Chaoli Guo[†]

School of Business, Faculty of Business and Economics, University of Hong Kong

September 15, 2010

ABSTRACT

We test and offer support to Merton's (1987) theory that difference in a stock's investor recognition affects its cost of capital. In the U.S. market, using the breadth of ownership among retail investors as a proxy for investor recognition, we show that a long-short portfolio based on the annual change of shareholder base earns a compounded annual abnormal return of 6.42% after controlling for the Fama-French three factors. These results are more pronounced among young, low visibility and high idiosyncratic volatility stocks, and are robust to various controls such as momentum, breadth of institutional ownership, analyst coverage, liquidity, idiosyncratic volatility, trading volume, accruals, capital investment, probability of informed trading (PIN), and retail investor sentiment. Moreover, we present evidence that the investor recognition effect can explain approximately 20% of the net equity issuance effect documented by Pontiff and Woodgate (2008).

JEL Classification: D83; G11; G12

Keywords: retail investor recognition, ownership breadth, net equity issuance

^{*} Corresponding author. Address: School of Business, Faculty of Business and Economics, University of Hong Kong, Pokfulam Road, Hong Kong. Tel: (852) 2219-4567; fax: (852) 2548-5471; email: ecchang@business.hku.hk.

[†] Tel: (852) 3921-2516; email: guochaoli@business.hku.hk.

1. Introduction

The acquisition of information and its dissemination to all economic units are central activities in capital markets. With limits to information diffusion, all investors cannot distill new information on securities and incorporate that information into prices instantaneously. Ample empirical evidence (French and Poterba, 1991; Coval and Moskowitz, 1999; Hong, et al., 2008) shows that many investors only know about and invest in a subset of the available securities in the investment community.

Merton (1987) proposes a capital market equilibrium model with incomplete information. When constructing optimal portfolios, investors include in their investment choices only those firms with which they are familiar, so the aggregate demand for each firm's stock comes from only a proportion of investors in the market. As investors differ in their "awareness" subsets, each firm's stock is held by a different proportion of investors. In equilibrium, a firm's degree of "investor recognition" is positively associated with the contemporaneous stock price and negatively associated with future expected return. Moreover, because firms held by restricted investors render their shareholders relatively undiversified compared to the optimal market portfolio, the higher risk premium for low visibility firms should be particularly large for stocks with high idiosyncratic volatility to compensate investors for bearing the additional undiversified risk.

The early empirical work that foreshadowed Merton's (1987) model dates back to Arbel et al. (1983), who document a "neglected firm" effect among stocks with few institutional holdings.¹

To facilitate information diffusion to investors, CFOs and investor relations experts in a firm can

¹ Many papers examine events that can eliminate market segmentation due to investment barriers and increase the size of a firm's investor base, such as the initial public offering as the origin of a public-listed firm's investor recognition (Kadlec and McConnell, 1994; Keckes, 2009), non-U.S. firms cross-listed on U.S. exchanges (Foerster and Karolyi, 1999; King and Segal, 2008), inclusion in the S&P 500 index (Chen et al., 2004), a reduction in the minimum trading unit (Amihud et al., 1999), and "sin" stocks restrictedly held by certain institutions (Hong and Kacperczyk, 2009).

adopt a variety of activities to increase the firm's visibility among potential investors, such as competing for analyst coverage (Womack, 1996; Irvine, 2003; Anantharaman and Zhang, 2008), market advertising (Grullon et al., 2004; Lou, 2008; Chemmanur and Yan, 2008), voluntary information disclosures (Botosan, 1997; Botosan and Plumlee, 2002), equity ownership of new institutional investors (O'Brien and Bhushan, 1990), and undertaking public seasoned equity offerings as opposed to rights offerings (Merton, 1987).

Despite the many event studies in search of an investor recognition premium, systematic tests on the empirical cross-sectional relation between investor recognition and expected stock returns have been rare until recently (Bodnaruk and Östberg, 2009; Lehavy and Sloan, 2008). Bodnaruk and Östberg (2009) use the numbers of retail and institutional investors in Swedish firms to proxy for investor recognition, and show that they are important negative predictors of cross-sectional difference in stock returns on the Stockholm Stock Exchange. Employing the change in the breadth of institutional investors in the U.S., Lehavy and Sloan (2008) find that expected stock return is decreasing in this proxy of investor recognition.

To devise a sharper and more powerful test of Merton's (1987) theory in the U.S. market, in this paper we propose a new proxy to measure investor recognition: change in the ownership breadth of *retail* investors, or, simply, shareholder base growth. Each year, the number of retail shareholders is reported in each listed company's 10-K filings and can be collected from the Compustat annual files. We believe that our proxy of investor recognition has merit in three respects.

First, Merton's (1987) incomplete information argument should hold particularly true for retail investors, who are more likely to be time constrained for doing research, with limited access to information and lack of expertise. Second, the availability of data on this proxy enables us to

increase the power of the test by examining a long period of stock returns over 30 years (1977-2008) and for an average of 3,287 firms per month in the U.S. market, as compared to 6 years (1995-2001) and average 243 public firms in Bodnaruk and Östberg's (2009) study of the Swedish market. Third, the breadth of ownership among retail investors provides intuitively better prediction of future returns than that among institutional investors, as in Lehavy and Sloan (2008). In a setting where stock prices are overvalued due to short-sale constraints, the percentage of institutional ownership also proxies for the availability of stocks for borrowing in short-sale activities, thus displaying positive stock return forecasting power (D'Avolio, 2002; Nagel, 2005). An increase in the retail shareholder base would naturally rule out this alternative explanation, as retail investors are hardly the source of stock lending for short-sales.

We show that growth in the retail shareholder base can strongly and negatively explain the cross-sectional difference in stock returns, and that the result is robust to an extensive set of controls and testing methods in a variety of subsamples. A long-short trading strategy that buys the portfolio of lowest shareholder base growth stocks and sells the portfolio of highest shareholder base growth stocks earns a compounded annual abnormal return of 6.42% after controlling for the Fama-French (1993) three-factors in the 12 months after portfolio formation. The shareholder base growth effect is more economically prominent within a particular group of stocks with high idiosyncratic volatility and low investor recognition. Moreover, we show that our main finding survives control for momentum, breadth and percentage of institutional ownership, analyst coverage, liquidity, idiosyncratic volatility, trading volume, accruals, capital investment, and the probability of informed trading (PIN) in a context of Fama-Macbeth regression.

Our study contributes to the growing literature on the determinants of stock price dynamics and lends empirical support to Merton's (1987) model in the U.S. market, by using a novel proxy

and showing a robust cross-sectional relation between retail investor recognition and expected stock returns. It corroborates Merton's (1987) proposition that expanding the breadth of investor cognizance and familiarity through advertising or investor relations is valuable, in the sense that it can reduce the firm's cost of capital.

Another contribution of this paper is that it provides an economic explanation for the net equity issuance puzzle documented by Pontiff and Woodgate (2008). Both theoretical and empirical work (Merton, 1987; Bodnaruk and Östberg, 2008) leads us to investigate the relation between the change in a firm's investor base and a firm's activities in SEOs, stock repurchases, and stock mergers and acquisitions. We show in Section 4 that the investor recognition hypothesis can explain about 20% of the stock return predictability from net equity issuance. At the same time, we conjecture that the investor recognition hypothesis can contain somewhat different information from what the new issues variable can imply. This is broadly supported by the evidence that shareholder base growth remains essentially strong when the equity issuance effect is less likely to apply.

A strand of literature related to our paper explores the return effect of stocks that temporarily attract more attention. Among the variables that have been used to measure investor attention are trading volume (Gervais et al., 2001; Hou et al., 2008), media news and headlines (Meschke, 2004; Yuan, 2008; Fang and Peress, 2009), extreme returns (Barber and Odean, 2008), and active search frequency on the Internet (Da et al., 2009). These evidences suggest that faced with the daunting task of choosing from among thousands of publicly traded firms, investors are easily attracted to firms that have recently grabbed their attention. We demonstrate in Section 5.2 that our proxy for investor recognition differs from these short-run attention variables and has distinct implications for stock returns.

Finally, there is a growing literature on retail investor sentiment that examines the impact of systematically correlated trading among retail investors on stock prices (e.g., Kumar and Lee, 2006; Hvidkjaer, 2008; Dorn et al., 2008; Kaniel et al., 2008; Barber et al., 2009). As a natural extension, it is plausible to infer that such investor herding behavior could cause the annual change in a firm's shareholder base. However, we argue that limited investor recognition reflects investors' limited capacity to process information and is thus not a behavioral bias. Given that the competing theories of investor recognition and investor sentiment predict stock returns in the same direction, we distinguish the two stories by examining firms' operating performances and returns around earnings announcements in portfolio sorts. In Section 5.2, we refute the alternative behavioral story by showing that the low returns to high shareholder base growth stocks are not driven by the irrationality of noise traders.

The remainder of this paper is organized as follows. In Section 2, we describe the data and our key variable for investor recognition. Section 3 presents the main hypothesis tests. Section 4 makes a further distinction between the investor recognition effect and the net equity issuance effect. The possibility of alternative risk- and mispricing-based explanations for our results is explored in Section 5, and Section 6 concludes the paper.

2. Data

Our initial sample includes all U.S.-domiciled ordinary common stocks listed on the NYSE, Amex, and NASDAQ from the Center for Research in Security Prices (CRSP) database. The accounting information is from the Compustat industrial files from 1975 to 2006. The institutional and mutual fund ownership data are from the 13f/s12 common stock holdings files compiled by CDA/Spectrum. The I/B/E/S summary recommendation database is obtained for computing analyst

coverage. We gathered the data on share issuance events, including SEO offerings, repurchase announcements, and merger announcements, from Thomson Financial's SDC Platinum. The yearly PIN estimates for all NYSE/Amex common stocks from 1983 to 2001 were obtained from <http://www.hvidkjaer.net>.

The variable of interest used to measure retail investor recognition is the year-on-year change of a firm's retail shareholder base, with the number of common/ordinary shareholders obtained from the Compustat annual files. The data are derived from the firms' 10-K filings, reported under Item 210(b) of SEC Regulation S-K. The number of shareholders for common equity is usually counted in brokers' names, so if multiple individuals hold a stock through the same brokerage house, they would be identified as one record by the firm. Compustat starts to report the data in 1975, covered by 93% of firms (with 98% of the total market capitalization) on average in our sample. We define the annual shareholder base growth as the log change of the number of shareholders from fiscal year end $t-1$ to fiscal year end t :

$$SBG_t = Ln(\text{Common Shareholders}_t) - Ln(\text{Common Shareholders}_{t-1}) \quad (1)$$

[Insert Fig. 1 Here]

Fig. 1 plots the average firm's level and change in the shareholder base over the sample period from 1975 to 2006.² Panel A shows that the annual median number of shareholders for all firms gradually decreases during the period. This is mainly driven by the public listings of new firms, which comprise the bulk of our sample. When we divide our sample into groups according to their years of listing, we find that each subgroup achieves steady growth in the shareholder base after the firms are traded on the stock market. To look at the phenomenon from another perspective,

² The mean (median) number of shareholders per firm is 16,939 (1,560) for the full sample, and the mean (median) for the log change of shareholders per firm is 3.61% (-1.37%).

a firm's shareholder base is highly correlated with its age – the number of years since the stock's first appearance in CRSP – at 0.51.

The consistent decrease in the average retail investor base for newly issued firms can be accounted for by two facts. First, institutional investors have largely increased their share of the common stock market over the past 30 years (Gompers and Metrick, 2001). Second, the development of the financial market has enabled a larger number of young firms to issue public equity at an earlier stage in their life cycles (Jovanovic and Rousseau, 2001; Fink et al., 2005), so that these firms have a relatively narrow investor base when they are initially exchange-traded.

The evidence shown in Panel B is in line with our interpretation of Panel A. While older firms that were already listed at the beginning of our sample period remain a stable investor base over time, newly listed firms enjoy a rapid expansion in their shareholder bases, especially in the 1980s and 1990s. More importantly, the shareholder base growth is most evident in the early years immediately after stocks are first listed on the public market.

Overall, Fig. 1 suggests that despite the increased ownership of institutional investors, the breadth of retail shareholders can still act as a good proxy for investor recognition.³ Moreover, we expect that the investor recognition effect will be more likely to be identified in the later period of our sample and among young, small, and low investor recognition stocks.

One caveat is that when we use a stock's shareholder base to proxy for its investor recognition, the measure tends to underestimate the true recognition effect. By simply observing that two investors do not own a stock, we cannot distinguish one who does not know about the stock from the other who researches but excludes the stock from her optimal portfolio. In the rationale of Miller's (1977) model with difference of opinions and short-sales constraints, some investors are

³ We show in Section 3.1 that a stock with a larger growth in its retail investor base also tends to be more broadly held by a larger number of institutional investors (mutual funds).

sitting on the “sideline” of the stock market because short-sales constraint prevents them from expressing their negative views by trading in the market. In line with this logic, an expansion in the investor base does not necessarily proxy for a higher degree of investor recognition, but could instead stand for a less severely binding short-sales constraint. When fewer pessimistic investors are shunned away from the market, prices are less likely to be overvalued relative to the fundamentals.⁴ Nevertheless, there is no reason to believe that our proxy’s underestimation of investor recognition will differ across stocks in a systematic manner. Note also that when the competing theories of Merton (1987) and Miller (1977) both apply, they predict stock returns in the opposite directions. Hence, we expect that if there is a bias due to our proxy’s measurement, the results for our empirical tests of Merton’s (1987) model will only be underestimated.

It is worth emphasizing that in the rest of the paper, we focus on the *change* rather than the *level* of shareholder numbers to carry out the test, for two reasons. First, the change variable provides a more comparable measurement for the cross-sectional difference in the degrees of investor recognition, due to the way in which our data are compiled. As mentioned previously, the true number of shareholders can be underestimated when multiple investors trade through a common brokerage, and such underestimation may not be uniform across stocks and could be greater among larger firms. Hence, we expect the change of shareholder numbers to be more representative of a firm’s changing condition of investor recognition, assuming that such underestimation will remain of a similar magnitude within a firm across time.

The second reason has to do with the correlation of our investor recognition variables with other stock return determinants. Merton (1987) argues that for a firm with limited investor base, the shadow cost of incomplete information is a joint function of its idiosyncratic volatility, market value,

⁴ Chen et al. (2002) find that change in the breadth of institutional investors positively, rather than negatively, predicts future stock returns.

and shareholder base. Given that a firm's shareholder base has a strong correlation with idiosyncratic volatility and firm size in our sample, examining the relation between the level of shareholder numbers and expected returns can be confounded by these related factors. By using the change of shareholder numbers, we can largely purge the confounding effects and specifically examine the incremental effect of enhancing investor recognition on lowering a firm's cost of capital.⁵

At the end of June each year between 1977 and 2007, we assign each stock to different portfolios based on its annual shareholder base growth at fiscal year end $t-1$, hold the portfolios from July in year t to June in year $t+1$, and then rebalance the portfolios annually. We exclude stocks with number of shareholders less than 100 and growth rate equal to zero⁶. To eliminate extreme outliers, data are also trimmed at the 1% and 99% percentiles of the shareholder base growth each year. Our final sample consists of 3,287 firms per month on average, which is comparable to those used in previous asset-pricing studies (e.g., Fama and French, 2008).⁷

3. Empirical results

This section presents tests of our major hypothesis that retail investor recognition should be priced at the cross-sectional stock returns, and be more economically prominent among young, low

⁵ The level of shareholder numbers has average cross-sectional correlation with idiosyncratic volatility and size at -0.30 and 0.64, respectively. For change of shareholder numbers, the correlation with idiosyncratic volatility and size drops to 0.04 and 0.03.

⁶ Some firms report the same number of shareholders for two consecutive years, probably due to the brokerage-counting method for data compiling. The number of these firms ranges from 108 to 403 each year during our sample period. On average, they only comprise 1.9% of the total market capitalization. We exclude this part of the sample from the sort-based analysis, but include them in the Fama-Macbeth regression tests later.

⁷ Following Fama and French (2008), to be included in our sample a firm should have a non-negative book equity at the end of fiscal year $t-1$, non-missing data to compute market capitalization for December of $t-1$ and June of t , and the data available to construct our investor recognition measure. Our primary test consists of firms from all industries. The results in this paper are unchanged or become even stronger if we exclude financial, utility firms, closed-end funds, and REITs.

visibility, and high idiosyncratic volatility stocks. We achieve this goal by performing portfolio sorts on both one- and two-dimensions, and then examining the return spreads between the high and low extreme portfolios.

3.1. *Descriptive statistics*

Table 1 reports summary statistics of firm characteristics across shareholder base growth (SBG_{t-1}) groups. Instead of sorting stocks equally into percentiles, we investigate in more detail when a firm's shareholder base either increases or decreases, because we expect that there is a difference in the degree of measuring investor recognition between positive and negative SBG_{t-1} .

We are aware that besides investor recognition, the negative SBG_{t-1} can derive from a number of potential roles. First, along with the increase in institutional ownership (Gompers and Metrick, 2001), a portion of a firm's common shares migrates into the hands of institutions from retail investors. Institutional investors are generally considered to be more sophisticated and hold a more diversified portfolio than individuals. Hence, in this case, the decrease in the retail shareholder base might not represent the severity of under-diversification and loss of investor awareness, but could be due to the increasing participation of professional fund managers.⁸ Second, the negative SBG_{t-1} can also arise from the fact that in a market with difference of opinions, some existing shareholders consider the firm's stock as a short position in their optimal portfolio choices, so they simply sell the stock in hand. Nevertheless, if a firm with reductions in all classes of investors does not actively seek to raise its visibility in the investment community and re-expand its shareholder base, then it could fade out of attention for a larger number of investors and result in an increased cost of capital.

⁸ We further address this issue in Section 5.1 by controlling for the percentage of institutional ownership in testing the investor recognition hypothesis.

In short, whereas a firm's improving visibility can be convincingly viewed as the major driving force behind the positive SBG_{t-1} , the inference from the negative SBG_{t-1} may be less clear. Consequently, we expect the positive SBG_{t-1} to carry more information for testing the investor recognition hypothesis than the negative SBG_{t-1} . Hence, stocks with positive and negative SBG_{t-1} are sorted into quartiles, respectively.

[Insert Table 1 Here]

In Table 1, characteristic variables are reported as the time-series averages of yearly cross-sectional means (or medians) within each portfolio sorted by SBG_{t-1} . There are, on average, 335 stocks for each positive SBG_{t-1} quartile and 488 stocks for each negative SBG_{t-1} quartile per month. The portfolio in the positive high (negative low) SBG_{t-1} quartile has a median of 54% (-17%) log change in the number of shareholders annually. All stocks with positive SBG_{t-1} during the year have, on average, a positive median SBG_{t-2} in the last year, and vice versa.

For firm fundamental and accounting information, we can see from columns SB_{t-2} , $Size$, and AGE that stocks in both the positive high and negative low SBG_{t-1} quartiles are smaller, younger firms that initially have fewer individual shareholders.⁹ In columns B/M , NS_{t-1} , ROE_{t-1} , $CAPEX_{t-1}$, and AGE , as SBG_{t-1} increases from negative to positive, the book-to-market ratio and firm age monotonically decrease, whereas firms' net equity issuance, return-on-equity, and capital investment generally increase, lending support to our argument in Section 2 that growth firms are more inclined to experience an expansion in investor recognition. Firms also spend a larger amount on advertising relative to sales (as reported in column Adv_{t-1}) when SBG_{t-1} increases, consistent with the finding of Grullon et al. (2004).

⁹ To avoid a spurious relation between SBG_{t-1} and SB_{t-1} (both measured at the end of fiscal year $t-1$), we look at SB_{t-2} one year before the SBG_{t-1} is measured.

In the columns related to “Market variables”, stocks with a larger change in the ownership breadth of retail investors also enjoy a larger contemporaneous change in the ownership breadth of institutional investors, denoted as IBG_{t-1} . In columns IO_{t-1} and NA , stocks in both the positive high and negative low SBG_{t-1} quartile portfolios are likely to have lower percentages of institutional ownership and smaller numbers of analysts covering the stocks, presumably because the institutions and analysts tend to follow large-sized firms. Finally, as denoted in columns $IDIVOL_{t-1}$, TO_{t-1} , and $VOLUME_{t-1}$, with the exception in the negative low SBG_{t-1} quartile, stocks with negative SBG_{t-1} generally have lower idiosyncratic volatility, stock turnover and trading volume¹⁰ than stocks with positive SBG_{t-1} .

3.2. *Univariate analysis*

We now turn to understanding the cross-sectional relation between shareholder base growth and expected stock returns. Following the sorting method in Section 3.1, we assign stocks according to SBG_{t-1} at the end of June each year and hold the portfolios over the next 12 months, for both equal-weighted (EW) and value-weighted (VW) schemes. The monthly average raw return and excess returns adjusted for CAPM and the Fama-French (1993) three factors for each portfolio are calculated and reported in Table 2.

[Insert Table 2 Here]

Consistent with our hypothesis, monthly average returns decrease monotonically with SBG_{t-1} across the portfolios. Stocks in the negative low SBG_{t-1} quartile earn an average EW (VW) return of 1.54% (1.32%) per month, and conversely, stocks in the positive high SBG_{t-1} quartile earn an

¹⁰ Because Nasdaq is a dealer market with double counting of dealer buys and sells, the trading volume and turnover of stocks on Nasdaq and NYSE/Amex are not directly comparable. In this paper, the volume and turnover of Nasdaq stocks are divided by two.

average EW (VW) return of 0.95% (0.75%) per month. A long-short portfolio (Neg L - Pos H) that goes long in the negative low SBG_{t-1} quartile and short in the positive high SBG_{t-1} quartile can produce raw and risk-adjusted returns that are all statistically significant at the 1% level. For example, when stocks are equally weighted, the monthly abnormal return of the long-short portfolio after controlling for the Fama-French three factors is 52 basis points and more than 4.65 standard errors from zero. Moreover, the long-short portfolio derives its alpha from both the long and short legs, with 0.20% ($t = 2.23$) for the negative low SBG_{t-1} quartile and -0.33% ($t = 2.65$) for the positive high SBG_{t-1} quartile. For the value-weighted strategy, whereas the FF-3 alpha on the long-short portfolio is 34 basis points per month ($t = 2.71$), it mainly comes from the short end at -0.25% ($t = 2.50$). The magnitude of the monthly risk premium can be translated into a compounded annual return of 6.42% (4.16%) for the EW (VW) portfolio, which is highly economically significant, given that our portfolios are *annually* rebalanced.¹¹

As explained previously, we hypothesize that a stronger negative cross-sectional relation should be found in the positive SBG_{t-1} portfolios. In unreported analysis, within the positive regime of SBG_{t-1} , the differences in the FF-3 alpha between the lowest (Pos L) and highest (Pos H) quartiles of positive SBG_{t-1} are 0.41% ($t = 3.55$) for EW and 0.27% ($t = 1.87$) for VW, both of which are statistically significant. However, within the negative regime of SBG_{t-1} , the differences in the FF-3 alpha between the two extreme quartiles (Neg L and Neg H) are not significant for either portfolio weighting scheme. This is largely in line with our previous interpretation. To examine the effect of positive SBG_{t-1} more closely, by treating all stocks with negative SBG_{t-1} as one group, we form an alternative long-short portfolio (Neg – Pos H) that goes long in all stocks with negative SBG_{t-1} and

¹¹ We test the risk-adjusted return of the long-short portfolio after controlling for Chen, Novy-Marx and Zhang's (2010) q -theory three-factor model. The monthly alpha is -0.27% ($t = -2.17$) for EW but loses statistical significance for VW (although still on the negative sign of -0.16%). We explore this issue further in footnote 14.

short in the positive high SBG_{t-1} quartile stocks. At the bottom of Table 2, we show that this alternative portfolio yields raw and excess returns of similar magnitude with our first long-short portfolio (Neg L – Pos H), with the FF-3 alphas at 0.52% ($t = 5.26$) for EW and 0.27% ($t = 2.59$) for VW.

[Insert Fig. 2 Here]

To examine consistency in the shareholder base growth effect over time, we plot the monthly average raw returns across SBG_{t-1} portfolios in time series, as illustrated in Fig. 2. The two extreme quartiles with positive high (Pos H) and negative low (Neg L) SBG_{t-1} are presented, as well as the spread between these two series (Neg L – Pos H). Monthly returns are averaged within each year and reported for both equal-weighted and value-weighted methods. The graphs show that negative low SBG_{t-1} stocks generally outperform positive high SBG_{t-1} stocks. Over the 32-year sample period, returns of the long-short portfolio are positive for 23 years and 72% of the time, for both weighting schemes.

[Insert Table 3 Here]

Table 3 investigates the robustness of the investor recognition effect by performing portfolio sorts in subsamples. We first formally test the time-series consistency of the effect in two subperiods. We then seek to examine the effect within different firm size groups. In the tests above, we form value-weighted portfolios with the aim to mitigate the concern that the effect is restricted to very small stocks. This is not enough, however. Table 1 shows that stocks within both the largest increase and decrease SBG_{t-1} quartiles are likely to be smaller in market size, so sorting stocks based on SBG_{t-1} in the *whole* sample and examining the returns on the long-short portfolio between the two extremes could still pick up many small stocks and neglect the large stocks, which comprise the majority of the total market capitalization. In the following discussion, for brevity, we focus on

Panel A of Table 3 when equal-weighted returns are used, as the value-weighted returns in Panel B have similar implications.

First, we perform a subperiod analysis in the first two rows of Panel A, in which we report returns for SBG_{t-1} -sorted portfolios during 1977-1992 and 1993-2008. The FF-3 alpha spreads on the long-short portfolio (Neg L – Pos H) are statistically significant in both periods. More importantly, the magnitude is greater in the later period 1993-2008. This confirms the result in Fig. 2 and our previous conjecture that the investor recognition effect should increase in strength over time, when more young and small firms are publicly listed on the market and have a greater incentive to expand their shareholder bases to further reduce the cost of capital.

Second, as an initial attempt to control for firm size, we examine stocks traded only on the NYSE/Amex. We find that excluding NASDAQ stocks does not alter the baseline result. Stocks in the positive high SBG_{t-1} quartile have an average FF-3 alpha of -0.45% per month. The FF-3 alpha for the long-short portfolio (Neg L – Pos H) remains large in magnitude, at 0.49% per month, with a t -statistic of 4.50.

Next, we test the interaction of the investor recognition effect with firm size by conducting a two-way sort. Specifically, we first form Small, Medium, and Large size groups ranked on market capitalization, and then within each size group, we sort stocks into quartiles based on negative and positive SBG_{t-1} separately. The three size groups are determined by the 30th and 70th percentiles of NYSE market capitalization at the end of June of each year t . We find that in each size group, the positive high SBG_{t-1} quartile has a lower FF-3 alpha than the other quartiles. Moreover, the FF-3 alpha of the long-short portfolio is most pronounced among the smallest stock subsample, at 0.50% per month ($t = 4.13$). While this is consistent with the notion that the investor recognition effect

should work especially for small-sized firms, we show that the long-short portfolio has a statistically significant FF-3 alpha even among the large-sized group at 0.32% ($t = 2.64$).

In the last two rows, we perform additional tests by excluding stock returns in January and eliminating stocks with average prices of less than \$5. Again, we find that neither the January returns nor the bid-ask bounces for penny stocks can explain the shareholder base effect. In summary, we can conclude that change in the investor base is a powerful and robust explanatory variable for the cross-sectional difference in expected stock returns.

3.3. *Interaction effect with other stock characteristics*

In this section, to further test Merton's (1987) model, we examine the investor recognition effect within a particular firm characteristic group. As implied by Merton's model, because a firm with only a subset of investors is required to offer additional returns to compensate investors for being imperfectly diversified, the investor recognition premium should be higher for a firm with greater idiosyncratic risk. Moreover, expanding a firm's shareholder base should be particularly important for a stock that originally bears low visibility among retail investors. We adopt four variables to measure firm visibility: analyst coverage, institutional ownership, number of individual shareholders, and firm age. We hypothesize that the investor recognition effect should be most pronounced among stocks with high idiosyncratic volatility, fewer analysts following, low fraction of institutional ownership, low shareholder base, and young age.

[Insert Table 4 Here]

In Table 4, we conduct a similar two-way sort as before. Each month we first sort stocks into three terciles ranked on the firm characteristics, and then within each tercile we sort stocks based on SBG_{t-1} , as we did in the baseline case. The long-short portfolio that goes long in the negative low

SBG_{t-1} quartile and short in the positive high SBG_{t-1} quartile stocks is formed within each characteristic tercile. Average monthly returns on the long-short portfolios within the three characteristic-ranked subgroups are calculated, as is the difference of returns between the high and low characteristic groups.

The evidence shown in Table 4 consistently supports our hypothesis. In Panel A, for EW portfolios, both the raw and risk-adjusted returns on the long-short portfolio are largest in absolute magnitude and statistical significance within the high *IDIVOL* group. This is even more so when VW portfolios are used. The FF-3 alpha on the long-short VW portfolio in the high *IDIVOL* group is remarkably high at 1.06% per month ($t = 3.43$), 1.00% greater than that in the low *IDIVOL* group.

In Panels B, C, D, and E, returns on the long-short portfolios broadly decrease with analyst coverage, institutional ownership, shareholder base, and firm age in absolute magnitude, especially for the VW portfolios. The long-short VW portfolios in small *NA*, low *IO*, low *ShareBase*, and young *AGE* group can yield both statistically and economically significant monthly FF-3 alphas at 1.02%, 1.14%, 0.74%, and 0.37%, respectively.

The overall message from this section is that change in the size of a firm's retail investor base can robustly explain the cross-sectional difference of expected stock returns. The return predictability is particularly strong for firms with initially low level of investor visibility and increasing shareholder base, which confirms Merton's (1987) proposition that expanding the breadth of investor cognizance and familiarity through advertising or investor relations activities is valuable for reducing the firm's cost of capital.

4. Shareholder base effect and net equity issuance effect

Pontiff and Woodgate (2008) and McLean et al. (2009) document a strong predictive power of share issuances for cross-sectional stock returns, both in the U.S. and international equity markets. A well-established strand of literature argues that the share issuance effect is caused by investors' underreaction to managers' market timing activities (Ritter, 2003; Greenwood and Hanson, 2010). Recently, several papers have aimed to explain this using an investment-based argument (Lyandres et al., 2008; Chen et al., 2010). In this paper, we posit a new hypothesis to (partly) explain the net equity issuance anomaly — the investor recognition hypothesis.

Firms take the shareholder base into consideration when they decide their equity issuance and repurchase policies. For example, Merton (1987) asserts that when planning for an SEO, negotiated underwriting incurs much higher underwriting expenses than rights offering. However, with negotiated underwriting, a firm can take advantage of investment banks' investor distribution channels to market new equity to other-than-existing shareholders, thereby expanding the firm's investor base. Consider another example. Bodnaruk and Östberg (2008) show that in regard to payout policies, firms with low shareholder base choose to pay dividends rather than repurchase shares to avoid further reducing the investor base.

[Insert Table 5 Here]

In Table 5, we regress SBG_t on the contemporaneous net equity issuance (NS_t) during fiscal year t , as well as with the occasional events that can significantly broaden or narrow a firm's investor base, in a Fama-Macbeth cross-sectional framework. SEO , REP , MAT , and MAA denote dummy variables for whether the firm experiences an SEO, a repurchase, being a target or an acquirer in a stock M&A in fiscal year $t-1$, t and $t+1$, as indicated in their subscripts. As reported in model 1, NS_t is highly correlated with SBG_t at the cross section, with the highest t -statistic in all models. In model 2, firms undertaking an SEO in year $t-1$ and year t realize a considerable increase

in SBG_t of year t , and a larger investor base of year t can further support the success of a future SEO in year $t+1$. In model 3, Firms undertaking a share repurchase in the last and current year achieve a slight decrease in SBG_t this year, but are less likely to repurchase shares next year, consistent with Bodnaruk and Östberg's (2008) viewpoint. Model 6 shows that during the current year t , on average, an SEO increases a firm's shareholder base by 12.8%, a repurchase decreases SBG_t by 1.4%, being a target in an M&A decreases SBG_t by 14.9%, and being an acquirer increases SBG_t by 5.9%.

In light of the strong correlation between shareholder base growth and net equity issuance, we explore the explanatory power of SBG for the equity issuance effect, by examining the NS -sorted portfolio returns before and after controlling for SBG . For stock returns from July of year t to June of year $t+1$, SBG and NS are both defined as the log change in the number of shareholders and shares outstanding from fiscal year end $t-2$ to $t-1$.

[Insert Table 6 Here]

In Panel A.1 of Table 6, we first form equal-weighted quartile portfolios by sorting stocks based on NS . A high minus low NS -spread portfolio yields monthly raw return at -0.66% and abnormal return at -0.86% (-0.66%) after adjusting for the CAPM (FF-3) model. Next, in Panel A.2, we control for the SBG effect by first sorting stocks into quartiles based on SBG and then, within each SBG group we sort stocks into quartiles based on NS . Returns on the NS -sorted quartile portfolios are then averaged across the four SBG groups and reported. Compared to the returns in one-way sorts, average abnormal returns in both the high and low NS quartiles become smaller in absolute magnitude. Specifically, monthly raw and abnormal returns on the high minus low NS -spread portfolio become -0.54% and CAPM (FF-3)-adjusted -0.71% (-0.53%). We define $\Delta\alpha$ as the difference of high minus low NS -spread returns between the one- and two-way sorts. Column $\Delta\alpha$ indicates that after controlling for SBG , the decreases in NS spread returns (in absolute terms) are all

statistically significant on raw and risk-adjusted basis. By measuring $|\Delta\alpha|$ as a percentage of the original *NS*-spread returns $|\alpha|$ in the single-sort, the last column $|\Delta\alpha|/|\alpha|$ shows that *SBG* accounts for about 20% of the *NS*-return relation on FF-3 risk-adjusted basis.¹²

Yet, despite the link between shareholder base growth and equity issuance, we argue that the two variables can contain different information for explaining the cross-sectional difference in stock returns. Whereas *NS* could to some extent proxy for the private information owned by the firm and manifested through managers' market timing activities, *SBG* generally measures the degree of acknowledgement by the majority of investors in the market, who acquire information about the firm mainly through public channels. This visibility of the public information is valuable in the sense that it can affect investors' demand for the firm's equity and increase its firm value. Firms can implement many strategies to raise investor awareness, such as voluntary information disclosure and analyst coverage initiation (as discussed previously), without necessarily going for a new equity issuance.

Consequently, we expect the shareholder base effect to be particularly stronger when a firm's total number of shares outstanding remains stable. In Panel B of Table 6, we run monthly Fama-Macbeth cross-sectional regressions of individual stock returns on lagged *SBG* and *NS* (defined the same as before), as well as typical controls used in previous literature, including market cap, book-to-market equity, and 12-month lagged returns (Jegadeesh and Titman, 1993). The time-series means of the estimated coefficients and intercept are reported, and the *t*-statistics are calculated based on time-series standard errors of the monthly slopes and corrected for Newey-West (1987) autocorrelation with 12-month lags. First, in model 1, the strong statistical significance of

¹² In a similar vein, Lyandres *et al.* (2008) find that adding an investment factor into standard factor regressions can reduce the composite issuance effect by 40%.

SBG in the presence of *Size*, *B/M*, and *MOM* confirms our result of portfolio sorts in Section 3.2. In model 2, when we put both *SBG* and *NS* together with *Size*, *B/M*, and *MOM* to jointly explain the stock returns, our *SBG* measure survives the control of *NS* at the 10% significance level in the full sample. In model 3, we include the same set of variables as in model 2 but exclude stocks that have experienced SEOs, repurchases, and stock mergers in the past one year. This yields a stronger result on the coefficient of *SBG* than in model 2, which is now significant at the 5% level. Finally, with the inclusion of only stocks with zero change in *NS*, as shown in model 3, *SBG* retains its predictive power at the 10% significance level.

5. Alternative risk- and mispricing-based explanations

In this section, we perform a list of robustness checks to rule out alternative risk- or behavioral-based explanations for the investor recognition effect we document in the paper. We first run Fama-Macbeth cross-sectional regressions to control for a set of firm characteristics, which have been shown to have predictive power for cross-sectional stock returns and may be correlated with our proxy for investor recognition. We then proceed to examine firms' operating performances and returns around quarterly earnings announcements before and after the portfolio sorts, to further distinguish the competing hypotheses between retail investor recognition and retail investor sentiment. In general, we refute these alternative explanations and affirm that the shareholder base growth effect is consistent with Merton's (1987) market equilibrium model with incomplete information.

5.1. Fama-Macbeth cross-sectional regressions

In the following tests, we run monthly cross-sectional regressions of stock returns on shareholder base growth and a set of control variables, as we did in Panel B of Table 6. We first consider *Size*, *B/M*, and *MOM* as controls in the base set of tests, and then the other variables. As with the sorting method, we perform the regressions on all firms and separately on small, medium, and large size-sorted groups. We find that our measure of retail investor recognition is not subsumed by these alternative return determinants.

[Insert Table 7 Here]

Size, *Book-to-market*, *Momentum*. In model 1, after we add *Size*, *B/M*, and *MOM* in the regressions, *SBG* remains a strong predictor for stock returns across all size groups in Panels A, B, C, and D. Note also that *B/M* actually loses forecasting power within large firms, and *Size* exhibits a steady decrease in significance from the small to large groups. Surprisingly, *SBG* possesses a similar explanatory power with *MOM*, even though the *SBG* variable is constructed with six to eighteen month lags and is only updated annually rather than monthly for *MOM*.

Breadth of Institutional Ownership. Presumably, the limitation of investor cognizance can apply to both retail and institutional investors, though we argue that it should particularly work for retail investors. In model 2, we add growth in the ownership breadth of mutual funds (*IBG*) besides the base set of controls in regressions. Consistent with the fact that the return predictability of *IBG* is inconclusive in the literature (Chen et al., 2002; Lehavy and Sloan, 2008), the coefficients on *IBG* are insignificant when we have *SBG* in the regressions. More importantly, *SBG* retains its economical and statistical significance in predicting future stock returns across the full sample and size-sorted subsamples.

Analyst Coverage. Security analysts are considered as an important channel to increase the firm's information diffusion and visibility in the market (Womack, 1996; Irvine, 2003;

Anantharaman and Zhang, 2008). In model 3, controlling for the number of analysts following, we find that in all firms and size subsamples, the magnitudes and statistical significances of the coefficients on *SBG* are not materially affected as compared to the baseline results in model 1, with *t*-statistics ranging from -2.25 to -3.19 in the four panels.

Percentage of Institutional Ownership. As previously discussed, a decrease in the retail shareholder base can be a mechanical consequence of an increase in institutional ownership, and vice versa. Gompers and Metrick (2001) show that the increased equity holdings of institutional investors positively predict returns and can account for the superior performance of large firms during the 1980s and 1990s. However, we show in model 4 that by controlling for the *level* of *IO*, the explanatory power of *SBG* remains robust in all panels. In an unreported analysis, we also find that the results on *SBG* are unchanged by controlling for the *change* of *IO*.

Liquidity. Mounting empirical evidence suggests that a stock's liquidity improves when the firm begins to have an expansion in its shareholder base (Kadlec and McConnell, 1994; Amihud et al., 1999; Grullon et al., 2004). We adopt stock turnover as a rough proxy for liquidity and include it in the regressions of model 5. The average coefficients on *SBG* survive the control of turnover and maintain a strong explanatory power for stock returns.¹³

Idiosyncratic Volatility. Ang et al. (2006, 2009) find a puzzling negative relation between idiosyncratic volatility (*IDIVOL*) and future returns. In Section 3.3, we show that returns on the *SBG*-spread portfolios are presented across *IDIVOL*-sorted groups. In model 6, we further test the relation in the regressions controlling for *IDIVOL*. In all samples, the coefficients on *SBG* are unaffected in both magnitude and statistical significance.

¹³ We use Amihud's (2002) illiquidity measure (*ILLIQ*) as another proxy for liquidity, and the results on *SBG* still hold. *ILLIQ* is defined as the ratio of daily absolute return to (dollar) trading volume on that day, averaged over the past 12 months and used to predict returns in the next month. Trading volume is divided by two for Nasdaq stocks.

Trading Volume. Barber and Odean (2008) show that individual investors are attracted to and are net buyers of stocks experiencing high abnormal trading volumes. Gervais et al. (2001) show that these stocks have high returns in the following month, because increased attention raises demand for them over the short horizon. As indicated in Table 1, our measure of investor recognition is quite persistent over time. If a positive *SBG* in year $t-1$ proxies for expected higher investor attention in year t , manifested in the form of high trading volume, then the positive relation between returns and volume in the one-month window of year t can obscure the negative relation between returns and *SBG* that we document in this paper. Consistent with this conjecture, in model 7, the addition of trading volume leads to stronger inference on *SBG*. For example, in Panel A, the coefficient on *SBG* increases in both absolute magnitude and statistical significance, from -0.392 ($t = -3.05$) in baseline model 1 to -0.453 ($t = -3.93$) in model 7.

Profit and Accruals. Apart from trading volume, retail investors can be attracted to a stock because of its noteworthy operating performance, such as return on equity. However, investors tend to pay less attention to information in earnings that requires greater cognitive processing ability, and thus underreact to information contained in accruals (Hirshleifer and Teoh, 2003). Considering the return explanatory power of profitability and accruals documented by Cohen et al. (2002) and Sloan (1996), in models 8 and 9 we show that controlling for *PROFIT* and *ACCR* does not materially alter the results on *SBG*. This corroborates our supposition that the negative relation between returns and *SBG* is derived from increased investor awareness, rather than retail investors' inattention to complicated earnings disclosures.

Capital Investment. In Merton's (1987) model, a firm's expansion in investor base and increase in capital investment tend to coincide, arising from the same driving force – a reduced cost of capital. The neoclassical q -theory of investment and the real options theory advanced by

Cochrane (1991) and Carlson et al. (2004) postulate a negative relation between real investment and expected returns, and they are attested by ample empirical studies from Titman et al. (2004), Anderson and Garcia-Feijoo (2006), and Chen, Novy-Marx and Zhang (2010). However, when we add *CAPEX* to the base set of controls in model 10, the statistical significance of *SBG* remains strong across all panels, while *CAPEX* is statistically insignificant in medium- and large-sized groups. This implies that our variable of shareholder base growth offers an alternative measure for the change in a firm's cost of capital in a more timely fashion, and hence contributes to the determination of expected returns beyond what capital investment can imply.¹⁴

Probability of Information-based Trading (PIN). Easley et al. (2002) show that *PIN* can positively predict future stock returns. When a firm's ownership structure shifts from a limited number of blockholders to a large number of investors, a concomitant decrease in *PIN* is anticipated, because most retail investors obtain information about a stock through public channels and thus stand a higher chance of being uninformed traders. Using the data of *PIN* for stocks traded on NYSE/Amex, the evidence shown in model 11 refutes the competing premise. For all subsamples, the coefficients on *SBG* retain their economical and statistical significance.¹⁵

5.2. *Other mispricing-based tests*

Our results so far show that returns of stocks sorted on shareholder base growth cannot be explained by the Fama-French (1993) three-factor model, and neither can they be explained by a

¹⁴ As *SBG* and *CAPEX* carry in part the same information about a firm's cost of capital, controlling *CAPEX* in the regressions may reduce the explanatory power of *SBG*. This is also supported in the data. For example, in all firms, the coefficient on *SBG* drops in both absolute magnitude and statistical significance, from -0.392 ($t = -3.05$) in baseline model 1 to -0.261 ($t = -2.19$) in model 10. This result is also consistent with the reduced magnitudes of *SBG*_{*t*-1} spread returns after adjusting for the Chen *et al.* (2010) three-factor model, as discussed in footnote 11.

¹⁵ We only include NYSE/Amex stocks in model 11. It is reasonable to argue that NASDAQ stocks may have a higher probability of informed trading due to NASDAQ's less stringent listing requirements, but they cannot be included in the test due to the constraints on constructing the *PIN* measure.

variety of well-documented stock return determinants. However, could they be caused by some market inefficiencies such as behavioral biases or market mispricing? In this section, we attempt to refute such alternative explanations in three respects.

[Insert Fig. 3 Here]

First, we show that distinct from short-run attention-grabbing variables such as trading volume, media news and headlines, or extreme returns, shareholder base growth is quite a persistent measure and can reliably predict stock returns for a variety of holding periods. In Panel A of Fig. 3, we plot the average shareholder base growths around portfolio formation year. Specifically, at the end of each year, we sort stocks based on the annual *SBG* during the fiscal year (we denote this as Year -1) as in Section 3.1, and we calculate the annual median *SBG* every year for 10 years around Year -1 for each *SBG*-sorted portfolio. Finally, each portfolio's *SBGs* in the (-5, +5) year window are averaged across time and plotted. We can see that stocks with positive *SBG* in Year -1 tend to experience a sustainable growth in their shareholder bases in future years, though at a more moderate rate.¹⁶ Similar patterns can be found among stocks with negative *SBG*. The stability of shareholder base growth over a long horizon distinguishes itself from the aforementioned attention variables.

We also examine the robustness of the shareholder base growth effect for alternative measurement and return prediction horizons. In an unreported table, when we sort portfolios using the total change of shareholder base from fiscal year $t-3$ ($t-4$) to $t-1$ and hold the portfolios from July of year t to June of year $t+1$, the long-short portfolio (as in Section 3.2) can still generate a significant FF-3 spread alpha of 0.33% (0.33%) per month. Moreover, when we use portfolio sorts

¹⁶ We also construct a probability transition matrix for *SBG* in the (-5, +5) year window. Stocks within the top quartile of positive *SBG* in portfolio formation Year -1 have on average 65% (57%) probability of maintaining a positive shareholder base growth 1 (2) year(s) after portfolio formation.

based on the annual change of shareholder base in year $t-1$ (as in our baseline case) to predict stock returns from July of year $t+1$ ($t+2$) to June of year $t+2$ ($t+3$), the long-short portfolio can produce a significant monthly FF-3 spread alpha of 0.30% (0.24%) in the second (third) year after portfolio formation.

Second, the burgeoning literature on retail investor sentiment casts doubt on the rationality of retail investors' buying activities. Different from investor recognition, a few behavioral biases can cause the clustering of retail investors buying a stock. For example, investors are likely to be subject to extrapolative errors, whereby they extrapolate past firm (accounting) performance too far into the future and correct their errors later (Lakonishok, Shleifer, and Vishny, 1994). Firms that are welcomed and bought by these overly-optimistic investors may not be able to sustain the same profitability in the future, as compared to the high growth they originally see.

As a result, we examine the firm operating performance in the $(-5, +5)$ year window around the portfolio formation year, as we did for the shareholder base growth. Operating margin is defined as operating income before depreciation divided by contemporaneous net sales. Panel B of Fig. 3 indicates that although stocks with positive *SBG* have, on average, higher operating margin than stocks with negative *SBG*, each portfolio's operating margin remains relatively stable and does not appear to change significantly before and after the portfolio formation year (Year -1). A rigorous statistical test ascertains that the differences in operating margins between Year -1 and Year +1 are not significant for both positive high and negative low *SBG* portfolios.

Third, if investors are subject to extrapolative errors, it is possible that firms' future earnings are not in line with investors' original expectation. To test whether investors are later surprised by earnings news, following La Porta et al. (1997) and Cooper et al. (2008), we examine stock returns around quarterly earnings announcement dates (EADs) during Year +1. Specifically, we compare

the mean daily return for the 3 days around the EADs with that for all non-EADs in each *SBG*-sorted portfolio.¹⁷

In an unreported analysis, for stocks in the top quartile of positive *SBG*, the mean daily EAD and non-EAD returns are 0.08% and 0.07% respectively, with the difference at an insignificant *t*-statistic of 0.45. The equally distributed daily returns among EADs and non-EADs during the post-portfolio-formation year have far-reaching implications, because they confirm the conjecture that the low returns to high *SBG* stocks are based on reduced risk premium, rather than the extrapolation theory that investors overreact to past firm performance.

In contrast, for stocks in the bottom quartile of negative *SBG*, the mean EAD and non-EAD returns are 0.19% and 0.08% respectively. The 11 basis point difference in daily returns is statistically significant at the 1% level ($t = 4.27$). The distinct stock outperformance during EADs can be interpreted as earnings announcements serving as a salient event that can raise investors' short-term attention, especially for firms with very low recognition in the market.

6. Conclusion

In this paper, we have provided robust evidence that change in the retail shareholder base can negatively explain the cross-sectional difference in stock returns, lending empirical support to Merton's (1987) hypothesis that limits to information diffusion play an important role in the capital market. The premise of Merton's model, which specifies that the dissemination of information may not be perfectly efficient due to the limited processing ability of market participants, applies to

¹⁷ EADs are gathered from the quarterly Compustat data. There are usually 3 to 4 EADs in Year +1, and we require at least three daily returns around EADs each year.

many situations. Our paper joins this strand of research by showing that this information dissemination differential is priced at the cross section of equity market.

Moreover, as firms take their investor bases into consideration when issuing and repurchasing their own stocks, we argue and present empirical evidence that the investor recognition theory can partly explain the net equity issuance effect documented by Pontiff and Woodgate (2008) and McLean et al. (2009).

Finally, although it is hard to completely rule out the possibility that our proxy reflects not investor recognition but some kind of behavioral biases on the part of sentiment investors, which operate in a way that they are partly captured by the shareholder base growth variable, our extensive robustness checks generally support the major hypothesis that is at the core of Merton's (1987) rational equilibrium model with incomplete information.

References

- Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5, 31-56.
- Amihud, Y., Mendelson, H., Uno, J., 1999. Number of shareholders and stock prices: Evidence from Japan. *Journal of Finance* 54, 1169-1184.
- Anantharaman, D., Zhang, Y., 2008. Cover me: Managers' response to decreases in analyst coverage. Working paper, Columbia Business School.
- Anderson, C.W., Garcia-Feijoo, L., 2006. Empirical evidence on capital investment, growth options, and security returns. *Journal of Finance* 61, 171-194.
- Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61, 259-299.
- Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: international and further U.S. Evidence. *Journal of Financial Economics* 91, 1-23.
- Arbel, A., Carvell, S., Strebel, P., 1983. Giraffes, institutions and neglected firms. *Financial Analysts Journal* 39, 57-63.
- Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies* 21, 785-818.
- Barber, B.M., Odean, T., Zhu, N., 2009. Do retail trades move markets? *The Review of Financial Studies* 22, 151-186.
- Bodnaruk, A., Östberg, P., 2008. The shareholder base and payout policy. Working paper, University of Notre Dame and Norwegian School of Economics and Business Administration.
- Bodnaruk, A., Östberg, P., 2009. Does investor recognition predict returns? *Journal of Financial Economics* 91, 208-226.
- Botosan, C.A., 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72, 323-349.
- Botosan, C.A., Plumlee, M.A., 2002. A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research* 40, 21-40.
- Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance* 59, 2577-2603.
- Chemmanur, T.J., Yan, A., 2009. Advertising, attention, and stock returns. Working paper, Boston College and Fordham University.

- Chen, H., Noronha, G., Singal, V., 2004. The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. *Journal of Finance* 59, 1901-1929.
- Chen, J., Hong, H., Stein, J.C., 2002. Breadth of ownership and stock returns. *Journal of Financial Economics* 66, 171-205.
- Chen, L., Novy-Marx, R., Zhang, L., 2010. An alternative three-factor model. Working paper Washington University in St. Louis, University of Chicago, and University of Michigan.
- Cochrane, J.H., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance* 46, 209-237.
- Cohen, R.B., Gompers, P.A., Vuolteenaho, T., 2002. Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of Financial Economics* 66, 409-462.
- Cooper, M.J., Gulen, H., Schill, M.J., 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63, 1609-1651.
- Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54, 2045-2073.
- Da, Z., Engelberg, J., Gao, P., 2009. In search of attention. Working paper, University of Notre Dame and University of North Carolina.
- D'Avolio, G., 2002. The market for borrowing stock. *Journal of Financial Economics* 66, 271-306.
- Dorn, D., Huberman, G.U.R., Sengmueller, P., 2008. Correlated trading and returns. *Journal of Finance* 63, 885-920.
- Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? *Journal of Finance* 57, 2185-2221.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3-56.
- Fama, E.F., French, K.R., 2008. Dissecting anomalies. *Journal of Finance* 63, 1653-1678.
- Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. *Journal of Finance* 64, 2023-2052.
- Fink, J., Fink, K., Grullon, G., Weston, J.P., 2005. IPO vintage and the rise of idiosyncratic risk. Working paper, Rice University, Houston, TX.
- Foerster, S.R., Karoly, G.A., 1999. The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *Journal of Finance* 54, 981-1013.

- French, K.R., Poterba, J.M., 1991. Investor diversification and international equity markets. *The American Economic Review* 81, 222-226.
- Gervais, S., Kaniel, R., Mingelgrin, D.H., 2001. The high-volume return premium. *Journal of Finance* 56, 877-919.
- Gompers, P.A., Metrick, A., 2001. Institutional investors and equity prices. *The Quarterly Journal of Economics* 116, 229-259.
- Greenwood, R.M., Hanson, S.G., 2010. Characteristic timing. Working paper, Harvard University.
- Grullon, G., Kanatas, G., Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies* 17, 439-461.
- Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36, 337-386.
- Hong, H., Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93, 15-36.
- Hong, H., Kubik, J.D., Stein, J.C., 2008. The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics* 90, 20-37.
- Hou, K., Peng, L., Xiong, W., 2008. A tale of two anomalies: The implications of investor attention for price and earnings momentum. Working paper, Ohio State University, City University of New York, and Princeton University.
- Hvidkjaer, S., 2008. Small trades and the cross-section of stock returns. *The Review of Financial Studies* 21, 1123-1151.
- Irvine, P.J., 2003. The incremental impact of analyst initiation of coverage. *Journal of Corporate Finance* 9, 431-451.
- Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48, 65-91.
- Jovanovic, B., Rousseau, P.L., 2001. Why wait? A century of life before IPO. *The American Economic Review* 91, 336-341.
- Kadlec, G.B., McConnell, J.J., 1994. The effect of market segmentation and illiquidity on asset prices: Evidence from exchange listings. *Journal of Finance* 49, 611-636.
- Kaniel, R.O.N., Saar, G., Titman, S., 2008. Individual investor trading and stock returns. *Journal of Finance* 63, 273-310.
- Kecskes, A., 2009. Initial public offerings: The origin of investor recognition? Working paper, Virginia Polytechnic Institute and State University.

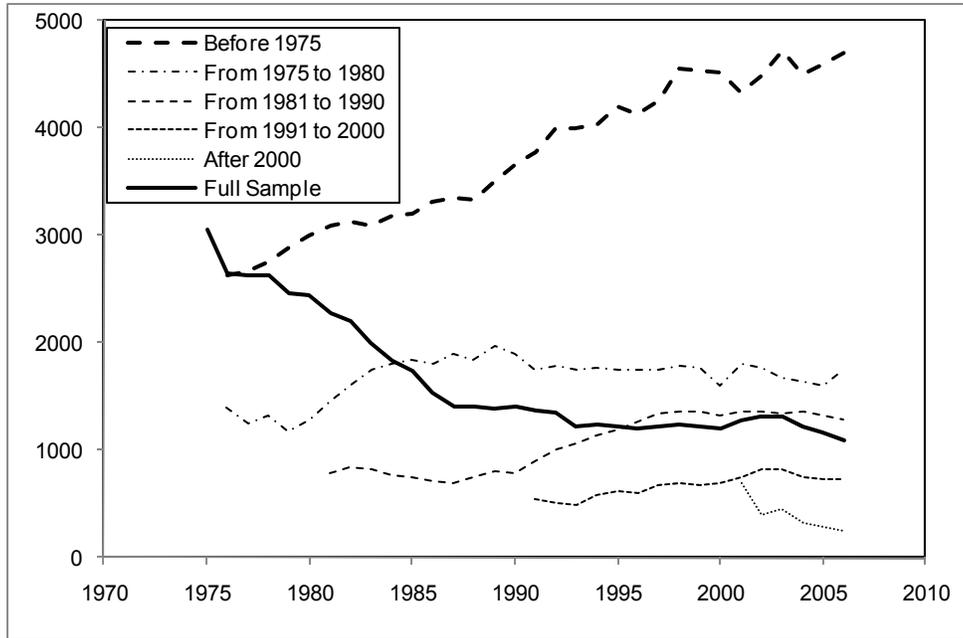
- King, M.R., Segal, D., 2009. The long-term effects of cross-listing, investor recognition, and ownership structure on valuation. *The Review of Financial Studies* 22, 2393-2421.
- Kumar, A., Lee, C.M.C., 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451-2486.
- La Porta, R., Lakonishok, J., Shleifer, A., Vishny, R., 1997. Good news for value stocks: Further evidence on market efficiency. *Journal of Finance* 52, 859-874.
- Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49, 1541-1578.
- Lehavy, R., Sloan, R.G., 2008. Investor recognition and stock returns. *Review of Accounting Studies* 13, 327-361.
- Lou, D., 2008. Attracting investor attention through advertising. Working paper, London School of Economics and Political Science.
- Lyandres, E., Sun, L., Zhang, L., 2008. The new issues puzzle: Testing the investment-based explanation. *The Review of Financial Studies* 21, 2825-2855.
- McLean, R.D., Pontiff, J., Watanabe, A., 2009. Share issuance and cross-sectional returns: International evidence. *Journal of Financial Economics* 94, 1-17.
- Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42, 483-510.
- Meschke, F., 2004. CEO interviews on CNBC. Working paper, University of Minnesota.
- Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. *Journal of Finance* 32, 1151-1168.
- Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics* 78, 277-309.
- Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.
- O'Brien, P.C., Bhushan, R., 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28, 55-76.
- Pontiff, J., Woodgate, A., 2008. Share issuance and cross-sectional returns. *Journal of Finance* 63, 921-945.
- Ritter, J.R., 2003. Investment banking and securities issuance. In: Constantinides G.M., Harris M., Stulz R.M. (Ed.), *Handbook of Economics and Finance: Corporate Finance*, Vol. 1A. Elsevier Science B.V.

Sloan, R.G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289-315.

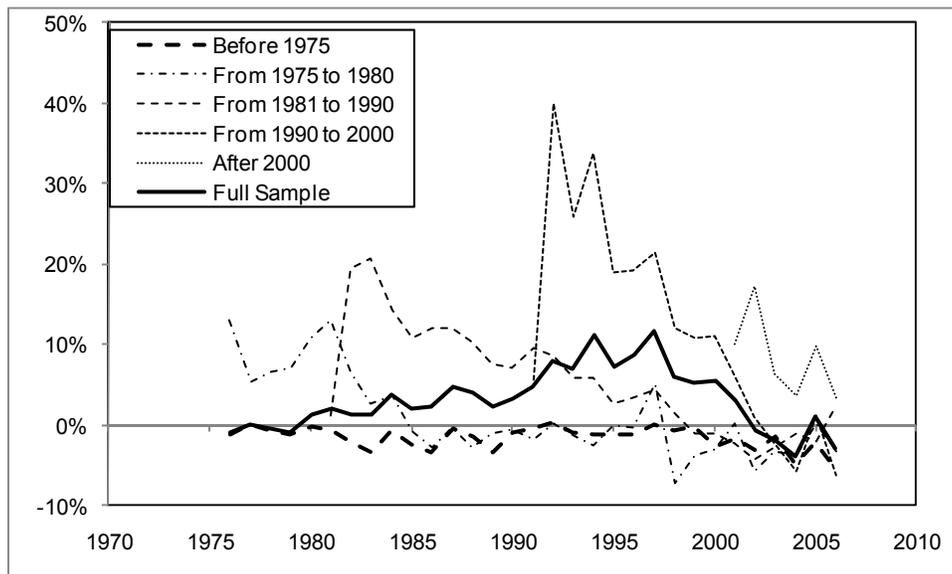
Titman, S., Wei, K.C.J., Feixue, X., 2004. Capital investments and stock returns. *Journal of Financial and Quantitative Analysis* 39, 677-700.

Womack, K.L., 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51, 137-167.

Yuan, Y., 2008. Attention and trading. Working paper, University of Iowa.

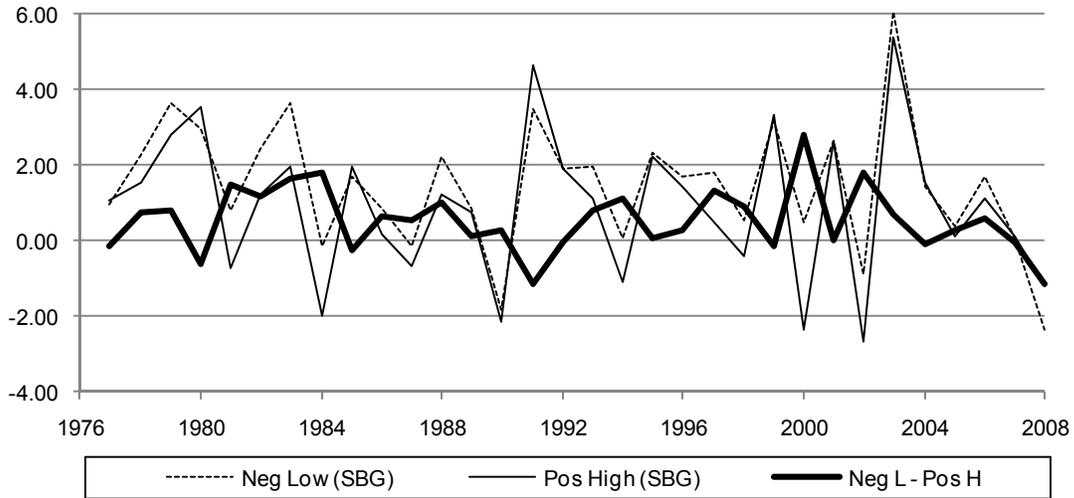


Panel A: Time series of cross-sectional median shareholder base

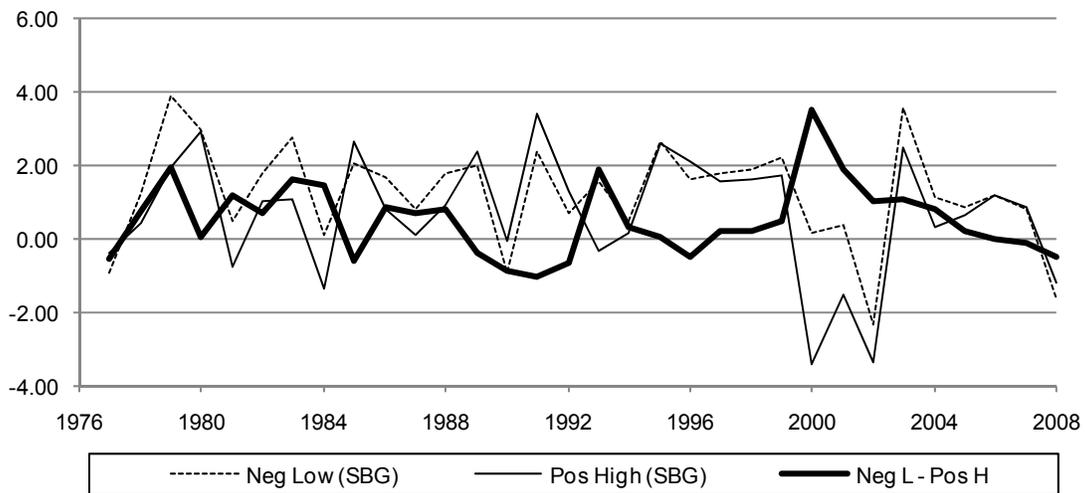


Panel B: Time series of cross-sectional mean shareholder base growth

Fig. 1. Time series of level and change in shareholder base. These two figures plot the time series of summary statistics for the shareholder base and annual growth rate from 1975 to 2006, both for the whole sample and by sub-period listing groups. Firms are divided into groups based on their years of listing: before 1975, 1975-1980, 1981-1990, 1991-1990, and 2000-2006, according to their time of first appearance in CRSP. Panel A represents the median number of shareholders and Panel B represents the mean annual log change in number of shareholders.

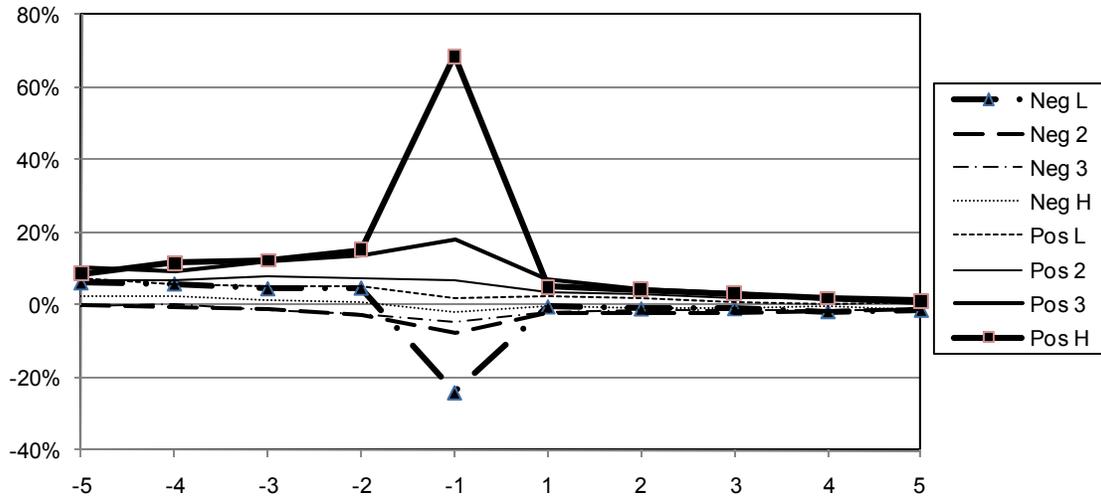


Panel A: Equal-weighted monthly return sorted by shareholder base growth

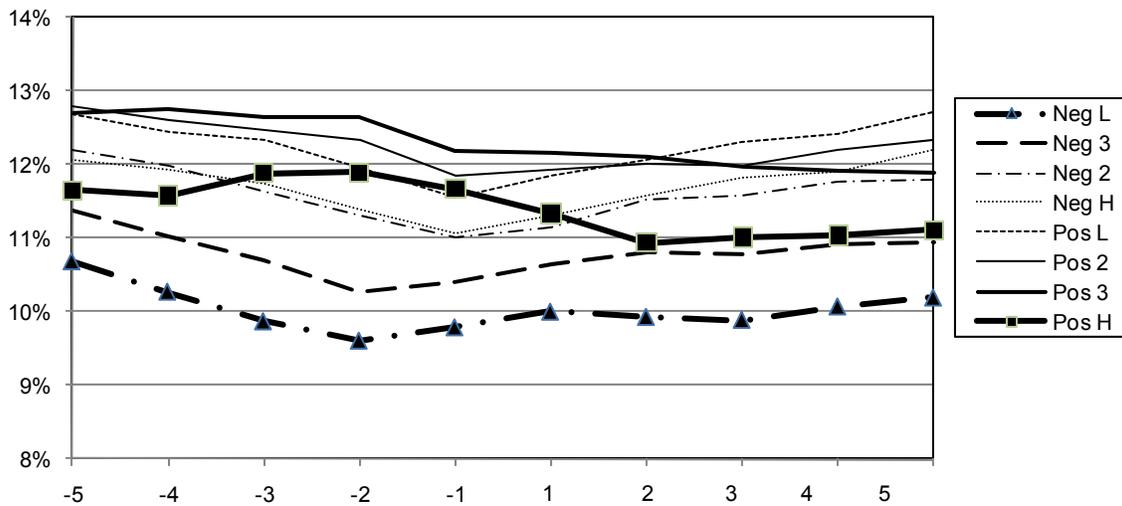


Panel B: Value-weighted monthly return sorted by shareholder base growth

Fig. 2. Time series of average monthly returns for shareholder base growth portfolios. At the end of June of each year t from 1977 to 2007, stocks are sorted into quartiles based on the negative and positive shareholder base growth at the fiscal year-end $t-1$, respectively. These two figures first calculate the time series of average monthly returns for the long/short leg and the hedge portfolio that goes *long* in stocks with the greatest decrease in the number of shareholders and goes *short* in stocks with the greatest increase in the number of shareholders, and then the average returns across each year are depicted. All portfolios are annually rebalanced and cover the period from July 1977 to June 2008. The vertical axis represents the monthly return in percentage terms and the horizontal axis denotes the time. The upper panel is for equal-weighted portfolios and the lower panel is for value-weighted portfolios.



Panel A. Annual median shareholder base growth in Year (-5,+5)



Panel B. Annual median operating margin in Year (-5,+5)

Fig. 3. Shareholder base growth and accounting performance around event time. At the end of June of each year t from 1977 to 2007, stocks are sorted into quartiles based on the negative and positive shareholder base growth at the fiscal year-end $t-1$, respectively. The time-series means of the cross-sectional average shareholder base growth and operating margins are reported every year for 10 years around the portfolio formation year (Year -1). Operating margin is defined as operating income before depreciation divided by contemporaneous net sales. Panel A plots the annual median shareholder base growth for portfolios 5 years before and after the portfolio formation. Panel B plots the time-series average of the annual median operating margin around the formation time.

Table 1. Descriptive statistics for shareholder base growth portfolios

At the end of June of each year t from 1977 to 2007, stocks with negative and positive shareholder base growth are sorted into quartiles, respectively. Shareholder base growth (SBG_{t-1}) is defined as the natural log of the ratio of the number of shareholders from the fiscal year ending in calendar year $t-1$ divided by the number of shareholders in $t-2$. The statistics in the first and second columns labeled N and $MktSh$ measure the average number of firms and the average percentage of the total market capitalization in each group. SBG_{t-2} is the log change in number of shareholders from fiscal year end $t-3$ to fiscal year end $t-2$. SB_{t-2} (in thousands) denotes the number of shareholders at fiscal year end $t-2$. $Size$, in millions of dollars, is price times shares outstanding at the end of June of year t . B/M is the ratio of the book value of equity at the fiscal year end in $t-1$ over the market value of equity at the end of December of $t-1$. The rest of the table reports firm characteristics measured at the end of fiscal year $t-1$, except for NA . NS_{t-1} represents the log change of split-adjusted shares outstanding from $t-2$ to $t-1$. ROE_{t-1} is income before extraordinary items in $t-1$ divided by average common equity over $t-2$ to $t-1$. $Advt_{t-1}$ is advertising expenses scaled by total sales in $t-1$. $CAPEX_{t-1}$ denotes capital expenditures in $t-1$ scaled by net property, plant and equipment from $t-2$. AGE is the number of years since the stock's first appearance in CRSP. IBG_{t-1} measures the change in the breadth of ownership among institutional investors at the fiscal year end $t-1$, as described in Chen, Hong and Stein (2002). IO_{t-1} is the fraction of shares outstanding held by institutions at the fiscal year end $t-1$. NA is the number of analysts making fiscal year- t earnings forecast at the end of year $t-1$. $IDIVOL_{t-1}$ is the annualized idiosyncratic volatility relative to the Fama-French (1993) model, computed from daily stock returns in each month and averaged over the fiscal year $t-1$. TO_{t-1} and $VOLUME_{t-1}$ are monthly stock turnover and trading volume (in millions), divided by two for Nasdaq stocks and averaged over the fiscal year $t-1$. The numbers in columns SBG_{t-1} , SBG_{t-2} , SB_{t-2} , NS_{t-1} , ROE_{t-1} , $Advt_{t-1}$, $CAPEX_{t-1}$, and AGE are time-series averages of yearly cross-sectional medians, while the numbers in columns $Size$, B/M , IBG_{t-1} , IO_{t-1} , NA , $IDIVOL_{t-1}$, TO_{t-1} , and $VOLUME_{t-1}$ are time-series averages of yearly cross-sectional means. Only firms with the number of shareholders greater than 100 and non-zero change in SBG_{t-1} are included.

		Fundamentals							Accounting variables					Market variables					
		N	MktSh	SBG _{t-1}	SBG _{t-2}	SB _{t-2}	Size	B/M	NS _{t-1}	ROE _{t-1}	Advt _{t-1}	CAPEX _{t-1}	AGE	IBG _{t-1}	IO _{t-1}	NA	IDIVOL _{t-1}	TO _{t-1}	VOLUME _{t-1}
Negative	Low	483	6.4%	-0.17	-0.03	1.92	708	0.95	0.005	0.083	0.0185	0.22	8.6	0.0001	0.33	5.6	0.45	0.09	3.4
	2	487	12.1%	-0.08	-0.05	2.02	1342	1.00	0.003	0.091	0.0178	0.19	12.3	0.0000	0.37	6.9	0.40	0.07	4.6
	3	491	18.3%	-0.05	-0.04	2.57	2025	0.98	0.003	0.099	0.0186	0.19	14.7	0.0000	0.36	8.0	0.39	0.06	5.1
	High	488	17.9%	-0.02	-0.02	2.53	1979	0.97	0.004	0.097	0.0199	0.20	12.5	0.0002	0.32	7.7	0.42	0.06	4.7
Positive	Low	334	13.5%	0.02	0.00	2.25	2293	0.85	0.008	0.106	0.0211	0.23	9.9	0.0004	0.34	7.9	0.43	0.08	6.1
	2	336	12.5%	0.07	0.02	1.98	2058	0.78	0.010	0.109	0.0210	0.26	8.5	0.0006	0.35	7.5	0.43	0.09	6.8
	3	335	12.1%	0.17	0.05	1.74	1973	0.69	0.018	0.113	0.0209	0.31	6.3	0.0007	0.36	7.1	0.45	0.10	6.6
	High	333	7.2%	0.54	0.05	1.11	1201	0.64	0.036	0.104	0.0212	0.41	4.4	0.0008	0.32	5.6	0.49	0.13	5.4
Pos H-Neg L				0.71	0.07	-0.81	493	-0.31	0.031	0.020	0.0027	0.19	-4.18	0.0007	-0.01	0.05	0.04	0.04	1.9
t(Spread)				(14.64)	(7.38)	(-11.53)	(3.84)	(-7.31)	(7.48)	(3.59)	(4.36)	(11.46)	(-8.27)	(6.56)	(-1.30)	(0.27)	(4.35)	(6.16)	(3.33)

Table 2. Portfolios sorted by shareholder base growth

At the end of June of each year t from 1977 to 2007, we form equal-weighted and value-weighted portfolios into quartiles based on the negative and positive change in number of shareholders in fiscal year $t-1$, respectively. The portfolios are held from July in year t to June in year $t+1$ and rebalanced annually. Stock returns are calculated from July 1977 to June 2008. Stocks with the greatest decrease in number of shareholders are in portfolio “Negative Low 1” and stocks with the greatest increase in number of shareholders are in portfolio “Positive High 4”. The row “Pos H - Neg L” represents the return spread between the “Positive High 4” and “Negative Low 1”. The row “Pos H - Neg” represents the return spread between portfolio “Positive High 4” and all stocks with decreases in number of shareholders. Monthly average raw return and Jensen’s alphas with respect to the CAPM and Fama-French (1993) three-factor model are presented. Robust Newey-West (1997) t -statistics with 12-month lags are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

		<u>Equal-Weighted</u>			<u>Value-Weighted</u>		
		Mean	CAPM Alpha	FF3 Alpha	Mean	CAPM Alpha	FF3 Alpha
Negative	Low 1	1.54*** (5.19)	0.43** (2.10)	0.20** (2.23)	1.32*** (5.38)	0.18** (2.04)	0.09 (1.03)
	2	1.56*** (5.90)	0.50*** (2.75)	0.22*** (2.83)	1.18*** (5.40)	0.13* (1.74)	0.04 (0.42)
	3	1.44*** (5.94)	0.43** (2.45)	0.12* (1.84)	1.12*** (5.77)	0.14* (1.94)	0.02 (0.26)
	High 4	1.53*** (6.09)	0.50*** (2.78)	0.25*** (3.17)	1.08*** (5.35)	0.12 (1.07)	0.00 (0.04)
Positive	Low 1	1.40*** (5.18)	0.34* (1.70)	0.08 (0.78)	1.08*** (5.07)	0.08 (0.71)	0.01 (0.17)
	2	1.33*** (4.83)	0.22 (1.12)	0.01 (0.13)	1.05*** (4.35)	-0.01 (-0.13)	0.02 (0.22)
	3	1.26*** (4.23)	0.10 (0.54)	-0.06 (-0.56)	1.07*** (4.04)	-0.05 (-0.58)	0.07 (0.76)
	High 4	0.95*** (2.78)	-0.27 (-1.29)	-0.33*** (-2.65)	0.75** (2.33)	-0.47*** (-3.13)	-0.25** (-2.50)
Pos H - Neg L		-0.59*** (-4.61)	-0.70*** (-5.59)	-0.52*** (-4.65)	-0.57*** (-3.09)	-0.65*** (-3.39)	-0.34*** (-2.71)
Pos H - Neg		-0.56*** (-3.98)	-0.74*** (-5.58)	-0.52*** (-5.26)	-0.38** (-2.00)	-0.59*** (-3.06)	-0.27*** (-2.59)

Table 3. FF-3 alphas of portfolios sorted on shareholder base growth

This table shows the Fama-French (1993) three-factor adjusted returns for portfolios sorted within various subsamples. All portfolios are rebalanced annually. In rows labeled “Subperiod”, portfolio returns are calculated for two periods separately: 1977-1992 and 1993-2008. In the row “NASDAQ Excluded”, we sort stocks based on the change in number of shareholders in $t-1$, using only NYSE/Amex stocks. In rows labeled “Size Group”, at the end of June of each year t , we assign stocks to three groups based on the 30th and 70th percentiles of NYSE market capitalization. Small stocks are below the 30th percentile. Medium stocks are between the 30th and 70th percentile, and Large stocks are above the 70th percentile. Then, within each size group, we sort stocks based on the change in number of shareholders. In the last two rows, alphas are calculated by exclusion of January returns and stocks with average prices less than \$5 in June of each year t . Portfolio “Negative Low 1” refers to the portfolio of stocks with the greatest decrease in number of shareholders in $t-1$ and portfolio “Positive High 4” refers to the portfolio of stocks with the greatest increase in number of shareholders in $t-1$. The column “Pos H - Neg L” represents the return spread between portfolio “Positive High 4” and “Negative Low 1”. Panel A presents results for equal-weighted portfolios and Panel B for value-weighted portfolios. Robust Newey-West (1997) t -statistics for the “Pos H - Neg L” portfolios are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

		Panel A: Equal-Weighted									
		Ranking on Shareholder Base Growth								Pos H -	t
		Negative				Positive				Neg L	(Spread)
		1(Low)	2	3	4(High)	1(Low)	2	3	4(High)		
Subperiod	1977-1992	0.06	0.15*	0.09	0.09	-0.09	-0.06	-0.22**	-0.41***	-0.47***	(-3.37)
	1993-2008	0.33**	0.31**	0.18*	0.41***	0.24	0.07	0.09	-0.26	-0.59***	(-3.28)
NASDAQ Excluded		0.05	0.03	-0.06	0.04	-0.15*	-0.06	-0.18**	-0.45***	-0.49***	(-4.50)
	Small	0.25**	0.29**	0.27***	0.38***	0.15	0.09	-0.06	-0.25*	-0.50***	(-4.13)
Size Group	Medium	0.01	0.13	-0.06	0.03	-0.03	-0.09	-0.08	-0.33**	-0.34**	(-2.20)
	Large	0.06	0.04	-0.01	0.04	-0.03	-0.05	-0.02	-0.26**	-0.32***	(-2.64)
January Excluded		0.03	0.12	0.00	0.07	-0.12	-0.19*	-0.26***	-0.63***	-0.65***	(-5.43)
Price>\$5		0.09	0.11	0.02	0.06	-0.05	-0.05	-0.17*	-0.35***	-0.44***	(-4.16)
		Panel B: Value-Weighted									
		Ranking on Shareholder Base Growth								Pos H -	t
		Negative				Positive				Neg L	(Spread)
		1(Low)	2	3	4(High)	1(Low)	2	3	4(High)		
Subperiod	1977-1992	0.11	0.12	0.08	0.00	0.07	0.09	-0.03	-0.20*	-0.30*	(-1.73)
	1993-2008	0.07	-0.02	-0.02	0.04	-0.05	-0.07	0.13	-0.40**	-0.47**	(-2.33)
NASDAQ Excluded		0.02	0.01	-0.03	0.01	0.01	0.09	-0.01	-0.34***	-0.36***	(-3.29)
	Small	-0.01	0.06	0.03	0.07	-0.17	-0.21*	-0.33**	-0.39***	-0.38***	(-3.20)
Size Group	Medium	0.08	0.17*	-0.02	0.00	-0.06	-0.04	-0.09	-0.33**	-0.41***	(-2.78)
	Large	0.05	0.06	-0.01	0.00	0.06	0.01	0.06	-0.15	-0.20	(-1.57)
January Excluded		0.09	0.07	0.02	0.03	0.00	-0.05	-0.04	-0.40***	-0.49***	(-3.77)
Price>\$5		0.09	0.02	0.01	0.03	0.02	0.01	0.08	-0.27***	-0.36***	(-3.03)

Table 4. Shareholder base growth portfolio interacted with firm characteristics

This table examines the profitability of a trading strategy based on the shareholder base growth in subsamples of firms sorted by various firm characteristics. A dependent-sorting scheme is adopted by first allocating stocks into three terciles ranked on specific firm characteristics, and then within each tercile, we sort stocks into quartiles based on negative and positive *SBG* of year *t-1* separately. Average monthly return on a hedge portfolio, which denotes the spread between the positive high *SBG* and negative low *SBG* stocks, is calculated in each characteristic subgroup, as is the difference of returns between the high and low characteristic groups. In Panel A, idiosyncratic volatility (*IDIVOL*) is computed from daily stock returns relative to the Fama-French (1993) model in the previous month. In Panel B, *NA* denotes the number of analysts making fiscal year-*t* earnings forecast at the end of year *t-1*. In Panel C, *IO* represents institutional ownership at the end of the previous quarter. Panel C and Panel D compute the return spreads conditional on the number of shareholders at fiscal year end *t-2* (*ShareBase*) and the number of years since the stock's first appearance in CRSP at fiscal year end *t-1* (*AGE*). Robust Newey-West (1997) *t*-statistics are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Equal-Weighted			Value-Weighted		
Mean	CAPM Alpha	FF3 Alpha	Mean	CAPM Alpha	FF3 Alpha
Panel A: First sorted by IDIVOL, then by SBG					
Low					
-0.23**	-0.28***	-0.21**	-0.24	-0.27*	-0.06
(-2.45)	(-3.06)	(-2.42)	(-1.55)	(-1.80)	(-0.41)
Medium					
-0.50***	-0.60***	-0.42***	-0.57***	-0.66***	-0.35*
(-4.06)	(-4.97)	(-3.87)	(-2.77)	(-3.18)	(-1.82)
High					
-0.65***	-0.78***	-0.60***	-1.16***	-1.27***	-1.06***
(-4.12)	(-5.25)	(-3.88)	(-3.50)	(-3.79)	(-3.43)
High-Low					
-0.42***	-0.50***	-0.39**	-0.92***	-0.99***	-1.00***
(-2.63)	(-3.25)	(-2.45)	(-2.70)	(-2.94)	(-3.01)
Panel B: First sorted by NA, then by SBG					
Small					
-0.62***	-0.72***	-0.62***	-0.97***	-1.08***	-1.02***
(-4.13)	(-4.80)	(-4.40)	(-4.10)	(-4.78)	(-4.03)
Medium					
-0.41**	-0.47**	-0.22	-0.49*	-0.58**	-0.27
(-2.31)	(-2.56)	(-1.39)	(-1.93)	(-2.26)	(-1.43)
Large					
-0.45***	-0.55***	-0.35***	-0.39**	-0.45**	-0.17
(-3.16)	(-3.67)	(-2.81)	(-2.15)	(-2.33)	(-1.20)
Large-Small					
0.17	0.17	0.27*	0.58**	0.63**	0.85***
(1.07)	(1.04)	(1.69)	(2.01)	(2.13)	(2.84)

Equal-Weighted			Value-Weighted		
Mean	CAPM Alpha	FF3 Alpha	Mean	CAPM Alpha	FF3 Alpha
Panel C: First sorted by IO, then by SBG					
Low					
-0.98*** (-5.05)	-1.12*** (-6.15)	-0.95*** (-4.35)	-1.21*** (-3.13)	-1.44*** (-3.67)	-1.14*** (-3.25)
Medium					
-0.50*** (-3.22)	-0.67*** (-4.52)	-0.42*** (-3.04)	-0.39 (-1.47)	-0.52* (-1.95)	-0.18 (-0.72)
High					
-0.26* (-1.96)	-0.36*** (-2.68)	-0.18 (-1.45)	-0.35** (-2.13)	-0.41** (-2.39)	-0.27 (-1.62)
High-Low					
0.71*** (3.25)	0.76*** (3.50)	0.77*** (3.23)	0.85** (2.21)	1.03** (2.52)	0.87** (2.32)
Panel D: First sorted by ShareBase, then by SBG					
Low					
-0.48*** (-3.06)	-0.62*** (-4.20)	-0.51*** (-3.29)	-0.82*** (-3.19)	-1.03*** (-4.15)	-0.74*** (-3.19)
Medium					
-0.44*** (-2.96)	-0.57*** (-4.15)	-0.40*** (-2.90)	-0.48** (-2.14)	-0.54** (-2.47)	-0.32* (-1.66)
High					
-0.65*** (-4.85)	-0.65*** (-4.71)	-0.45*** (-3.38)	-0.39** (-2.31)	-0.42** (-2.45)	-0.15 (-0.95)
High-Low					
-0.16 (-0.87)	-0.03 (-0.17)	0.06 (0.34)	0.43 (1.55)	0.61** (2.33)	0.59** (1.98)
Panel E: First sorted by AGE, then by SBG					
Young					
-0.38** (-2.48)	-0.53*** (-3.61)	-0.36** (-2.31)	-0.54** (-2.30)	-0.73*** (-3.19)	-0.37* (-1.87)
Medium					
-0.62*** (-4.63)	-0.69*** (-5.38)	-0.50*** (-3.91)	-0.52** (-2.45)	-0.56*** (-2.60)	-0.29 (-1.53)
Old					
-0.21* (-1.77)	-0.22* (-1.87)	-0.12 (-1.08)	-0.40** (-2.54)	-0.34** (-2.20)	-0.26* (-1.71)
Old-Young					
0.17 (0.92)	0.31* (1.73)	0.23 (1.21)	0.14 (0.49)	0.39 (1.47)	0.10 (0.40)

Table 5. Shareholder base growth on net equity issuance, SEO, repurchase, and stock M&A

From 1976 to 2006, Fama-Macbeth cross-sectional regressions are run each year with each firm's shareholder base growth (SBG_t) as the dependent variable. SBG_t is defined as the change in the natural log of number of shareholders from the fiscal year-end $t-1$ to t . NS_t represents the log change of split-adjusted shares outstanding from $t-1$ to t . SEO , REP , MAT and MAA are dummy variables for whether the firm experiences an SEO, a share repurchase, being a target or an acquirer in a stock merger and acquisition for the fiscal year $t-1$, t and $t+1$, as denoted in their subscripts. The time-series means of the estimated coefficients and intercept are reported and average R^2 is adjusted for degrees of freedom. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Model	Int	NS_t	SEO_{t-1}	SEO_t	SEO_{t+1}	REP_{t-1}	REP_t	REP_{t+1}	MAT_{t-1}	MAT_t	MAT_{t+1}	MAA_{t-1}	MAA_t	MAA_{t+1}	R^2
1	0.019*** (3.08)	0.389*** (7.60)													4.49%
2	0.024*** (3.98)		0.090*** (2.97)	0.118*** (6.00)	0.029** (2.27)										1.38%
3	0.036*** (4.88)					-0.019** (-2.36)	-0.014** (-2.18)	0.020** (2.53)							-0.02%
4	0.035*** (5.17)								0.064 (1.45)	-0.152*** (-4.33)	-0.031* (-1.88)				0.43%
5	0.033*** (4.83)											0.055 (1.00)	0.078** (2.23)	0.049 (1.63)	0.07%
6	0.030*** (4.75)			0.128*** (7.19)			-0.014** (-2.17)			-0.149*** (-4.22)			0.059* (1.77)		1.43%
7	0.027*** (4.26)		0.090*** (2.92)	0.115*** (5.77)	0.027** (2.12)	-0.016** (-2.05)	-0.014* (-1.97)	0.016** (2.05)	0.063 (1.44)	-0.151*** (-4.30)	-0.029* (-1.74)	0.026 (0.46)	0.056 (1.66)	0.04 (1.28)	1.86%
8	0.014** (2.55)	0.375*** (6.32)	0.090*** (3.12)	0.056** (2.21)	0.022* (1.75)	-0.006 (-0.79)	-0.004 (-0.62)	0.018** (2.30)	0.046 (1.03)	-0.153*** (-4.53)	-0.02 (-1.33)	0.019 (0.33)	0.04 (1.18)	0.036 (1.12)	5.60%

Table 6. Stock returns on shareholder base growth and net equity issuance

In Panel A, at the end of June of each year t from 1977 to 2007, we first form equal-weighted quartile portfolios by sorting stocks based on NS , and calculate the monthly average raw and risk-adjusted returns for each quartile as well as the spread between the highest and lowest NS quartile portfolios. Next, we control for the SBG effect, by first sorting stocks into quartiles based on SBG , and then within each quartile, sorting stocks into quartiles based on NS . The four NS portfolios are then averaged over each of the four SBG portfolios, and equal-weighted portfolio returns are reported. The column $\Delta\alpha$ denotes the difference of spread portfolio returns between the one-way and two-way sorts, and the column $|\Delta\alpha|/|\alpha|$ denotes the difference as a percentage of the spread returns on the one-way sorts. For stock returns from July of year t to June of $t+1$, SBG and NS are the log change in the number of shareholders and shares outstanding from fiscal year-end $t-2$ to $t-1$.

Panel B shows monthly Fama-Macbeth cross-sectional regressions of stock returns on SBG and NS for three different samples. Model 1 and 2 includes all firms; Model 3 excludes stocks that experience an SEO, a repurchase or a stock M&A during the 12 months before the monthly returns are measured; and Model 4 includes only stocks with zero NS . For stock returns from July of year t to June of $t+1$, $Size$ is the logarithm of market capitalization in June of t (in millions); B/M is the natural log of ratio of book equity at the fiscal year-end $t-1$ over market equity in December of $t-1$. To predict returns in month j , MOM is the buy-and-hold return from month $j-12$ to $j-2$. The time-series means of the estimated coefficients and intercept are reported, and average R^2 is adjusted for degrees of freedom. The t -statistics, shown in brackets, are calculated based on time-series standard errors of the monthly slopes and corrected for Newey-West (1987) autocorrelation with 12-month lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Portfolios Sorted by NS

1. One-way Sorts							
	1(Low)	2	3	4(High)	4 - 1		
Mean	1.65*** (6.74)	1.57*** (5.93)	1.50*** (5.23)	0.99*** (2.93)	-0.66*** (-3.51)		
CAPM Alpha	0.65*** (3.35)	0.55*** (2.74)	0.37* (1.92)	-0.21 (-1.00)	-0.86*** (-4.95)		
FF3 Alpha	0.35*** (3.68)	0.23*** (2.67)	0.20** (2.21)	-0.32*** (-2.95)	-0.66*** (-5.03)		
2. Controlling for SBG							
	1(Low)	2	3	4(High)	4 - 1	$\Delta\alpha$	$\Delta\alpha / \alpha$
Mean	1.61*** (6.58)	1.54*** (5.62)	1.45*** (5.14)	1.08*** (3.32)	-0.54*** (-3.35)	-0.12*** (-3.04)	18%
CAPM Alpha	0.61*** (3.16)	0.50** (2.48)	0.32* (1.75)	-0.10 (-0.48)	-0.71*** (-4.61)	-0.15*** (-4.47)	17%
FF3 Alpha	0.31*** (3.35)	0.20** (2.30)	0.14* (1.73)	-0.22** (-2.23)	-0.53*** (-4.47)	-0.13*** (-4.51)	20%

Panel B: Fama-Macbeth Regressions

Model	Int	SBG	NS	Size	B/M	MOM	R²	obs
1	1.684*** (3.53)	-0.392*** (-3.05)		-0.143** (-2.46)	0.278*** (2.69)	0.473*** (3.49)	2.41%	3503
2	1.779*** (3.70)	-0.213* (-1.81)	-1.865*** (-5.74)	-0.149*** (-2.60)	0.250*** (2.73)	0.455*** (3.28)	2.56%	3503
3	1.801*** (3.72)	-0.264** (-2.17)	-1.844*** (-5.24)	-0.155*** (-2.67)	0.253*** (2.77)	0.438*** (3.08)	2.47%	3134
4	1.869*** (4.01)	-0.484* (-1.72)		-0.226*** (-3.36)	0.348*** (3.33)	0.196 (1.30)	1.81%	428

Table 7. Fama-Macbeth cross-sectional regressions

This table shows average slopes and their t -statistics from monthly cross-sectional return regressions for both the whole sample and sub-size groups. Size groups are defined by assigning firms into one of three groups (small, medium, and large) using the 30th and 70th percentiles of NYSE market capitalization breakpoints in June of year t . SBG , $Size$, B/M , and MOM are defined the same as in Table 6. For stock returns from July of year t to June of $t+1$, $PROFIT$ denotes the income before extraordinary items minus dividends on preferred stock plus income statement deferred taxes, divided by book equity at $t-1$; $ACCR$ is the change in operating working capital from $t-2$ to $t-1$ divided by book equity for $t-1$, where operating working capital is current assets minus cash and short-term investments minus current liabilities plus debt in current liabilities; $CAPEX$ represents capital expenditures in $t-1$ scaled by net property, plant and equipment from $t-2$. To predict returns in month j , TO and $VOLUME$ are the logarithm of stock turnover and trading volume (in millions) in month $j-1$, divided by two for Nasdaq stocks; $IDIVOL$ is the annualized idiosyncratic volatility relative to the Fama-French (1993) model, computed from daily stock returns in month $j-1$ and in percentage terms. IBG is the change in the breadth of ownership among institutional investors, and IO is the fraction of shares outstanding held by institutions, both measured at the end of last quarter. NA is the number of analysts making fiscal year- t earnings forecast at the end of year $t-1$. PIN denotes the probability of informed trading at the end of calendar year $t-1$ from Easley, Hvidkjaer, and O'Hara (2008). All accounting variables are winsorized at the 1% and 99% percentiles. Only firms with the number of shareholders greater than 100 are included, with the monthly average number of stocks in the regression shown in the last column. The sample period is from July 1977 to June 2008 for all controls except IBG and IO (February 1980 to June 2008) and PIN (January 1984 to December 2002). The time-series means of the estimated coefficients and intercept are reported, and average R^2 is adjusted for degrees of freedom. The t -statistics, shown in brackets, are calculated based on time-series standard errors of the monthly slopes and corrected for Newey-West (1987) autocorrelation with 12-month lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. All Firms

Model	Int	SBG	Size	B/M	MOM	IBG	NA	IO	TO	IDIVOL	VOLUME	PROFIT	ACCR	CAPEX	PIN	R ²	obs
1	1.684*** (3.53)	-0.392*** (-3.05)	-0.143** (-2.46)	0.278*** (2.69)	0.473*** (3.49)	2.41%	3525
2	1.642*** (3.45)	-0.411*** (-3.23)	-0.141** (-2.45)	0.283*** (2.81)	0.467*** (3.54)	11.291 (1.59)	2.53%	3525
3	2.055*** (3.93)	-0.400*** (-3.16)	-0.257*** (-3.45)	0.253** (2.44)	0.502*** (3.77)	.	0.041*** (4.39)	2.61%	3525
4	1.681*** (3.59)	-0.377*** (-2.96)	-0.170*** (-3.08)	0.272*** (2.70)	0.477*** (3.52)	.	.	0.437** (2.56)	2.59%	3525
5	1.724*** (3.57)	-0.411*** (-3.19)	-0.141** (-2.31)	0.273*** (2.72)	0.532*** (3.73)	.	.	.	-0.918 (-1.20)	3.06%	3317
6	1.696*** (4.28)	-0.378*** (-3.13)	-0.136*** (-2.86)	0.261*** (2.73)	0.491*** (4.06)	-0.322 (-1.29)	3.04%	3524
7	2.287*** (2.82)	-0.453*** (-3.93)	-0.244** (-2.09)	0.309*** (3.40)	0.504*** (3.73)	0.140* (1.66)	3.48%	3313
8	1.666*** (3.85)	-0.402*** (-3.12)	-0.141*** (-2.95)	0.282*** (3.18)	0.450*** (3.39)	0.069 (0.39)	.	.	.	2.71%	3521
9	1.713*** (3.55)	-0.356*** (-2.79)	-0.144** (-2.45)	0.299*** (2.94)	0.437*** (3.28)	-0.459*** (-2.91)	.	.	2.39%	3134
10	1.829*** (3.91)	-0.261** (-2.19)	-0.152*** (-2.64)	0.254** (2.59)	0.447*** (3.35)	-0.300*** (-3.68)	.	2.50%	3283
11	0.108 (0.19)	-0.369*** (-2.64)	0.038 (0.63)	0.205** (2.47)	0.776*** (4.17)	1.182* (1.80)	3.06%	1630

Panel B. Small Size Firms

Model	Int	SBG	Size	B/M	MOM	IBG	NA	IO	TO	IDIVOL	VOLUME	PROFIT	ACCR	CAPEX	PIN	R ²	obs
1	2.309*** (4.36)	-0.351** (-2.53)	-0.335*** (-4.02)	0.290*** (2.61)	0.428*** (3.56)	1.79%	2277
2	2.422*** (4.60)	-0.370*** (-2.66)	-0.358*** (-4.27)	0.283*** (2.63)	0.427*** (3.65)	9.484 (0.42)	1.83%	2277
3	2.657*** (4.54)	-0.390*** (-2.87)	-0.436*** (-4.5)	0.255** (2.27)	0.463*** (3.97)	.	0.065*** (3.03)	1.87%	2277
4	2.336*** (4.54)	-0.335** (-2.43)	-0.393*** (-4.96)	0.266** (2.49)	0.434*** (3.62)	.	.	0.840*** (4.09)	1.88%	2277
5	2.344*** (4.33)	-0.434*** (-2.95)	-0.337*** (-3.84)	0.271** (2.41)	0.482*** (3.73)	.	.	.	-0.153 (-0.12)	2.26%	2101
6	2.350*** (5.59)	-0.331** (-2.52)	-0.327*** (-4.94)	0.270** (2.58)	0.453*** (4.27)	-0.364 (-1.53)	2.43%	2276
7	2.963*** (3.43)	-0.480*** (-3.76)	-0.434*** (-3.23)	0.322*** (3.29)	0.449*** (3.73)	0.168* (1.82)	2.74%	2098
8	2.275*** (4.85)	-0.361*** (-2.59)	-0.330*** (-4.81)	0.300*** (3.30)	0.405*** (3.41)	0.037 (0.21)	.	.	.	2.07%	2273
9	2.328*** (4.41)	-0.317** (-2.35)	-0.333*** (-4.04)	0.311*** (2.80)	0.408*** (3.39)	-0.446*** (-2.62)	.	.	1.83%	2096
10	2.431*** (4.62)	-0.209* (-1.68)	-0.340*** (-4.03)	0.272** (2.47)	0.410*** (3.46)	-0.266*** (-3.58)	.	1.87%	2172
11	0.847 (0.96)	-0.366* (-1.88)	-0.168 (-1.19)	0.247** (2.09)	0.865*** (4.48)	1.375 (1.46)	2.33%	716

Panel C. Medium Size Firms

Model	Int	SBG	Size	B/M	MOM	IBG	NA	IO	TO	IDIVOL	VOLUME	PROFIT	ACCR	CAPEX	PIN	R ²	obs
1	0.64 (1.34)	-0.355** (-2.25)	0.004 (0.06)	0.184* (1.74)	0.835*** (3.58)	3.78%	796
2	0.737 (1.51)	-0.356** (-2.25)	-0.012 (-0.21)	0.174* (1.69)	0.843*** (3.73)	2.389 (0.24)	4.02%	796
3	0.974* (1.79)	-0.352** (-2.25)	-0.065 (-0.8)	0.162 (1.56)	0.848*** (3.72)	.	0.018 (1.52)	4.22%	796
4	0.687 (1.42)	-0.334** (-2.13)	-0.018 (-0.29)	0.181* (1.72)	0.832*** (3.56)	.	.	0.230 (1.11)	4.37%	796
5	0.707 (1.39)	-0.278* (-1.80)	0.002 (0.03)	0.210** (2.09)	0.873*** (3.76)	.	.	.	-1.093 (-1.22)	4.94%	765
6	1.384*** (3.29)	-0.313** (-2.00)	-0.054 (-1.08)	0.11 (1.20)	0.825*** (3.97)	-1.606*** (-3.85)	5.02%	796
7	0.623 (0.81)	-0.279* (-1.82)	0.007 (0.07)	0.223** (2.18)	0.876*** (3.91)	-0.013 (-0.20)	5.23%	765
8	0.708 (1.52)	-0.406** (-2.52)	-0.022 (-0.39)	0.200* (1.92)	0.814*** (3.58)	0.805*** (2.77)	.	.	.	4.33%	795
9	0.533 (1.05)	-0.297* (-1.79)	0.021 (0.34)	0.204** (2.08)	0.756*** (3.24)	-0.109 (-0.65)	.	.	3.82%	666
10	0.675 (1.42)	-0.304** (-2.14)	0.007 (0.12)	0.166** (1.97)	0.773*** (3.30)	-0.198 (-1.31)	.	4.28%	706
11	-0.909 (-1.36)	-0.352** (-2.09)	0.217*** (2.64)	0.185** (2.14)	0.735*** (3.20)	0.124 (0.14)	3.14%	518

Panel D. Large Size Firms

Model	Int	SBG	Size	B/M	MOM	IBG	NA	IO	TO	IDIVOL	VOLUME	PROFIT	ACCR	CAPEX	PIN	R ²	obs
1	1.009* (1.76)	-0.646*** (-3.18)	-0.05 (-0.85)	0.074 (0.78)	0.675** (2.58)	6.28%	452
2	1.121** (2.00)	-0.654*** (-3.26)	-0.066 (-1.12)	0.066 (0.71)	0.651** (2.50)	3.303 (0.54)	6.63%	452
3	1.268** (2.17)	-0.626*** (-3.19)	-0.097 (-1.34)	0.067 (0.71)	0.679*** (2.60)	.	0.011 (1.55)	6.84%	452
4	0.826 (1.52)	-0.598*** (-2.90)	-0.054 (-0.91)	0.082 (0.87)	0.705*** (2.70)	.	.	0.346 (1.43)	7.11%	452
5	0.937* (1.69)	-0.687*** (-3.22)	-0.045 (-0.78)	0.069 (0.79)	0.708*** (2.70)	.	.	.	0.451 (0.41)	8.15%	450
6	1.188** (1.99)	-0.610*** (-3.13)	-0.064 (-1.06)	0.04 (0.45)	0.730*** (3.05)	-0.525 (-0.81)	7.93%	452
7	1.055 (1.44)	-0.669*** (-3.18)	-0.059 (-0.64)	0.08 (0.89)	0.702*** (2.79)	0.013 -0.22	7.39%	450
8	1.032* (1.80)	-0.644*** (-3.15)	-0.061 (-1.04)	0.096 (0.98)	0.630** (2.49)	0.498 (1.29)	.	.	.	6.91%	452
9	0.942 (1.53)	-0.677*** (-3.10)	-0.041 (-0.63)	0.056 (0.61)	0.624** (2.50)	-0.182 (-0.82)	.	.	6.64%	371
10	1.029* (1.71)	-0.594*** (-2.75)	-0.051 (-0.82)	0.049 (0.60)	0.611** (2.52)	-0.048 (-0.22)	.	7.31%	405
11	0.411 (0.48)	-0.369* (-1.68)	0.026 (0.30)	0.01 (0.10)	0.612* (1.87)	-0.633 (-0.60)	5.53%	396