European Accounting Review, 2023
https://doi.org/10.1080/09638180.2023.2265975

Check for updates

Investor Base Size and Underreaction-Consistent Stock Return Anomalies

STEFAN ANCHEV* and NICHA LAPANAN**

*BI Norwegian Business School, Norway **University of Agder, Norway*

(*Received: 28 December 2021; accepted: 14 September 2023*)

ABSTRACT We find that several well-documented underreaction-consistent stock return anomalies, such as those based on stocks' earnings-to-price ratios, returns on assets and past returns, arise and persist only among stocks with smaller (institutional) investor bases, which are presumably stocks that are neglected by investors. These results are driven by the short side of our long-short trading strategies (i.e., by the seemingly overpriced stocks from the bottom quantiles of the anomaly variables), they appear even after controlling for several stock characteristics (e.g., market capitalization and institutional ownership) and potential risk factors, and they are considerably more pronounced during periods with more information and/or less technology. Overall, these findings suggest that the incomplete dissemination of (negative) information across investors helps in explaining the occurrence and the persistence of the anomalies.

Keywords: (Institutional) Investor base size; Underreaction-consistent stock return anomalies; Information dissemination

JEL classifications: G12; G14; G23

1. Introduction

In the highly influential Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966), the investor bases of all stocks would consist of all of the investors in the market. In contrast, even today, the mean institutional investor base of the stocks listed on the three major stock exchanges in the United States (US) consists of just around 5% of all of the identifiable institutional investors in the market. Given this seemingly considerable discrepancy between the CAPM and the data, in this paper, we examine empirically whether the size of stocks' investor bases helps in explaining several stock return anomalies.

The size of the investor base of a certain stock could be relevant in this respect because it is presumably related to the degree to which investors know about and follow that stock. Indeed, this seems logical because, even if investors do not regularly follow a given stock, they must at least know about that stock before they can take their positions in it. More importantly, in the CAPM, the result above would arise largely because of the assumption that, without incurring any costs, all investors know about and follow all of the available securities. This implies that

**Corresponding Author*: Stefan Anchev, Department of Accounting and Operations Management, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. E-mail: stefan.anchev@bi.no.

Paper accepted by Beatriz García Osma.

all of the available information about those securities is instantaneously disseminated across all investors, which means that they take their positions while having complete information. Under these and several other conditions (see, e.g., Fama & French, 2004; Perold, 2004), the CAPM suggests that stock returns should be predictable only with their market betas.

In contrast to this prediction, during the last 50–60 years, approximately 450 stock characteristics besides market beta have been found that seem to predict stock returns (see, e.g., Hou et al., 2020). Despite the abundance of these so-called anomalies and their widespread use by investment professionals, however, the reasons for them are still far from being clear. One such reason could be that the assumption in the CAPM that all investors know about and follow all of the available securities is significantly violated in reality.[1] This could be so if investors face costs of obtaining information (Merton, 1987) and/or if they have limited attention (Peng & Xiong, 2006). Regardless of the exact constraint(s) on investors' behavior, unlike in the CAPM, all of the available information about all of the available securities would not be instantaneously disseminated across all investors, which means that they would have to take their positions while having incomplete information. Under these conditions, the prices of stocks can underreact to both positive and negative information for a prolonged period of time, which could be the reason why their returns seem to be predictable with other stock characteristics than market beta. Thus, to the extent that the size of the investor base of a certain stock is related to the degree to which investors know about and follow that stock, underreaction-consistent stock return anomalies should arise and persist mostly among stocks with smaller investor bases.

To test this prediction, we use data on the stocks listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and NASDAQ from the end of the first quarter of 1980 to the end of the first quarter of 2019. To our knowledge, data on the size of these stocks' complete investor bases are, at least to researchers, unavailable. Hence, as a proxy for the size of the complete investor base of a certain stock, we use the size of the institutional investor base of that stock, calculated as the number of institutional investors with long positions in it. With respect to the stock return anomalies, we examine one valuation anomaly based on the earnings-to-price ratios of stocks (as in Basu (1977)), one profitability anomaly based on the returns on assets of stocks (as in Haugen and Baker (1996)) and one momentum anomaly based on the past returns on stocks (as in Jegadeesh and Titman (1993)). These anomalies are examined primarily because they are some of the very few underreaction-consistent anomalies that have been repeatedly reproduced in the prior literature, even with very recent data (see, e.g., Hou et al., 2020; Linnainmaa & Roberts, 2018).

Given that the underlying mechanism that we seek to investigate is ultimately related to the degree of information dissemination across investors, several features of our analyzes are important to note. First, the size of the institutional investor base of a certain stock is highly correlated with several other characteristics of that stock, such as its stock exchange, market capitalization and institutional ownership, of which the last two have already been found to help

[1] Quite a few empirical observations seem to support this conjecture. First, most individual and institutional investors have been found to have stock portfolios with only a fraction of all of the available stocks (see, e.g., Goetzmann & Kumar, 2008; Griffin & Xu, 2009). Of course, these findings can be attributed to other factors as well, some of which are taxes and/or transactions costs. There is additional evidence, however, that, in contrast to the principles of portfolio optimization in the CAPM, most investors allocate larger stock portfolio weights to stocks with which they are more likely to be familiar, such as stocks that are more local to them (see, e.g., Coval & Moskowitz, 1999; Ivković & Weisbenner, 2005) or stocks that receive more attention in the financial media (see, e.g., Barber & Odean, 2008). Second, many firms do take costly actions with the goal of introducing their stocks to investors. For example, firms list their stocks on one or more stock exchanges and they pay others, such as investment banks, to promote them. Lastly, several instances have been documented in which stock prices seem to react to certain information more strongly when that information is more likely to be obtained by a larger number of investors (see, e.g., Huberman & Regev, 2001).

in explaining stock return anomalies in the context of other underlying mechanisms (see, e.g., Griffin & Lemmon, 2002; Nagel, 2005). Therefore, to control for the effects of these other stock characteristics, our analyzes exclude the stocks with the smallest market capitalizations and, instead of using the raw size of stocks' institutional investor bases, our main analyzes exploit the residuals from estimating cross-sectional regressions of this variable on the aforementioned other stock characteristics.[2] Second, our analyzes entail institutional investors, which, on average, are likely to have more resources (e.g., technology) and, hence, fewer difficulties in obtaining information than other investors (e.g., individuals). Finally, our analyzes involve stock return anomalies based on stocks' earnings and past returns, which are probably some of the most scrutinized pieces of information by the financial media and/or by investors in general. Thus, considering all of these aspects of our analyzes, it is worth noting that they are likely to be biased toward finding statistically and economically insignificant results.

Despite this bias, however, the results from our analyzes do support our prediction. In fact, they indicate that the stock return anomalies arise and persist only among stocks with smaller investor bases. Specifically, among the stocks with the smallest residual institutional investor bases, the raw quarterly return on a long-short trading strategy based on the earnings-to-price ratios of stocks is 2.23%. In contrast, among the stocks with the largest residual institutional investor bases, the return is not even positive; it is -0.11%. Consequently, the differential between these returns is 2.34%. The corresponding return differentials for the strategies based on stocks' returns on assets and past returns are 1.88% and 2.72%, respectively. These patterns keep appearing after controlling for various other stock characteristics and potential risk factors, they also appear over longer holding periods and they are highly statistically significant (even from the perspective of Harvey et al. (2016)). Hence, they are consistent with the hypothesis that the incomplete dissemination of information across investors helps in explaining the occurrence and the persistence of the anomalies.

In addition to these main findings, our analyzes indicate that the effect of the residual size of stocks' institutional investor bases on the stock return anomalies arises almost completely on the short side of the trading strategies (i.e., among the seemingly overpriced stocks in the bottom quantiles of the anomaly variables). This suggests that the prices of less-known and less-followed stocks (hereinafter, neglected stocks) underreact to negative information more than they underreact to positive information, which makes intuitive sense in a theory predicated on the incomplete dissemination of information across investors. That is, if managers, for whatever reason (e.g., their compensation and/or reputation), prefer the prices of their firms' stocks to be higher rather than lower, then it is reasonable to assume that their incentives for disseminating negative information would be weaker than their incentives for disseminating positive information (Kothari et al., 2009). In that case, they are likely to expend less of their resources (e.g., time and/or effort) into disseminating the former than into disseminating the latter. To do so, they could, for instance, reduce their voluntary disclosures and/or their presence in the financial media. Under these conditions, therefore, the dissemination of negative information across investors can be more incomplete than the dissemination of positive information.

Apart from these explanations for our results, one alternative explanation for them could be that stocks with smaller residual institutional investor bases are stocks with higher transaction costs, especially in terms of short selling. In the theory of Miller (1977), the prices of such stocks would underreact predominantly to negative information not because many investors do

[2]To be clear, we are interested in analyzing the effects of the size of the institutional investor bases of stocks, while holding all else equal, including their institutional ownership. For instance, consider two otherwise similar stocks, both with an institutional ownership of 45%, where one of them has 25 institutional investors while the other has 180 such investors. Our interest then is in exploring if the variation of this kind helps in explaining the stock return anomalies.

not know about them and do not follow them, but because many pessimistic investors with negative opinions find it too costly to take short positions in them. Transaction costs in general, however, and those of short selling in particular, have been found to be higher among stocks with smaller market capitalizations and/or lower institutional ownership (see, e.g., D'Avolio, 2002; Stoll & Whaley, 1983). Thus, it is important to reiterate that our results emerge after excluding the stocks with the smallest market capitalizations and after controlling for these stock characteristics. This is also the case even when controlling for the share turnover of stocks, which has conventionally been thought of as being closely related to transaction costs. Hence, although we cannot definitively rule out that this underlying mechanism, which has found empirical support before (see Nagel, 2005), contributes toward our results, it is unlikely that they are driven by it.

To explore our interpretation of the results further, we additionally seek to examine whether they vary with the amount of information in the market. If our results truly reflect the incomplete dissemination of information across investors, then it seems sensible to expect that they would be more pronounced during periods when there is more information to be disseminated. In line with this reasoning, we find a strong seasonality in the return differentials presented above in that they are almost fully realized during the first two calendar quarters; that is, when the trading strategies are implemented at the end of the first and the fourth calendar quarters, which is when nearly all of the public information for a given current or most recent calendar year (which is equivalent to the fiscal year of most firms in the US) is available. In light of these results, it is worth noting that the underlying mechanism of the kind discussed here could help in explaining the so-called turn-of-the-year (or January) effect (see, e.g., Keim, 1983; Reinganum, 1983).

Our analyzes also involve investigating if the results differ over time. Considering the various technological developments during our sample period (e.g., in terms of computing power and/or the internet), it is likely that the costs of acquiring information faced by investors have decreased and/or that the constraints on their attention have weakened over that period. If this is true, then investors could know about and follow an increasing number of stocks over time. Thus, it seems plausible that the dissemination of information across investors would have been more incomplete during the earlier part of our sample period than during its later part. In congruence with these arguments, we find that our results are considerably more pronounced over the first half of that period. It is important to note that, since the stock return anomalies that we study have indeed been reproduced with very recent data (see Linnainmaa & Roberts, 2018), these results are unlikely to reflect the finding in McLean and Pontiff (2016) that, on average, such anomalies become less pronounced after their original publication in the academic literature.

Last but not least, in a placebo test, we have examined the accrual anomaly of Sloan (1996) as well, which is one of the best-known accounting-based stock return anomalies and for which a commonly proposed explanation in the prior literature has been the overreaction of stock prices to new information. However, if this explanation is correct and if our results truly reflect the underreaction to new information of the prices of stocks, then the size of their investor bases should be irrelevant for this anomaly. As expected, this is exactly what we find, which serves both as an additional support for our story, but also as an important caveat to it.

Overall, our paper contributes to the empirical asset pricing literature on cross-sectional stock return anomalies.[3] Specifically, it provides an explanation for several well-documented

[3]See, for example, Hong et al. (2000), Griffin and Lemmon (2002), Nagel (2005), Fama and French (2008), Stambaugh et al. (2012), Hou et al. (2015), Edelen et al. (2016), Fama and French (2016), Harvey et al. (2016), Keloharju et al. (2016), McLean and Pontiff (2016), Novy-Marx and Velikov (2016), Cao et al. (2017), Stambaugh and Yuan (2017),

underreaction-consistent anomalies, which evolves around the size of stocks' investor bases. Indeed, while several papers find that this stock characteristic has implications for stock returns (see, e.g., Bodnaruk & Ostberg, 2009; Richardson et al., 2012), to our knowledge, no paper has, thus far, empirically examined its implications for stock return anomalies. Hence, for researchers who study such anomalies, especially those consistent with the underreaction of stock prices to new information, the implication of our findings is that they should consider the size of the investor bases of stocks in their analyzes.

Finally, it is crucial to emphasize that even though our findings suggest that several prominent stock return anomalies reflect a mispricing, this mispricing seems to be concentrated among stocks with smaller investor bases. Moreover, the mispricing appears to have diminished significantly during the last 20 years or so. Therefore, probably the most important implication of our paper is that stock markets seem to be even more informationally efficient than previously documented (see, e.g., Hou et al., 2020; Linnainmaa & Roberts, 2018).

2. Data

2.1. Institutional Investor Base Size

To our knowledge, data on the size of the complete investor bases of the stocks in the US are, at least to researchers, unavailable. As a matter of fact, the number of investors with long positions in a certain stock from Compustat (i.e., Compustat variable CSHR) is obtained from the 10-K form of that stock's issuing firm, in which only the number of investors who keep their shares in their own accounts, but not in the accounts of their nominees, is reported. Hence, as a proxy for the size of the complete investor bases of stocks, in our analyzes, we use the size of their institutional investor bases, for which more complete data are available in the Thomson Reuters Institutional Holdings (13F) Database.

These data, which have been widely used by researchers (see, e.g., Edelen et al., 2016; Lewellen, 2011; Nagel, 2005), are described in more detail in Gompers and Metrick (2001). Here, it suffices to say that the data are obtained from 13F forms, which have been required at the end of each calendar quarter by the US Securities and Exchange Commission (SEC) since the late 1970s from all (domestic and foreign) institutional investors with equity portfolios worth at least 100 million US dollars (USD). Institutional investors with such portfolios are required to report all of their long stock positions of at least 10,000 shares or USD 200,000.

Despite being highly detailed, the database from Thomson Reuters has several limitations. First, it does not contain institutional investors that do not meet the aforementioned requirements. However, even though these investors could be quite numerous, many of them are likely to be considerably less wealthy than those that do meet the requirements. Thus, while this limitation prevents us from identifying all of the institutional investors in a given stock, it prompts us to focus on the most economically relevant institutional investors in that stock. Second, the database does not contain institutional investors that know about and follow a certain stock, but that have neither long nor short positions in that stock. This limitation, however, which admittedly could lead to a large error in the identification of such investors, is difficult to overcome. Finally, the database does not contain institutional investors that have only short positions in a stock. Nevertheless, considering the many potential impediments to taking such positions and their relatively low number and value in the US (D'Avolio, 2002; Dechow et al., 2001; Duffie

---

Birru (2018), Engelberg et al. (2018), Linnainmaa and Roberts (2018), Calluzzo et al. (2019), Guo et al. (2020), Hou et al. (2020) and Chen et al. (2023).

et al., 2002; Koski & Pontiff, 1999),[4] the error in the identification of the size of the institutional investor bases of stocks due to this limitation is likely to be small.

Notwithstanding these limitations, we believe that the data from Thomson Reuters do allow for a highly precise identification of the size of stocks' institutional investor bases. Therefore, calculated using these data, the main explanatory variable in our analyzes, $\mathrm{InsIBS}_{i,t}$, denotes the institutional investor base size of a stock and it is defined as the number of institutional investors with long positions in stock i at the end of quarter t.

2.2. Other Data

The data on all of the other stock characteristics used in our analyzes are from the Center for Research in Security Prices (CRSP) and Compustat. To make these data compatible with those from Thomson Reuters, we use observations at the end of calendar quarters and, when merging firms' stock market data with their quarterly accounting data, we assume that the latter become publicly available three months after each firm's fiscal quarter-end. This implies that, at the end of calendar quarter t, we use the quarterly accounting data on stock i's issuing firm from its fiscal quarter that ends three to six months before the end of calendar quarter t.

Thus, using the data from CRSP and Compustat, we calculate three variables related to the valuation, profitability and momentum stock return anomalies that we examine. For stock i at the end of quarter t, the denotations and the definitions of these variables are as follows (see also the Appendix). First, $\mathrm{E/P}_{i,t}$ denotes the earnings-to-price ratio of the stock and it is defined as the ratio of the earnings before extraordinary items and after preferred dividends (Compustat variable IBCOMQ) at the end of quarter t to $\mathrm{M}_{i,t}$ (i.e., the market capitalization of the stock). Second, $\mathrm{ROA}_{i,t}$ denotes the return on assets of the stock and it is defined as the ratio of the earnings before extraordinary items (Compustat variable IBQ) at the end of quarter t to the mean of the total assets (Compustat variable ATQ) at the end of quarters t and $t-1$. Lastly, $\mathrm{RET}_{i,t:t-4}$ denotes the past return on the stock and it is defined as the total return (CRSP variable RET) from the end of quarter $t-4$ to the end of quarter t (similarly as in Jegadeesh and Titman (1993), the return over the last month in quarter t is excluded from the calculation and, in the case of a delisting, the delisting return (CRSP variable DLRET), adjusted as in Shumway (1997) and Shumway and Warther (1999), is included in it). Higher values of these (or similarly calculated) variables have been found to predict higher stock returns.

We examine the stock return anomalies based on these variables because of several reasons. First, these anomalies are consistent with the underreaction of stock prices to new information. Second, they are some of the very few such anomalies that have been reproduced repeatedly in the prior literature, even when using very recent data. Indeed, despite the abundance of claimed anomalies in the US, the vast majority of them have been found to be either irreplicable or irreproducible (see, e.g., Hou et al., 2020; Linnainmaa & Roberts, 2018).[5] Finally, the calculation of the variables above does not lead to great loss of data due to missing values on the necessary CRSP and Compustat variables.

---

[4]Dechow et al. (2001) document that, from 1976 to 1993, more than 98% of the stocks listed on the NYSE and AMEX had less than 5% of their outstanding shares sold short. Moreover, although investors can effectively take short positions by using derivatives (e.g., futures and/or options), Koski and Pontiff (1999) report that around 80% of the US equity mutual funds in their sample do not use such securities at all.

[5]Hou et al. (2020), for example, find that 65% of the 452 stock return anomalies that they analyze are not statistically meaningful at the 5% level when using the standard critical t-value of ± 1.96. In addition, at the same level, when compared to the critical t-value proposed by Harvey et al. (2016) of ± 2.78, which reflects an adjustment for multiple testing, 82% of those anomalies are found to be statistically unimportant.

3. Empirical Analyzes, Results and Potential Explanations

3.1. Sample, Descriptive Statistics and Correlations

In our analyzes, we use a sample of primary common stocks listed on the NYSE, AMEX and NASDAQ from the end of the first quarter of 1980 to the end of the first quarter of 2019. As in the prior literature (e.g., Fama & French, 2008; Hou et al., 2020; Nagel, 2005), these stocks are required to have nonnegative book values of common equity. Moreover, since the stocks with the smallest market capitalizations have been found to be highly influential when studying stock return anomalies (see, e.g., Fama & French, 2008; Hou et al., 2020), following Nagel (2005), we exclude the stocks in the bottom cross-sectional quintiles of $M_{i,t}$ (which are determined using the stocks listed only on the NYSE and AMEX). Finally, the stocks are required to have non-missing values on $InsIBS_{i,t}$, $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$. Therefore, after imposing all of these requirements, our final sample consists of 419,131 stock-quarter observations, covering a total of 13,245 stocks over 157 quarters.

Table 1 presents descriptive statistics and correlations. In Panel A, the presented statistics are the time-series means of the cross-sectional statistics. The mean stock has 135.13 institutional investors, as can be seen from the mean of $InsIBS_{i,t}$. The median of this variable, however, shows that the median stock has a quite smaller institutional investor base, with 84.66 institutional investors, meaning that $InsIBS_{i,t}$ is positively skewed.

Next, Panel B presents the means of $InsIBS_{i,t}$ over time and across stock exchanges and industries, where the industries are determined using the first two digits of the Standard Industrial Classification code. Regardless of their stock exchanges and industries, stocks' institutional investor bases seem to have increased during the last 40 years. Of course, this could be so simply because of the increased number of institutional investors in the stock market over that period (Lewellen, 2011; Sias et al., 2006). Across the stock exchanges, however, there are notable differences between the means of the size of stocks' institutional investor bases. For example, at the end of 2000Q1, the means of $InsIBS_{i,t}$ among the stocks listed on the NYSE, AMEX and NASDAQ are 163.45, 32.34 and 61.61, respectively. Although present, the variation in the means of this variable across the four industries with the most stocks (i.e., the manufacturing, utility, finance and service industries) seems to be much smaller.

Finally, Panel C presents Pearson correlations, where the presented correlations are the time-series means of the cross-sectional correlations. These correlations reveal positive associations between the size of the institutional investor bases of stocks and their market capitalizations and institutional ownership. In particular, the correlations of $InsIBS_{i,t}$ with $\ln(M_{i,t})$ and $InsOWN_{i,t}$, respectively, are 0.85 and 0.38. Stocks with larger institutional investor bases further have higher earnings-to-price ratios ($E/P_{i,t}$), returns on assets ($ROA_{i,t}$) and share turnover ($TURN_{i,t}$). They are also older ($AGE_{i,t}$) and have lower idiosyncratic volatility of their past returns ($IVOL_{i,t:t-4}$).

To explore the high correlation between $InsIBS_{i,t}$ and $\ln(M_{i,t})$ in more detail, in Figure 1, we plot these two variables against each other and, as expected, the relation between them appears to be nonlinear. Specifically, the relation of the size of the institutional investor bases of stocks with their market capitalizations seems to be exponential.

Lastly, the autocorrelations, which are untabulated for brevity, indicate that the size of a stock's institutional investor base is almost perfectly autocorrelated. In particular, the autocorrelation of $InsIBS_{i,t}$ is 0.99. Thus, due to lack of variation in this variable in the time series for a given stock, we focus on its relations with the other variables of our interest in the cross section of stocks.

3.2. Determinants of Institutional Investor Base Size

The correlations in Panel C of Table 1 are unconditional. Because of this and the fact that the results from our analyzes ultimately depend on the ability of $InsIBS_{i,t}$ to serve as a good proxy

Table 1. Descriptive statistics and correlations.

Panel A: Descriptive statistics

Variable	Number of obs.	Mean	Standard deviation	Minimum	25th percentile	Median	75th percentile	Maximum
$\text{InsIBS}_{i,t}$	419,131	135.13	146.18	3.36	51.07	84.66	158.74	842.79
$\text{InsIBS-DED}_{i,t}$	406,028	2.29	2.82	0.00	0.27	1.46	3.09	13.95
$\text{InsIBS-TRA}_{i,t}$	406,028	34.03	29.09	0.50	14.23	25.08	45.28	149.80
$\text{InsIBS-QIX}_{i,t}$	406,028	92.34	107.92	1.97	32.93	54.89	104.19	631.93
$\text{E/P}_{i,t}$	419,131	0.01	0.04	− 0.20	0.00	0.01	0.02	0.09
$\text{ROA}_{i,t}$	419,131	0.01	0.03	− 0.17	0.00	0.01	0.02	0.08
$\text{RET}_{i,t:t-4}$	419,131	0.21	0.48	− 0.58	− 0.08	0.12	0.37	2.30
$\text{ACC}_{i,t}$	369,933	− 0.01	0.04	− 0.20	− 0.03	− 0.01	0.00	0.13
$\text{M}_{i,t}$ (in millions)	419,131	3,306.53	8,508.28	123.00	289.21	710.07	2,163.26	61,435.64
$\text{InsOWN}_{i,t}$	419,131	0.50	0.23	0.01	0.34	0.52	0.67	0.96
$\text{InsOWN-DED}_{i,t}$	406,028	0.03	0.05	0.00	0.00	0.00	0.03	0.31
$\text{InsOWN-TRA}_{i,t}$	406,028	0.12	0.09	0.00	0.05	0.10	0.16	0.41
$\text{InsOWN-QIX}_{i,t}$	406,028	0.34	0.17	0.01	0.22	0.35	0.47	0.72
$\text{TURN}_{i,t}$	419,131	0.29	0.26	0.02	0.12	0.22	0.38	1.43
$\text{BETA}_{i,t:t-4}$	419,131	1.17	1.09	− 1.30	0.46	1.04	1.75	4.64
$\text{IVOL}_{i,t:t-4}$	419,131	0.10	0.06	0.03	0.06	0.09	0.13	0.33
$\text{AGE}_{i,t}$	419,131	18.52	16.52	1.17	6.65	13.77	24.55	73.80

Panel B: Means of $\text{InsIBS}_{i,t}$ over time and across stock exchanges and industries

	Stock exchange			Industry				
Quarter_t	NYSE	AMEX	NASDAQ	Manufacturing	Utility	Finance	Service	Other
1980Q1	48.76	5.63	14.23	41.82	42.58	36.26	18.84	24.31
1985Q1	82.41	15.98	20.37	54.79	66.73	55.39	29.36	38.48
1990Q1	119.73	28.25	35.47	79.87	93.42	74.27	49.71	60.61
1995Q1	126.16	33.54	39.82	85.80	99.86	62.81	64.13	67.69
2000Q1	163.45	32.34	61.61	109.63	139.55	93.04	78.82	95.47
2005Q1	229.97	49.49	101.77	168.15	188.18	135.63	132.88	163.65
2010Q1	265.53	69.55	142.16	213.64	229.64	182.48	171.59	206.37
2015Q1	346.90	80.20	199.65	309.73	335.11	243.01	243.79	240.80
2019Q1	400.95	111.06	246.32	358.16	423.26	291.85	339.50	263.90

(*Continued*)

Table 1. Continued.

Panel C: Pearson correlations

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. $\text{InsIBS}_{i,t}$													
2. $\text{InsIBS-DED}_{i,t}$	0.84												
3. $\text{InsIBS-TRA}_{i,t}$	0.93	0.77											
4. $\text{InsIBS-QIX}_{i,t}$	1.00	0.84	0.90										
5. $\text{E/P}_{i,t}$	0.12	0.08	0.13	0.12									
6. $\text{ROA}_{i,t}$	0.16	0.08	0.18	0.15	0.62								
7. $\text{RET}_{i,t:t-4}$	− 0.01	− 0.04	0.07	− 0.03	0.12	0.12							
8. $\text{ACC}_{i,t}$	0.01	0.00	− 0.01	0.01	0.19	0.17	− 0.00						
9. $\ln(\text{M}_{i,t})$	0.85	0.72	0.86	0.84	0.15	0.20	0.05	0.01					
10. $\text{InsOWN}_{i,t}$	0.38	0.34	0.46	0.33	0.08	0.16	− 0.02	− 0.01	0.38				
11. $\text{InsOWN-DED}_{i,t}$	0.07	0.28	0.05	0.06	− 0.02	− 0.02	− 0.05	− 0.01	0.09	0.29			
12. $\text{InsOWN-TRA}_{i,t}$	0.14	0.09	0.35	0.08	0.02	0.09	0.16	− 0.02	0.16	0.64	− 0.01		
13. $\text{InsOWN-QIX}_{i,t}$	0.41	0.34	0.44	0.40	0.11	0.18	− 0.10	− 0.01	0.39	0.87	0.04	0.34	
14. $\text{TURN}_{i,t}$	0.13	0.09	0.28	0.10	− 0.08	− 0.00	0.11	− 0.03	0.17	0.28	− 0.02	0.43	0.17
15. $\text{BETA}_{i,t:t-4}$	− 0.05	− 0.04	0.01	− 0.06	− 0.11	− 0.07	0.04	− 0.03	− 0.06	0.04	− 0.00	0.13	0.00
16. $\text{IVOL}_{i,t:t-4}$	− 0.31	− 0.27	− 0.25	− 0.33	− 0.29	− 0.27	0.16	− 0.07	− 0.37	− 0.18	− 0.05	0.08	− 0.28
17. $\ln(\text{AGE}_{i,t})$	0.38	0.33	0.32	0.39	0.13	0.12	− 0.05	0.04	0.37	0.17	0.06	− 0.08	0.26

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. In Panel A (C), the presented statistics (correlations) are the time-series means of the cross-sectional statistics (correlations). All variables are winsorized at their cross-sectional 1st and 99th percentiles. The variable definitions are presented in the Appendix.

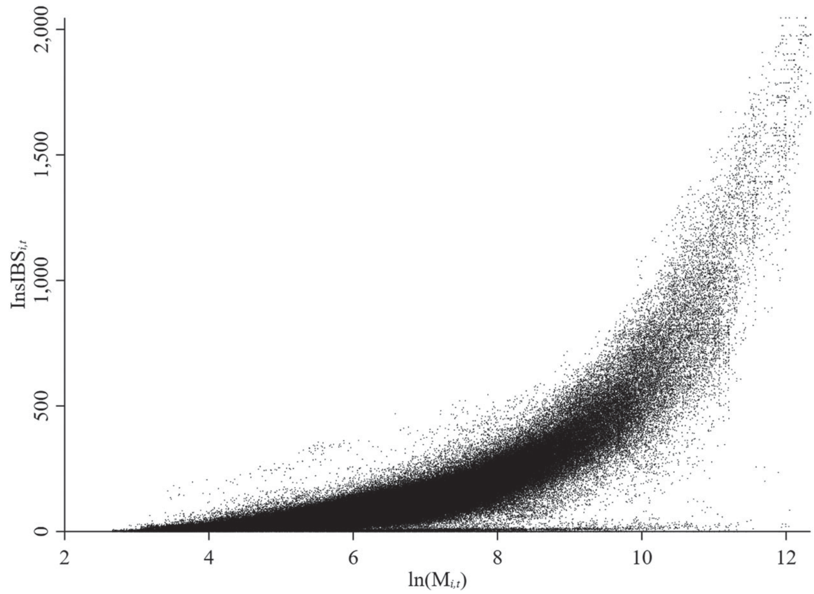

Figure 1. Scatterplot of institutional investor base size against market capitalization.

for the size of the complete investor bases of stocks, before proceeding further, it is important to investigate the construct validity of this variable with more rigor. The results from our analyzes of the determinants of the size of stocks' institutional investor bases are presented in Table 2.

All models present the results from estimating Fama-MacBeth regressions (as in Fama and MacBeth (1973)). These models are first estimated at the end of each quarter. The presented coefficients are then the time-series means of the estimated coefficients. The *t*-statistics (presented in parentheses) are based on the time series of the estimated coefficients and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. Since $\text{InsIBS}_{i,t}$ is bounded by zero and skewed, the dependent variable in all of the models is $\ln(1 + \text{InsIBS}_{i,t})$.

In light of the potentially nonlinear relation of $\text{InsIBS}_{i,t}$ with $\ln(\text{M}_{i,t})$, Model 1 includes this variable and its squared term as independent variables. While the coefficient on the former is positive, the one on the latter is negative, indicating a concave relation between the size of the institutional investor bases of stocks and their market capitalizations. Such relation, however, does not appear in Model 2, in which $\text{InsOWN}_{i,t}$ is included as well. That is, in this model, the coefficients on both $\ln(\text{M}_{i,t})$ and its squared term are positive, which is an additional indication that the relation between this variable and $\text{InsIBS}_{i,t}$ is truly exponential. Moreover, the positive coefficient on $\text{InsOWN}_{i,t}$ is another signal that the size of the institutional investor bases of stocks is positively related to their institutional ownership. All of these relations continue to appear when adding $\text{TURN}_{i,t}$ in Model 3 and several other variables in Models 4 and 5. The results from these models suggest that, on average, stocks with larger institutional investor bases also have higher share turnover ($\text{TURN}_{i,t}$). Additionally, these stocks have higher earnings-to-price ratios ($\text{E/P}_{i,t}$) and market betas ($\text{BETA}_{i,t:t-4}$), and they are older ($\text{AGE}_{i,t}$). In contrast, such stocks have lower past returns ($\text{RET}_{i,t:t-4}$) and lower idiosyncratic volatility of their past returns ($\text{IVOL}_{i,t:t-4}$). All of these results are statistically reliable at the 5% level.

Further, considering the variation in $\text{InsIBS}_{i,t}$ across stock exchanges, it should be noted that all of the models in Table 2 include stock exchange fixed effects. As can be seen from Model 5, adding industry fixed effects has a small impact on the results. Irrespective of the fixed effects

Table 2. Determinants of institutional investor base size.

	Dependent variable: $\ln(1 + \text{InsIBS}_{i,t})$				
Independent variable	1	2	3	4	5
$\ln(\text{M}_{i,t})$	0.73	0.20	0.21	0.22	0.22
	(12.70)	(2.19)	(2.38)	(2.44)	(2.40)
$\ln(\text{M}_{i,t})^2$	− 0.01	0.02	0.02	0.02	0.02
	(− 3.73)	(3.27)	(3.48)	(3.04)	(3.03)
$\text{InsOWN}_{i,t}$		1.84	1.81	1.75	1.77
		(15.61)	(14.81)	(14.55)	(14.01)
$\text{TURN}_{i,t}$			0.14	0.27	0.27
			(3.44)	(4.52)	(4.60)
$\text{E/P}_{i,t}$				0.40	0.42
				(2.30)	(2.91)
$\text{ROA}_{i,t}$				0.20	0.19
				(1.46)	(1.30)
$\text{RET}_{i,t:t-4}$				− 0.12	− 0.12
				(− 7.73)	(− 7.60)
$\text{BETA}_{i,t:t-4}$				0.02	0.02
				(3.65)	(5.31)
$\text{IVOL}_{i,t:t-4}$				− 0.30	− 0.27
				(− 1.98)	(− 1.76)
$\ln(\text{AGE}_{i,t})$				0.08	0.07
				(14.34)	(12.81)
Constant	0.01	0.55	0.55	0.46	0.54
	(0.11)	(3.15)	(3.26)	(2.73)	(3.70)
Additional control variables	No	No	No	No	No
Stock exchange/Industry fixed effects	Yes/No	Yes/No	Yes/No	Yes/No	Yes/Yes
Number of quarters	157	157	157	157	157
Number of observations	419,131	419,131	419,131	419,131	419,092
Mean adjusted R^2 (in percent)	69.51	82.76	82.96	84.16	84.56

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. All models are first estimated at the end of each quarter. The presented coefficients are then the time-series means of the estimated coefficients. The t-statistics (presented in parentheses) are based on the time series of the estimated coefficients and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

used, however, most of the variation in $\text{InsIBS}_{i,t}$ seems to be explained by $\text{M}_{i,t}$ and $\text{InsOWN}_{i,t}$. Specifically, while the mean adjusted R^2 of Model 2 is 82.76%, those of Models 3–5 are just around 1–2% higher. Despite these high mean adjusted R^2s, however, the results in Table 2 should be interpreted with caution because they could reflect only associations of $\text{InsIBS}_{i,t}$ with the other variables. Undoubtedly, an empirical investigation that would be more informative of the causal relations between these variables requires a more rigorous consideration of their presumably endogenous nature, especially with respect to the potential reverse causality between them. Nonetheless, the results seem quite sensible. Therefore, we believe that $\text{InsIBS}_{i,t}$ does serve as a good proxy for the size of stocks' complete investor bases.

3.3. Raw Returns

Two methodological approaches for identifying patterns in the cross section of stock returns have been commonly used in the prior literature. The first approach involves examining the returns on long-short portfolios of stocks, formed by first sorting stocks into quantiles on the basis of their characteristics and then by simultaneously taking long and short positions in the stocks from the extreme quantiles. In contrast, the second approach involves studying the returns on individual stocks, by estimating Fama-MacBeth regressions. Since each of these approaches has its own advantages and disadvantages, in our analyzes, we use both of them.

Namely, following Hong et al. (2000) and Nagel (2005), we use sorts, implemented with the residuals from estimating Fama-MacBeth regressions. In particular, instead of using $\text{InsIBS}_{i,t}$ as the primary sorting variable, our main analyzes exploit the residuals from Model 2 in Table 2. The variable containing these residuals is denoted $\text{InsIBS}^{r2}_{i,t}$.

This approach is chosen due to several reasons. First, given the relations of $M_{i,t}$ and $\text{InsOWN}_{i,t}$ with $\text{InsIBS}_{i,t}$, we cannot simply sort stocks on the basis of this variable. In fact, doing so would be largely equivalent to sorting on the basis of $M_{i,t}$ and $\text{InsOWN}_{i,t}$, both of which have already been found to be helpful in explaining stock return anomalies because of other underlying mechanisms than the one studied here (see, e.g., Griffin & Lemmon, 2002; Nagel, 2005). In contrast, by removing the effects of these variables on $\text{InsIBS}_{i,t}$ through the Fama-MacBeth regressions, our approach enables us to use the part of this variable that is unrelated to $M_{i,t}$ and $\text{InsOWN}_{i,t}$. Second, because of the use of sorts, this approach does not impose any functional forms on the relations between $\text{InsIBS}_{i,t}$ and the other variables of our interest. Lastly, our approach allows us to control for a relatively large number of variables simultaneously, which is usually impossible by using only sorts because forming stock portfolios with a sufficient number of stocks is ordinarily feasible by sorting them on the basis of two to three of their characteristics at a time.

Before we move on to our analyzes of the stock portfolio returns, it is important to explore the effectiveness of our approach in more detail. Hence, Panel A in Table 3 presents the means of $\text{InsIBS}_{i,t}$, $M_{i,t}$ and $\text{InsOWN}_{i,t}$ across the quintiles of $\text{InsIBS}^{r2}_{i,t}$. In this panel, at the end of each quarter, the stocks are first sorted into quintiles on the basis of $\text{InsIBS}^{r2}_{i,t}$. The presented means are then the time-series means of the cross-sectional means, which are calculated at the end of each quarter and for each of the quintiles of $\text{InsIBS}^{r2}_{i,t}$.

From the bottom to the top quintile of $\text{InsIBS}^{r2}_{i,t}$, the means of $\text{InsIBS}_{i,t}$ increase almost monotonically, from 88.24 to 145.36, or by nearly 65%. Although there is some variation in the means of $M_{i,t}$ and $\text{InsOWN}_{i,t}$ across all of the quintiles of $\text{InsIBS}^{r2}_{i,t}$, this variation is considerably smaller than the variation in these variables across all of the stocks in our sample (see Panel A in Table 1). More importantly, across the extreme quintiles of $\text{InsIBS}^{r2}_{i,t}$, the variation in the means of $M_{i,t}$ is almost nonexistent and the variation in the means of $\text{InsOWN}_{i,t}$ is quite small (and indeed much smaller than that in the means of $\text{InsIBS}_{i,t}$). Interestingly, the mean of $\text{InsOWN}_{i,t}$ is somewhat higher in the bottom quintile of $\text{InsIBS}^{r2}_{i,t}$ than in its top quintile. Nevertheless, across these quintiles, the results suggest that our approach is effective in preserving a substantial variation in the size of the institutional investor bases of stocks, while holding their market capitalizations and institutional ownership fairly constant.

Further, the Pearson correlations of $\text{InsIBS}^{r2}_{i,t}$ with the other variables used in our analyzes are presented in Panel B. In this panel, the presented correlations are the time-series means of the cross-sectional correlations. These correlations indicate that, apart from the dependent variable in Model 2 of Table 2 (i.e., $\ln(1 + \text{InsIBS}_{i,t})$), $\text{InsIBS}^{r2}_{i,t}$ is not particularly related to any of the other variables. Most importantly, the residual size of the institutional investor bases of stocks seems to be completely unrelated to their market capitalizations and institutional ownership. Indeed, the correlations of $\text{InsIBS}^{r2}_{i,t}$ with both $\ln(M_{i,t})$ and $\text{InsOWN}_{i,t}$ are almost exactly zero.

Finally, Panel C presents the raw quarterly stock portfolio returns based on the quintiles of $\text{InsIBS}^{r2}_{i,t}$ and $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$, respectively. Here, at the end of each quarter, the stocks are first sorted into quintiles on the basis of $\text{InsIBS}^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$, respectively. Lastly, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $\text{InsIBS}^{r2}_{i,t}$, and the equal-weighted total raw returns over the next quarter on these and the

Table 3. Raw quarterly portfolio returns.

Panel A: Means across quintiles of $InsIBS^{r2}_{i,t}$

	$InsIBS^{r2}_{i,t}$				
Variable	1	2	3	4	5
$InsIBS_{i,t}$	88.24	137.73	150.29	154.06	145.36
$M_{i,t}$ (in millions)	2,760.84	3,743.03	3,719.25	3,561.15	2,748.02
$InsOWN_{i,t}$	0.50	0.57	0.54	0.49	0.41

Panel B: Pearson correlations

Variable	$InsIBS_{i,t}$	$\ln(1+InsIBS_{i,t})$	$E/P_{i,t}$	$ROA_{i,t}$	$RET_{i,t:t-4}$	$\ln(M_{i,t})$	$InsOWN_{i,t}$	$TURN_{i,t}$	$BETA_{i,t:t-4}$	$IVOL_{i,t:t-4}$	$\ln(AGE_{i,t})$
$InsIBS^{r2}_{i,t}$	0.16	0.41	– 0.00	0.00	– 0.12	– 0.00	0.00	0.05	0.05	– 0.02	0.14

Panel C: Raw quarterly portfolio returns based on quintiles of $InsIBS^{r2}_{i,t}$ and $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$, respectively

	$InsIBS^{r2}_{i,t}$					
$E/P_{i,t}$	1	2	3	4	5	1 – 5
1	3.21	3.49	3.97	4.88	6.05	– 2.84
2	3.05	3.72	3.71	3.96	4.65	– 1.60
3	3.90	3.82	3.76	4.05	4.77	– 0.87
4	4.20	4.25	4.58	4.53	5.08	– 0.88
5	5.44	5.21	5.55	5.60	5.94	– 0.50
5 – 1	2.23	1.72	1.59	0.72	– 0.11	2.34
	(2.55)	(2.34)	(2.14)	(0.83)	(– 0.12)	(2.76)
Standard deviation	8.74	7.58	7.86	9.85	12.00	8.92
Mean number of stocks	191.71	201.97	202.87	201.68	196.78	388.49

(*Continued*)

Table 3. Continued.

	$\text{InsIBS}^{r2}_{i,t}$					
$\text{ROA}_{i,t}$	1	2	3	4	5	1 − 5
1	3.28	3.61	4.04	4.93	5.65	− 2.37
2	3.76	3.93	4.19	4.35	5.20	− 1.44
3	4.18	3.83	4.31	4.31	5.17	− 1.00
4	3.85	4.30	4.25	4.59	5.15	− 1.30
5	4.77	4.81	4.75	4.86	5.27	− 0.50
5 − 1	1.49	1.19	0.71	− 0.07	− 0.38	1.88
	(2.36)	(2.28)	(1.30)	(− 0.10)	(− 0.44)	(2.47)
Standard deviation	7.05	6.57	6.44	8.96	10.81	8.86
Mean number of stocks	191.41	202.76	203.30	202.21	197.90	389.31
$\text{RET}_{i,t:t-4}$						
1	2.70	3.28	3.70	4.28	5.69	− 3.00
2	3.64	3.56	4.20	4.33	5.42	− 1.79
3	4.10	3.94	4.17	4.49	5.24	− 1.15
4	4.49	4.33	4.40	4.70	4.80	− 0.31
5	5.05	5.43	5.14	5.26	5.33	− 0.28
5 − 1	2.35	2.14	1.45	0.98	− 0.36	2.72
	(2.60)	(2.46)	(1.74)	(1.17)	(− 0.31)	(2.58)
Standard deviation	10.29	9.24	8.84	9.12	13.32	11.01
Mean number of stocks	190.30	201.75	203.03	202.09	197.77	388.07

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. In Panel A, at the end of each quarter, these stocks are first sorted into quintiles on the basis of $\text{InsIBS}^{r2}_{i,t}$. The presented means are then the time-series means of the cross-sectional means, which are calculated at the end of each quarter and for each of the quintiles of $\text{InsIBS}^{r2}_{i,t}$. In Panel B, the presented correlations are the time-series means of the cross-sectional correlations. In Panel C, at the end of each quarter, the stocks are first sorted into quintiles on the basis of $\text{InsIBS}^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $\text{InsIBS}^{r2}_{i,t}$, and the equal-weighted total raw returns over the next quarter on these and the long-short portfolios are calculated. The presented returns (in percent) are the time-series means of these portfolio returns. The t-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

long-short portfolios are calculated.[6] The presented returns (in percent) are the time-series means of these portfolio returns. The *t*-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters.

For the trading strategy based on $\mathrm{E/P}_{i,t}$, the results show that, among the stocks in the bottom quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, while the raw quarterly return on the portfolio of the stocks in the bottom quintile of $\mathrm{E/P}_{i,t}$ is 3.21%, the one on the portfolio of the stocks in the top quintile of this variable is 5.44%. The differential between these returns, which is the return on the long-short portfolio, is 2.23% (*t*-statistic = 2.55). More importantly, from the bottom to the top quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, this return monotonically decreases. Actually, among the stocks in the top quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, the return is not even positive; it is −0.11% (*t*-statistic = −0.12). Hence, the differential between these returns is 2.34% (*t*-statistic = 2.76). Around 84% ($= \{[-0.87 - (-2.84)]/2.34\} \times 100$) of it arises on the short side (i.e., among the stocks in the bottom quintile of $\mathrm{E/P}_{i,t}$; in this calculation, as a benchmark, we use the −0.87% return differential between the portfolios in the middle quintile of $\mathrm{E/P}_{i,t}$, since they are neutral in terms of this variable).

Next, among the stocks in the bottom quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, the raw quarterly long-short portfolio return when the strategy is based on $\mathrm{ROA}_{i,t}$ is 1.49% (*t*-statistic = 2.36). As before, this return monotonically decreases from the bottom to the top quintile of $\mathrm{InsIBS}^{r2}_{i,t}$ and, among the stocks in the latter quintile, the return is negative (i.e., it is −0.38% (*t*-statistic = −0.44)). Thus, the differential between these returns is 1.88% (*t*-statistic = 2.47). The short side provides approximately 73% ($= \{[-1.00 - (-2.37)]/1.88\} \times 100$) of it.

Finally, if the strategy is based on $\mathrm{RET}_{i,t:t-4}$, the raw quarterly return on the long-short portfolio among the stocks in the bottom quintile of $\mathrm{InsIBS}^{r2}_{i,t}$ is 2.35% (*t*-statistic = 2.60). However, from the bottom to the top quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, this return again monotonically decreases. It is also negative among the stocks in the latter quintile. Specifically, among these stocks, the return is −0.36% (*t*-statistic = −0.31). Therefore, the differential between these returns is 2.72% (*t*-statistic = 2.58). About 68% ($= \{[-1.15 - (-3.00)]/2.72\} \times 100$) of it comes from the short side.

Overall, the results in Panel C of Table 3 indicate that the stock return anomalies based on stocks' earnings-to-price ratios, returns on assets and past returns are statistically and economically notable only among stocks with smaller residual institutional investor bases. As such, these results support our prediction that those anomalies arise and persist mostly among stocks with smaller investor bases. Hence, insofar as such stocks are neglected stocks, the results are consistent with the hypothesis that the incomplete dissemination of information across investors helps in explaining the occurrence and the persistence of the anomalies.

In particular, they could arise and persist because the assumption in the CAPM that all investors know about and follow all of the available securities is significantly violated in reality. This could be so if investors face fixed costs of information acquisition (Merton, 1987) and/or if their attention is constrained (Peng & Xiong, 2006). In spite of the specific restriction(s) on the behavior of investors, unlike in the CAPM, all of the available information about all of the available securities would not be instantaneously disseminated across all investors, which means

[6] Several issues about the formation of these stock portfolios and the calculation of their returns are important to note. First, among the stocks in a given quintile of $\mathrm{InsIBS}^{r2}_{i,t}$, the long-short portfolios are formed by simultaneously taking long and short positions in the stocks from the top and the bottom quintiles, respectively, of $\mathrm{E/P}_{i,t}$, $\mathrm{ROA}_{i,t}$ and $\mathrm{RET}_{i,t:t-4}$. Second, the number of stocks across all of the portfolios is nearly identical. Specifically, on average, each portfolio contains roughly 100 stocks, which implies that each long-short portfolio contains approximately 200 stocks. Finally, because equal-weighted portfolio returns can be significantly influenced by the returns on the stocks with the smallest market capitalizations, it is worth reiterating that these stocks, which are quite numerous but economically less relevant (Hou et al., 2020), are excluded from our analyzes.

that they would have to take their positions while having incomplete information. Under these conditions, it is possible for the prices of neglected stocks to underreact to both positive and negative information, in the sense that they can increase less than they should when there is positive information and they can decrease less than they should when there is negative information. As a result, these stocks can be mispriced (i.e., their prices can deviate from their true values) for an extended period of time, which could be the reason why their returns seem to be predictable with stock characteristics other than market beta.

For a more complete understanding of this explanation for our results, two aspects of it are important to keep in mind. First, as explained in Hong and Stein (2007), for stock prices to be affected by the incomplete dissemination of information across investors, their costs of obtaining information and/or their limited attention are insufficient. Indeed, what is also required is that investors are unsophisticated in another, more subtle way. Namely, when trading, investors should not fully consider that they can be at an information disadvantage (i.e., they should not infer the correct information from the trades of others). Second, as pointed out in Merton (1987), if some investors do not know about and do not follow a certain stock, then the disclosure of any information concerning that stock is unlikely to cause those investors to take positions in it. Thus, our explanation is pertinent to both public and private information as well as to stock-, firm-, industry- and market-specific information.

Furthermore, the results in Panel C of Table 3 show that the effect of the residual size of stocks' institutional investor bases on the stock return anomalies is concentrated almost entirely on the short side of the trading strategies (i.e., among the seemingly overpriced stocks from the bottom quintiles of the anomaly variables). Hence, by using only the stocks in the bottom quintiles of $\mathrm{E/P}_{i,t}$, $\mathrm{ROA}_{i,t}$ and $\mathrm{RET}_{i,t:t-4}$, investors could earn even higher returns than the return differentials presented above by simply taking long positions in the stocks from the top quintile of $\mathrm{InsIBS}^{r2}_{i,t}$ and short positions in the stocks from the bottom quintile of this variable. The raw quarterly return on each respective strategy would then be 2.84% (untabulated t-statistic $= 3.24$), 2.37% (untabulated t-statistic $= 3.01$) and 3.00% (untabulated t-statistic $= 3.35$). Importantly, these returns would be earned after controlling not only for $\mathrm{M}_{i,t}$ and $\mathrm{InsOWN}_{i,t}$, but also for $\mathrm{E/P}_{i,t}$, $\mathrm{ROA}_{i,t}$ and $\mathrm{RET}_{i,t:t-4}$, respectively.

Nonetheless, these results suggest that the underreaction to negative information of the prices of neglected stocks is greater than their underreaction to positive information, which is intuitively sensible in a theory based on the incomplete dissemination of information across investors. Specifically, for various reasons, such as their compensation and/or reputation, managers could prefer their firms' stock prices to be higher rather than lower. In that case, it seems plausible that they would have weaker incentives for disseminating negative information than for disseminating positive information (Kothari et al., 2009), which means that they are likely to expend less of their resources, such as time and/or effort, on the former than on the latter. Managers could do so, for example, by reducing their voluntary disclosures and/or their financial media presence. Thus, under these conditions, negative information can be disseminated across investors more incompletely than positive information.

3.4. Risk-Adjusted Returns

One alternative explanation for the patterns in Panel C of Table 3 could be that stocks with smaller residual institutional investor bases are riskier stocks. Hence, in Table 4, we check if the same patterns continue to appear among the risk-adjusted quarterly stock portfolio returns.

In this table, at the end of each quarter, the formation of the long-short portfolios and the calculation of their raw quarterly returns are the same as in Panel C of Table 3. However, the presented returns (in percent) are the alphas from estimating the CAPM, the three-factor model of Fama

Table 4. Risk-adjusted quarterly portfolio returns.

	CAPM alpha			FF3 alpha			PS4 alpha		
	$InsIBS^{r2}_{i,t}$			$InsIBS^{r2}_{i,t}$			$InsIBS^{r2}_{i,t}$		
$E/P_{i,t}$	1	5	1 − 5	1	5	1 − 5	1	5	1 − 5
1	− 0.86	1.31	− 2.16	− 0.53	1.60	− 2.13	− 0.55	1.57	− 2.12
5	2.38	2.71	− 0.32	1.96	2.16	− 0.20	1.89	2.14	− 0.25
5 − 1	3.24	1.40	1.84	2.49	0.56	1.93	2.44	0.57	1.87
	(4.01)	(1.65)	(2.40)	(3.45)	(0.91)	(2.68)	(3.71)	(0.94)	(2.94)
$ROA_{i,t}$									
1	− 0.74	0.98	− 1.73	− 0.36	1.35	− 1.71	− 0.42	1.32	− 1.74
5	1.24	1.66	− 0.42	1.57	1.83	− 0.26	1.55	1.85	− 0.30
5 − 1	1.98	0.68	1.31	1.93	0.48	1.45	1.97	0.53	1.44
	(3.21)	(0.89)	(1.92)	(2.98)	(0.71)	(2.33)	(2.95)	(0.77)	(2.28)
$RET_{i,t:t-4}$									
1	− 0.95	1.16	− 2.11	− 1.18	1.25	− 2.43	− 1.22	1.34	− 2.56
5	1.18	1.54	− 0.36	1.87	1.80	0.06	1.71	1.73	− 0.03
5 − 1	2.13	0.38	1.76	3.05	0.56	2.49	2.93	0.39	2.54
	(2.43)	(0.37)	(1.65)	(3.49)	(0.54)	(2.84)	(3.45)	(0.35)	(2.82)

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. At the end of each quarter, these stocks are first sorted into quintiles on the basis of $InsIBS^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $InsIBS^{r2}_{i,t}$, and the equal-weighted total raw returns over the next quarter on these and the long-short portfolios are calculated. The presented returns (in percent) are the alphas from estimating the CAPM, FF3 and PS4 models with the time series of these portfolio returns. The *t*-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

and French (1993) (denoted FF3) and the four-factor model of Pástor and Stambaugh (2003) (denoted PS4), where all models are estimated with the time series of those portfolio returns.[7] The *t*-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters.

Almost all of the long-short portfolio return differentials are positive and statistically discernible at the 5% level, which indicates that the premiums on our trading strategies are reliably higher among stocks with smaller $InsIBS^{r2}_{i,t}$. For instance, when the strategy is based on $E/P_{i,t}$, among the stocks in the bottom quintile of $InsIBS^{r2}_{i,t}$, the quarterly FF3 alpha of the long-short portfolio is 2.49% (*t*-statistic = 3.45). As before, from the bottom to the top quintile of $InsIBS^{r2}_{i,t}$, this alpha monotonically decreases and, among the stocks in the latter quintile, it is 0.56% (*t*-statistic = 0.91). Thus, the differential between these alphas is 1.93% (*t*-statistic = 2.68). Nearly all of the other premium differentials and their *t*-statistics, including those for the strategies based on $ROA_{i,t}$ and $RET_{i,t:t-4}$, are of a similar magnitude. Overall, the results in Table 4 suggest that the stock return anomalies that appear among stocks with smaller investor bases remain unexplained by the exposures of these stocks to the market, size, value and liquidity factors.

[7] When estimating these models, the returns on the one-month US treasury bill (RF_t) and the market ($MKTRF_t$), size (SMB_t) and value (HML_t) factors are obtained from the personal website of Kenneth R. French, whereas the returns on the liquidity factor (LIQ_t) are obtained from the personal website of Ľuboš Pástor.

3.5. Transaction Costs

Stocks with smaller residual institutional investor bases could also be stocks with higher transaction costs, especially when it comes to short selling. Miller (1977) hypothesizes that the prices of such stocks reflect mostly the positive opinions of optimistic investors who can and do take long positions, but they do not reflect the negative opinions of pessimistic investors who cannot and do not take short positions. If this is true, then these stocks should be principally overpriced (i.e., their prices should underreact mainly to negative information) and their returns should be more predictable with other stock characteristics than market beta. Transaction costs in general, however, and those of short selling in particular, have been found to be higher among stocks with smaller market capitalizations and/or lower institutional ownership.

Indeed, many stock return anomalies have been found to be more pronounced among stocks with smaller market capitalizations (see, e.g., Hou et al., 2020). Although the exact reason for this finding is not entirely clear, a few of the various proposed reasons for it are related to explicit and implicit transaction costs, both of which have been found to be higher among these stocks (see, e.g., Lesmond et al., 1999; Stoll & Whaley, 1983). For this reason, it is important to remember here that our results obtain after excluding the stocks with the smallest market capitalizations and after controlling for $M_{i,t}$. Thus, inasmuch as the market capitalization of a certain stock is related to the overall transaction costs of that stock, the effect of the residual size of the institutional investor bases of stocks on the stock return anomalies seems to be distinct from the effects of both of these other stock characteristics.

Further, while Duffie et al. (2002) suggest that institutional investors (especially those that take larger long positions over a longer period, such as pension funds) are the main lenders of shares, D'Avolio (2002) finds that the institutional ownership of US stocks explains most of the cross-sectional variation in their supply of lendable shares. This indicates that the supply of such shares is lower among stocks with lower institutional ownership, meaning that investors are likely to find it more difficult to sell these stocks short. Considering this and the theory of Miller (1977), Nagel (2005) proposes that an overpricing is more likely to arise among such stocks. He additionally argues that, if a certain stock becomes overpriced and investors have difficulties short-selling that stock, then the only way for the overpricing to disappear is for the investors who own the stock to be sophisticated enough to first realize that it is overpriced and then to sell some or all of their shares. Hence, since institutional investors are commonly presumed to be more sophisticated than other investors (e.g., individuals), Nagel (2005) also suggests that an overpricing is more likely to persist among stocks with lower institutional ownership. Consistent with these predictions, he further provides empirical evidence that the predictability of stock returns associated with stocks' market-to-book ratios, share turnover, dispersion of analysts' earnings forecasts and total volatility of past returns is more pronounced among such stocks. More importantly, most of the variation in the stock return predictability based on these stock characteristics is found to appear on the short side of the trading strategies, which is exactly where the impediments to taking short positions would limit arbitrage.

Although they may seem similar, our results differ considerably from those of Nagel (2005). First, they pertain to different stock return anomalies. Second, while all of the anomalies that we investigate have been reproduced recently, except for the anomaly based on stocks' market-to-book ratios, the same cannot be said about the other anomalies studied by Nagel (2005). Third, whereas he does not control for the size of the investor bases of stocks, our results emerge after controlling for their institutional ownership (i.e., $\text{InsOWN}_{i,t}$). Finally, and perhaps most importantly, to the extent that the institutional ownership of a certain stock is related to the transaction costs of short selling that stock, such costs seem to be lower (not higher) among stocks with smaller residual institutional investor bases. That is, even though we control for $\text{InsOWN}_{i,t}$, there is still some variation left in this variable across the quintiles of $\text{InsIBS}^{r2}_{i,t}$ (see Panel A in Table 3). Intriguingly, however, the mean of $\text{InsOWN}_{i,t}$ is moderately higher in the bottom quintile of $\text{InsIBS}^{r2}_{i,t}$ than in its top quintile, which indicates that the supply of lendable shares is higher among stocks with smaller residual institutional investor bases. Therefore, among these stocks, the short-sales constraints appear to be less severe, suggesting that our results are unlikely to reflect the underlying mechanism discussed in Miller (1977) and Nagel (2005).

3.6. Persistence

In this section, we analyze the persistence of our results, by examining whether they continue to appear over holding periods longer than one quarter. In particular, we repeat all of our analyzes as in Tables 3 and 4, but now we also calculate the returns on the long-short stock portfolios over the next two to four quarters after their formation. For brevity, Table 5 presents only the raw and the risk-adjusted premium differentials over the different holding periods.[8]

Almost all of the premium differentials are positive and statistically meaningful at the 5% level, which indicates that the premiums on our trading strategies are reliably and persistently higher among stocks with smaller $\text{InsIBS}^{r2}_{i,t}$. For instance, when the strategy is based on $\text{RET}_{i,t:t-4}$, among the stocks in the bottom quintile of $\text{InsIBS}^{r2}_{i,t}$, the untabulated yearly FF3 alpha of the long-short portfolio is 6.57% (t-statistic $= 2.24$). From the bottom to the top quintile of $\text{InsIBS}^{r2}_{i,t}$, this alpha monotonically decreases and, among the stocks in the latter quintile, the alpha is not even positive; it is -2.84% (t-statistic $= -0.88$). Hence, the differential between these alphas is 9.41% (t-statistic $= 3.01$). Nearly all of the other yearly premium differentials and their t-statistics, including those for the strategies based on $\text{E/P}_{i,t}$ and $\text{ROA}_{i,t}$, are similar in magnitude. Therefore, the results in Table 5 suggest that the stock return anomalies that arise among stocks with smaller investor bases persist for at least four quarters.

3.7. Seasonality

To explore our interpretation of the results further, in this section, we seek to investigate if they vary with the amount of information in the market. If our results truly reflect the incomplete dissemination of information across investors due to their costs of obtaining information and/or their limited attention, then it seems plausible that the constraints of this kind on investors' behavior would be more severe during periods when there is more information to be disseminated. In that case, our results should be more pronounced during such periods.

Relative to the second and the third quarters, during the first and the fourth quarters of each calendar year, the overall supply of information in the market is likely to be higher. Indeed, almost all of the public information for a given calendar year is usually available toward the end of the fourth quarter of that year. Moreover, because of this, around that time, the public information that is not yet available is likely to be easier to estimate with relatively high precision. In any case, the complete public information for a given calendar year is typically available by the end of the first quarter of the following calendar year.[9] Overall, therefore, as long as there are these kinds of seasonal variations in the amount of information in the market, our results should

---

[8] In this table, the holding periods beyond one quarter are overlapping. Thus, for a holding period of q quarters, during each quarter, the entire portfolio of each trading strategy consists of the portfolios formed at the end of each of the previous q quarters. At the end of quarter t, however, apart from opening new positions, the strategies involve closing the positions taken at the end of quarter $t-q$. Consequently, at the end of each quarter, while the weights allocated to $1/q$ of the stocks in the entire portfolio of each strategy are revised, the weights allocated to the remaining stocks are the same as those at the end of the prior quarter.

[9] It is important to note that the fiscal years of most firms in the US are equivalent to the calendar years and that the US SEC has been requiring firms to disclose their 10-K forms within 60–90 days after their fiscal year-ends. While the fiscal

Table 5. Raw and risk-adjusted premium differentials over different holding periods.

Holding period (in quarters)	Raw return	CAPM alpha	FF3 alpha	PS4 alpha
		Strategy with $E/P_{i,t}$		
1	2.34	1.84	1.93	1.87
	(2.76)	(2.40)	(2.68)	(2.94)
2	4.55	3.31	2.97	3.06
	(2.92)	(2.21)	(2.21)	(2.44)
3	6.69	4.89	4.09	4.12
	(3.00)	(2.35)	(2.32)	(2.43)
4	9.47	7.77	5.87	6.01
	(3.10)	(2.65)	(2.52)	(2.69)
		Strategy with $ROA_{i,t}$		
1	1.88	1.31	1.45	1.44
	(2.47)	(1.92)	(2.33)	(2.28)
2	4.18	2.91	2.85	2.78
	(2.92)	(2.41)	(2.56)	(2.42)
3	6.31	4.90	4.86	5.03
	(3.08)	(3.08)	(3.25)	(2.99)
4	8.62	7.73	7.48	8.05
	(3.41)	(3.93)	(3.79)	(3.64)
		Strategy with $RET_{i,t:t-4}$		
1	2.72	1.76	2.49	2.54
	(2.58)	(1.65)	(2.84)	(2.82)
2	5.71	3.60	5.02	4.41
	(2.85)	(2.16)	(3.47)	(3.31)
3	8.05	5.20	7.44	6.45
	(2.79)	(2.40)	(3.20)	(3.00)
4	10.64	6.81	9.41	7.54
	(2.78)	(2.56)	(3.01)	(2.76)

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. At the end of each quarter, these stocks are first sorted into quintiles on the basis of $InsIBS^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $InsIBS^{r2}_{i,t}$, and the equal-weighted total raw returns over the next one to four quarters on these and the long-short portfolios are calculated. The presented returns (in percent) are the $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$ premium differentials (i.e., the differentials in the long-short portfolio returns) between the stocks in the extreme quintiles of $InsIBS^{r2}_{i,t}$, calculated as the time-series means of the premium differentials and as the alphas from estimating the CAPM, FF3 and PS4 models with the time series of the premium differentials. The t-statistics (presented in parentheses) are based on the time series of the premium differentials and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

be more pronounced when the previously examined trading strategies are implemented at the end of the first and the fourth calendar quarters.

As a test of this prediction, we repeat our analyzes as in Table 3, but instead of calculating the time-series means of the long-short stock portfolio returns over all of the portfolio formation calendar quarters, we now calculate their time-series means and medians for each of those quarters. The raw quarterly premium differentials over the different portfolio formation calendar quarters are presented in Table 6.

years of around 68% of the firms in our sample are equivalent to the calendar years, about 81% of these firms have their fiscal year-ends during the first or the fourth calendar quarters.

Table 6. Raw quarterly premium differentials over different portfolio formation calendar quarters.

Portfolio formation calendar quarter	Mean raw return	Median raw return	Percentage of quarters with positive raw returns
	Strategy with $E/P_{i,t}$		
1	1.70	2.08	58.97
2	1.40	0.17	51.28
3	1.79	− 0.13	48.72
4	4.47	3.79	71.79
	Strategy with $ROA_{i,t}$		
1	1.90	1.46	58.97
2	1.11	0.05	51.28
3	1.45	0.59	56.41
4	3.05	3.53	66.67
	Strategy with $RET_{i,t:t-4}$		
1	2.67	2.99	61.54
2	0.25	0.61	58.97
3	2.85	0.04	53.85
4	5.10	5.80	66.67

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. At the end of each quarter, these stocks are first sorted into quintiles on the basis of $InsIBS^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $InsIBS^{r2}_{i,t}$, and the equal-weighted total raw returns over the next quarter on these and the long-short portfolios are calculated. The presented returns (in percent) are the $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$ premium differentials (i.e., the differentials in the long-short portfolio returns) between the stocks in the extreme quintiles of $InsIBS^{r2}_{i,t}$, calculated as the time-series means and medians of the premium differentials for each portfolio formation calendar quarter. The variable definitions are presented in the Appendix.

All of the premium differentials are cumulatively positive and much higher for the first and the fourth portfolio formation quarters than for the other quarters, indicating that our trading strategies earn considerably higher premiums among stocks with smaller $InsIBS^{r2}_{i,t}$ when they are implemented at the end of those quarters. This implies that these premiums are almost fully realized during the first two quarters. Although these findings are strongly supported by both the mean and the median premium differentials, the latter seem to offer a much clearer image. For example, when the strategy is based on $E/P_{i,t}$ and implemented at the end of the first and the fourth quarters, the median raw quarterly premium differentials are 2.08% and 3.79%, respectively. Conversely, the corresponding premium differentials for the second and the third quarters are substantially lower, 0.17% and −0.13%, respectively. Apart from these results, our analyzes reveal that, across all of the strategies, the premium differentials are positive for 58.97–71.79% of the first and the fourth quarters. This, however, is the case for 48.72–58.97% of the other quarters. Overall, the results in Table 6 agree with our prediction regarding the seasonality of the stock return anomalies that appear among stocks with smaller investor bases. As such, these results corroborate our interpretation of the main results that they reflect an underlying mechanism related to the degree of information dissemination across investors.

In contrast to this mechanism, one might wonder if our results could be driven by the seasonal trading patterns of institutional investors. Probably due to tax and/or image considerations, these investors have been found to sell more of their stocks with smaller market capitalizations (and lower past returns) during the fourth calendar quarter (see, e.g., Ng & Wang, 2004; Sikes, 2014). Thus, at the end of that quarter, such stocks are likely to have particularly smaller institutional investor bases (and lower institutional ownership), which means that they could dominate the stocks in the bottom quintile of $\text{InsIBS}_{i,t}^{r2}$ at that time. There are several reasons why, however, this is unlikely to be the case. First, since we control for $\text{M}_{i,t}$ and $\text{InsOWN}_{i,t}$, the market capitalizations and the institutional ownership of the stocks in the extreme quintiles of $\text{InsIBS}_{i,t}^{r2}$ are roughly the same (see Panel A in Table 3). Second, as shown above, all of the trading strategies earn considerably higher premiums among stocks with smaller $\text{InsIBS}_{i,t}^{r2}$ when they are implemented not only at the end of the fourth quarter, but also at the end of the first quarter. Finally, and perhaps most importantly, our results are not even consistent with the alternative explanation of the kind discussed here. Indeed, under this explanation, the stocks that are the most sold by institutional investors during the fourth quarter should be the ones that, at the end of that quarter, end up not only in the bottom quintile of $\text{InsIBS}_{i,t}^{r2}$, but also in the bottom quintiles of $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$, respectively. Hence, if the abnormal selling of these stocks disappears in the first quarter, then, during that quarter, they should not underperform the stocks in the top quintiles of any of these variables. For example, the strategy based on $\text{RET}_{i,t:t-4}$ should have produced results similar to those in Table 4 in Jegadeesh and Titman (1993), who find that the momentum premium is reliably negative in January (and nonexistent in February and March). Instead, our results suggest that, after controlling for the size of stocks' investor bases, this premium exists reliably only among stocks with smaller investor bases, that it is actually positive and that it is realized mainly during the first quarter. Overall, therefore, given that institutional investors' fourth-quarter trading has been one of the most commonly proposed explanations for the so-called turn-of-the-year (or January) effect (see, e.g., Keim, 1983; Reinganum, 1983), it is important to note that, while this investor behavior cannot explain our results, it seems that the incomplete dissemination of information across investors could help in explaining the turn-of-the-year effect.

3.8. Sample Subperiods

In this section, we examine whether our results differ over time. We do so because, given the various technological developments during our sample period (e.g., in terms of computing power and/or the internet), it is likely that investors' information acquisition costs have decreased and/or that their attention has become less constrained over that period. If this is indeed so, then the number of stocks that investors know about and follow could have increased over time. Hence, it seems sensible to expect that any information would have been disseminated across investors more incompletely during the earlier part of our sample period than during its later part. Consequently, our results should be more pronounced over the first half of that period.

To test this prediction, we repeat all of our analyzes as in Tables 3–5, but this time we split our sample period into two almost equal subperiods, one from 1980Q1 to 1998Q4 and another one from 1999Q1 to 2019Q1. Table 7 presents the raw and the risk-adjusted premium differentials over the different holding periods and these subperiods.

With respect to the trading strategies based on $\text{E/P}_{i,t}$ and $\text{ROA}_{i,t}$, regardless of the holding period, nearly all of the premium differentials are positive and statistically important at the 5% level during the period 1980Q1–1998Q4. This, however, is rarely the case over the period 1999Q1–2019Q1. These results can be most clearly seen from the FF3 and the PS4 alphas. Hence, the premiums on those strategies seem to be reliably and persistently higher among stocks with smaller $\text{InsIBS}_{i,t}^{r2}$, but considerably more so during the earlier subperiod. Over that period, when the strategy is based on $\text{ROA}_{i,t}$, for instance, the untabulated three-quarter

Table 7. Raw and risk-adjusted premium differentials over different holding and sample periods.

	Panel A: 1980Q1–1998Q4				Panel B: 1999Q1–2019Q1			
Holding period (in quarters)	Raw return	CAPM alpha	FF3 alpha	PS4 alpha	Raw return	CAPM alpha	FF3 alpha	PS4 alpha
	Strategy with $E/P_{i,t}$							
1	2.28	1.72	1.93	2.20	2.40	1.99	1.06	1.00
	(4.06)	(2.87)	(4.02)	(4.19)	(1.55)	(1.37)	(0.98)	(0.96)
2	3.89	2.45	2.32	3.40	5.18	4.18	2.07	2.03
	(2.91)	(1.86)	(2.38)	(3.16)	(1.89)	(1.56)	(1.06)	(1.03)
3	6.03	3.63	5.73	7.10	7.34	5.98	2.03	1.79
	(2.90)	(1.82)	(3.18)	(3.25)	(1.88)	(1.68)	(0.91)	(0.78)
4	7.90	5.09	6.80	9.20	11.02	9.89	1.54	1.18
	(2.96)	(2.14)	(2.71)	(3.19)	(2.03)	(2.01)	(0.61)	(0.44)
	Strategy with $ROA_{i,t}$							
1	1.74	1.47	2.12	2.47	2.00	1.36	0.61	0.61
	(2.45)	(2.12)	(3.37)	(4.23)	(1.52)	(1.08)	(0.61)	(0.58)
2	3.54	3.04	3.57	4.74	4.80	3.31	1.62	1.52
	(2.44)	(2.37)	(3.43)	(4.53)	(1.97)	(1.41)	(0.91)	(0.78)
3	5.56	5.33	7.59	9.52	7.04	5.17	2.30	2.25
	(2.56)	(2.98)	(4.01)	(4.38)	(2.04)	(1.70)	(0.97)	(0.88)
4	7.58	8.52	11.33	13.97	9.66	7.96	1.84	2.14
	(2.71)	(3.85)	(4.94)	(5.05)	(2.29)	(2.15)	(0.67)	(0.73)
	Strategy with $RET_{i,t:t-4}$							
1	0.54	− 0.19	0.78	0.94	4.79	3.79	3.70	3.88
	(0.55)	(− 0.19)	(0.79)	(0.90)	(2.98)	(2.28)	(2.92)	(2.82)
2	1.40	− 0.09	2.46	2.03	9.87	7.55	7.02	6.88
	(0.79)	(− 0.05)	(1.17)	(0.89)	(3.20)	(2.87)	(3.84)	(3.35)
3	2.67	2.09	6.50	7.74	13.29	9.33	7.71	6.63
	(1.01)	(0.86)	(1.62)	(1.76)	(2.90)	(2.64)	(3.14)	(2.74)
4	4.62	3.63	10.47	11.35	16.58	11.45	6.27	4.25
	(1.26)	(1.18)	(2.05)	(1.99)	(2.71)	(2.50)	(1.78)	(1.35)

Note: In Panel A (B), the sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 (1999Q1) to the end of 1998Q4 (2019Q1). At the end of each quarter, these stocks are first sorted into quintiles on the basis of $InsIBS^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $InsIBS^{r2}_{i,t}$, and the equal-weighted total raw returns over the next one to four quarters on these and the long-short portfolios are calculated. The presented returns (in percent) are the $E/P_{i,t}$, $ROA_{i,t}$ and $RET_{i,t:t-4}$ premium differentials (i.e., the differentials in the long-short portfolio returns) between the stocks in the extreme quintiles of $InsIBS^{r2}_{i,t}$, calculated as the time-series means of the premium differentials and as the alphas from estimating the CAPM, FF3 and PS4 models with the time series of the premium differentials. The *t*-statistics (presented in parentheses) are based on the time series of the premium differentials and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

FF3 alpha of the long-short portfolio among the stocks in the bottom quintile of $InsIBS^{r2}_{i,t}$ is 7.87% (*t*-statistic = 4.27). Yet again, this alpha monotonically decreases from the bottom to the top quintile of $InsIBS^{r2}_{i,t}$ and, among the stocks in the latter quintile, it is 0.28% (*t*-statistic = 0.14). Thus, the differential between these alphas is 7.59% (*t*-statistic = 4.01). During the later subperiod, the corresponding alphas are 1.69% (*t*-statistic = 1.45) and −0.60% (*t*-statistic = −0.26), and the differential between them is 2.30% (*t*-statistic = 0.97). For each subperiod, the magnitudes of almost all of the other three-quarter premium differentials and their *t*-statistics, including those for the strategy based on $E/P_{i,t}$, are similar.

The premium differentials regarding the trading strategy based on $\text{RET}_{i,t:t-4}$ are a bit less straightforward because they are sensitive to the holding period and the risk adjustment used. That is, over the period 1980Q1–1998Q4, when the holding period is one or two quarters, all of the premiums among the stocks in the extreme quintiles of $\text{InsIBS}_{i,t}^{r2}$ are positive and statistically reliable at the 5% level. Although nearly all of these premiums are higher among the stocks in the bottom quintile of $\text{InsIBS}_{i,t}^{r2}$ than among the stocks in its top quintile, the differentials between the premiums are not statistically notable. The major difference in these patterns when the holding period is three quarters is that the premiums are no longer statistically discernible among the stocks in the top quintile of $\text{InsIBS}_{i,t}^{r2}$. When the holding period is four quarters, however, the familiar patterns begin to emerge again and they can be most clearly seen from the FF3 and the PS4 alphas. For example, among the stocks in the bottom quintile of $\text{InsIBS}_{i,t}^{r2}$, the untabulated yearly FF3 alpha of the long-short portfolio is 5.91% (t-statistic $= 2.48$). From the bottom to the top quintile of $\text{InsIBS}_{i,t}^{r2}$, this alpha decreases and, among the stocks in the latter quintile, the alpha is not even positive; it is -4.56% (t-statistic $= -1.16$). Hence, the differential between these alphas is 10.47% (t-statistic $= 2.05$). The yearly PS4 alphas, the differential between them and their t-statistics are of a similar magnitude.

During the period 1999Q1–2019Q1, nearly all of the premium differentials with regard to the trading strategy based on $\text{RET}_{i,t:t-4}$ are positive and statistically significant at the 5% level. However, these results are not due to positive and statistically meaningful premiums among the stocks in the bottom quintile of $\text{InsIBS}_{i,t}^{r2}$. Indeed, although all of the premiums among these stocks are still positive, none of them are statistically important. Instead, the results are driven by the premiums among the stocks in the top quintile of $\text{InsIBS}_{i,t}^{r2}$, almost all of which are considerably negative and statistically unreliable. For instance, the untabulated three-quarter FF3 alphas of the long-short portfolios among the stocks in the bottom and the top quintiles of $\text{InsIBS}_{i,t}^{r2}$ are 3.55% (t-statistic $= 1.06$) and -4.16% (t-statistic $= -1.18$), respectively, which leads to the differential between these alphas of 7.71% (t-statistic $= 3.14$). In contrast to these results, the premium differentials calculated as the yearly FF3 and PS4 alphas are still positive, but they are no longer statistically notable. For example, among the stocks in the bottom and the top quintiles of $\text{InsIBS}_{i,t}^{r2}$, the untabulated yearly FF3 alphas of the long-short portfolios are 3.60% (t-statistic $= 0.88$) and -2.67% (t-statistic $= -0.58$), respectively, and the differential between them is 6.27% (t-statistic $= 1.78$). Therefore, similar to the strategies based on $\text{E/P}_{i,t}$ and $\text{ROA}_{i,t}$, the strategy based on $\text{RET}_{i,t:t-4}$ seems to also earn reliable and persistently higher premiums among stocks with smaller $\text{InsIBS}_{i,t}^{r2}$, but significantly more so during the earlier subperiod. Even though this inference is based solely on the yearly FF3 and PS4 alphas, it is important to note that, due to the overlapping holding periods and the potential risk factors considered when calculating these alphas, the tests for them are likely to be the most powerful.

Nevertheless, the demise of our results during the second half of the sample period could reflect the finding in McLean and Pontiff (2016) that, among 97 stock return anomalies, the mean anomaly is less pronounced after its original publication in the academic literature. The suggested explanation for this finding is that investors first learn about a certain anomaly from the literature and that they then trade to exploit that anomaly until it disappears. This, however, is unlikely to be the case with the anomalies studied here, which were originally published from 1977 to 1996 and for which a relatively large amount of data since then exist. Using these data, all of the anomalies that we examine have indeed been reproduced recently (see Linnainmaa & Roberts, 2018), which means that their economic and statistical significance has remained almost the same long after their original publication. Overall, thus, the results in Table 7 support our prediction with respect to the temporal variation in the statistical and the economic importance of the anomalies that appear among stocks with smaller investor bases.

3.9. Robustness Checks

3.9.1. Fama-MacBeth regressions

In this section, we seek to reproduce our results using Fama-MacBeth regressions, which enable us to analyze the returns on individual stocks (as opposed to the returns on portfolios of stocks). Therefore, for stock i at the end of quarter t, our dependent variable now is $\text{RET}_{i,t:t+q}$ (in percent), where q ranges from one to four, meaning that we study stock returns over periods of one to four quarters (without excluding any monthly returns). The independent variables involve $\ln(1+\text{InsIBS}_{i,t})$ and $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$, respectively, as well as their interaction terms. We also control for $\ln(\text{M}_{i,t})$ and $\text{InsOWN}_{i,t}$ and their own interactions with the three anomaly variables. Finally, the fixed effects of each stock exchange are controlled for too.

Before proceeding further, it is important to note that the results from estimating Fama-MacBeth regressions with multiple interaction terms can be quite sensitive to outliers. Thus, since our independent variables do not have well-behaved cross-sectional distributions, we estimate the regressions described above with the quintile ranks of those variables, which are determined at the end of each quarter and transformed so that their values range from zero to one (this transformation is performed before generating the interaction terms). The only exception here is $\ln(1+\text{InsIBS}_{i,t})$, which we use in its original form. We do so because of the high correlation between this variable and $\ln(\text{M}_{i,t})$, which makes their quintile ranks almost indistinguishable. Table 8 presents the results from these analyzes.

All of the estimated coefficients on $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$, $\text{RET}_{i,t:t-4}$ and their interaction terms with $\ln(1+\text{InsIBS}_{i,t})$ are statistically important at the 5% level. However, while the former are all positive, all of the latter are negative and smaller in magnitude. For instance, in Model 1, the coefficient on $\text{E/P}_{i,t}$ is 6.75 (t-statistic $=$ 3.79), whereas the one on the interaction term of this variable with $\ln(1+\text{InsIBS}_{i,t})$ is -1.96 (t-statistic $= -4.03$). From the bottom part of the table, we can see that the sum of these coefficients is also statistically different from zero at the 5% level (F-statistic $=$ 6.75). Note, however, that this is no longer the case in Models 3 and 4. Notwithstanding, these results suggest that, on average, among the stocks with the smallest market capitalizations and no institutional investors (for which $\ln(1+\text{InsIBS}_{i,t})$, $\ln(\text{M}_{i,t})$ and $\text{InsOWN}_{i,t}$ are jointly equal to zero), the differential between the quarterly returns on the stocks in the top and the bottom quintiles of $\text{E/P}_{i,t}$ is 6.75%. Holding all else equal, a one unit increase in $\ln(1+\text{InsIBS}_{i,t})$ is decreasing this differential by 1.96%. Importantly, the results from all of the other models are qualitatively the same. As such, they lend additional validity to our findings.

3.9.2. Types of institutional investors

Investors are likely to differ in terms of both their attention limitations and their information acquisition costs. Consequently, some investors could, on average, obtain more information than others. In turn, this could affect the incorporation of that information into the prices of stocks, depending on the type of the investors that know about and follow them. Hence, in this section, we explore if our results vary with the type of the institutional investor base size.

We do so by first estimating three different versions of Model 2 in Table 2, where $\text{InsIBS}_{i,t}$ and $\text{InsOWN}_{i,t}$ are replaced by their variants (denoted with DED, TRA and QIX for dedicated, transient and quasi-indexing institutional investors, respectively). These variables are calculated using Bushee's (1998) classification of institutional investors into those groups, which is available for the period from 1981 onward. We then repeat our analyzes as in Table 3, but now with the residuals from these regressions. The results are presented in Table 9.

Irrespective of the type of the residual institutional investor base size, all of the long-short portfolio return differentials for the trading strategies based on $\text{E/P}_{i,t}$ and $\text{ROA}_{i,t}$ are positive

Table 8. Fama-MacBeth regressions.

	$x_{i,t} = \mathrm{E/P}_{i,t}$				$x_{i,t} = \mathrm{ROA}_{i,t}$				$x_{i,t} = \mathrm{RET}_{i,t:t-4}$			
	Dependent variable: $\mathrm{RET}_{i,t:t+q}$ (in percent)											
	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 1$	$q = 2$	$q = 3$	$q = 4$
Independent variable	1	2	3	4	5	6	7	8	9	10	11	12
$x_{i,t}$	6.75	11.91	13.79	16.31	6.75	12.48	15.25	18.24	5.78	9.83	11.28	12.74
	(3.79)	(3.26)	(2.58)	(2.28)	(4.75)	(4.19)	(3.69)	(3.44)	(4.66)	(3.79)	(3.03)	(2.77)
$x_{i,t} \times \ln(1 + \mathrm{InsIBS}_{i,t})$	− 1.96	− 3.72	− 4.96	− 6.62	− 1.92	− 3.92	− 5.44	− 7.18	− 1.53	− 2.89	− 4.09	− 5.44
	(− 4.03)	(− 3.65)	(− 3.49)	(− 3.50)	(− 3.94)	(− 3.73)	(− 3.58)	(− 3.54)	(− 3.45)	(− 3.07)	(− 2.86)	(− 2.98)
$x_{i,t} \times \ln(\mathrm{M}_{i,t})$	6.94	13.26	18.83	25.45	6.26	12.40	17.63	23.43	3.23	6.31	9.98	14.14
	(4.41)	(4.68)	(4.84)	(4.94)	(4.06)	(4.01)	(3.81)	(3.78)	(2.73)	(2.79)	(2.97)	(3.22)
$x_{i,t} \times \mathrm{InsOWN}_{i,t}$	− 0.84	− 1.13	− 0.59	0.95	− 1.12	− 1.33	− 0.83	0.40	1.59	3.69	6.60	8.92
	(− 1.07)	(− 0.87)	(− 0.35)	(0.46)	(− 1.83)	(− 1.22)	(− 0.53)	(0.18)	(2.78)	(2.87)	(2.94)	(3.04)
$\ln(1 + \mathrm{InsIBS}_{i,t})$	1.73	3.35	4.78	6.57	1.64	3.34	4.86	6.66	1.59	2.94	4.16	5.63
	(4.28)	(3.98)	(4.22)	(4.54)	(4.00)	(3.81)	(3.90)	(4.12)	(4.82)	(4.30)	(4.19)	(4.33)
$\ln(\mathrm{M}_{i,t})$	− 8.98	− 17.64	− 26.10	− 35.47	− 8.52	− 16.98	− 25.02	− 33.72	− 7.46	− 14.45	− 21.44	− 28.99
	(− 5.48)	(− 5.64)	(− 5.98)	(− 6.31)	(− 5.42)	(− 5.43)	(− 5.54)	(− 5.77)	(− 6.40)	(− 6.69)	(− 7.04)	(− 7.30)
$\mathrm{InsOWN}_{i,t}$	0.13	− 0.26	− 1.28	− 2.94	0.03	− 0.55	− 1.59	− 3.15	− 1.29	− 2.99	− 5.26	− 7.33
	(0.22)	(− 0.25)	(− 0.92)	(− 1.69)	(0.07)	(− 0.59)	(− 1.29)	(− 1.91)	(− 2.92)	(− 3.27)	(− 3.72)	(− 4.03)
Constant	− 0.36	1.26	0.95	6.20	− 0.11	2.34	0.53	6.83	0.66	2.91	4.03	8.75
	(− 0.34)	(0.67)	(0.37)	(1.92)	(− 0.11)	(1.72)	(0.21)	(2.65)	(0.80)	(1.99)	(2.59)	(3.25)
Additional control variables	No	No	No	No	No	No	No	No	No	No	No	No
Stock exchange/Industry fixed effects	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
Number of quarters	156	155	154	153	156	155	154	153	156	155	154	153
Number of observations	395,733	380,805	367,914	356,222	395,733	380,805	367,914	356,222	395,733	380,805	367,914	356,222
Mean adjusted R^2 (in percent)	4.34	4.68	5.00	5.20	3.72	3.94	4.16	4.34	4.33	4.76	4.94	5.03
	F-statistics for sum of estimated coefficients											
$x_{i,t} + x_{i,t} \times \ln(1 + \mathrm{InsIBS}_{i,t})$	6.75	4.66	2.54	1.72	10.31	7.35	4.97	3.80	10.41	6.31	3.26	2.18

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. All models are first estimated at the end of each quarter. The presented coefficients are then the time-series means of the estimated coefficients. The t-statistics (presented in parentheses) are based on the time series of the estimated coefficients and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in Section 3.9.1. and the Appendix.

Table 9. Raw quarterly portfolio returns across different types of institutional investors.

	InsIBS-DED$^{r2}_{i,t}$			InsIBS-TRA$^{r2}_{i,t}$			InsIBS-QIX$^{r2}_{i,t}$		
E/P$_{i,t}$	1	5	1 − 5	1	5	1 − 5	1	5	1 − 5
1	2.74	5.86	− 3.12	2.99	6.75	− 3.76	3.63	5.93	− 2.30
5	5.28	5.46	− 0.18	5.39	6.02	− 0.63	5.25	5.46	− 0.20
5 − 1	2.54	− 0.40	2.94	2.40	− 0.73	3.13	1.62	− 0.47	2.09
	(2.82)	(− 0.48)	(4.70)	(2.70)	(− 0.72)	(3.62)	(1.71)	(− 0.49)	(2.38)
ROA$_{i,t}$									
1	2.78	5.71	− 2.93	3.09	6.48	− 3.38	3.57	5.60	− 2.03
5	4.46	4.64	− 0.18	4.74	5.35	− 0.61	5.00	4.86	0.14
5 − 1	1.68	− 1.07	2.75	1.65	− 1.13	2.77	1.43	− 0.74	2.17
	(2.18)	(− 1.43)	(4.11)	(2.63)	(− 1.22)	(3.54)	(2.18)	(− 0.80)	(2.77)
RET$_{i,t:t-4}$									
1	2.77	4.62	− 1.85	2.67	5.70	− 3.03	2.73	5.34	− 2.61
5	4.70	5.46	− 0.77	4.64	6.03	− 1.39	5.39	5.15	0.24
5 − 1	1.92	0.84	1.09	1.97	0.33	1.64	2.66	− 0.19	2.85
	(1.80)	(0.82)	(1.60)	(2.21)	(0.25)	(1.50)	(2.73)	(− 0.16)	(2.66)

t-statistics for equality of long-short portfolio return differentials

	DED vs. TRA	DED vs. QIX	TRA vs. QIX
E/P$_{i,t}$	− 0.25	1.01	2.00
ROA$_{i,t}$	− 0.03	0.69	1.08
RET$_{i,t:t-4}$	− 0.63	− 2.12	− 1.86

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1981Q4 to the end of 2019Q1. At the end of each quarter, these stocks are first sorted into quintiles on the basis of the variants of InsIBS$^{r2}_{i,t}$ (i.e., InsIBS-DED$^{r2}_{i,t}$, InsIBS-TRA$^{r2}_{i,t}$ and InsIBS-QIX$^{r2}_{i,t}$). The stocks in each of these quintiles are then sorted into quintiles on the basis of E/P$_{i,t}$, ROA$_{i,t}$ and RET$_{i,t:t-4}$, respectively. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of the variants of InsIBS$^{r2}_{i,t}$, and the equal-weighted total raw returns over the next quarter on these and the long-short portfolios are calculated. The presented returns (in percent) are the time-series means of these portfolio returns. The *t*-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

and statistically discernible at the 5% level. Regarding the strategy based on RET$_{i,t:t-4}$, the same can be said only about the one implemented with InsIBS-QIX$^{r2}_{i,t}$, even though the familiar patterns appear among the other types of institutional investors as well. Most importantly here, the *t*-statistics for the tests of the equality between the long-short portfolio return differentials (presented in the bottom part of the table) reveal no systematic patterns across the different types of institutional investors. One explanation for these results could be that the institutional investor bases of stocks are dominated by quasi-indexing institutions and that the sizes of all three types of the institutional investor bases are highly correlated (see Panels A and C in Table 1).

3.9.3. Accrual anomaly

In this section, as a placebo test, we examine probably the best-known accounting-based stock return anomaly—namely, the accrual anomaly of Sloan (1996)—which has been commonly explained in the prior literature with the overreaction of stock prices to new information. If

Table 10. Raw and risk-adjusted portfolio returns over different holding periods for the accrual anomaly.

	Raw return			CAPM alpha			FF3 alpha			PS4 alpha		
	$\text{InsIBS}^{r2}_{i,t}$			$\text{InsIBS}^{r2}_{i,t}$			$\text{InsIBS}^{r2}_{i,t}$			$\text{InsIBS}^{r2}_{i,t}$		
$\text{ACC}_{i,t}$	1	5	5 − 1	1	5	5 − 1	1	5	5 − 1	1	5	5 − 1
						Holding period: 1 quarter						
1	4.94	6.72	1.78	1.19	2.38	1.18	1.52	2.66	1.14	1.34	2.48	1.14
5	3.66	4.66	1.00	0.08	0.81	0.74	0.17	0.86	0.68	0.18	0.83	0.65
1 − 5	1.28	2.06	0.78	1.12	1.56	0.44	1.35	1.80	0.46	1.16	1.65	0.49
	(3.20)	(4.70)	(1.73)	(2.93)	(4.25)	(1.05)	(3.31)	(5.14)	(1.05)	(2.78)	(4.69)	(1.15)
						Holding period: 2 quarters						
1	9.50	12.14	2.63	1.97	3.54	1.57	2.54	3.55	1.01	2.29	3.36	1.07
5	7.56	8.97	1.41	0.73	1.52	0.79	0.69	1.38	0.69	0.55	1.33	0.78
1 − 5	1.94	3.17	1.23	1.25	2.03	0.78	1.85	2.17	0.32	1.74	2.03	0.28
	(2.48)	(3.80)	(1.27)	(1.81)	(3.31)	(1.02)	(2.29)	(3.28)	(0.43)	(2.19)	(3.11)	(0.33)
						Holding period: 3 quarters						
1	14.29	17.83	3.54	2.96	5.07	2.11	3.34	4.53	1.19	2.83	4.28	1.44
5	11.73	13.70	1.97	1.66	2.58	0.92	1.34	1.95	0.62	1.06	1.80	0.74
1 − 5	2.56	4.13	1.57	1.30	2.49	1.19	2.01	2.58	0.57	1.78	2.47	0.70
	(2.43)	(3.30)	(1.13)	(1.41)	(3.02)	(1.33)	(1.68)	(3.04)	(0.55)	(1.65)	(2.58)	(0.66)
						Holding period: 4 quarters						
1	18.73	23.41	4.68	3.90	6.78	2.88	3.90	5.08	1.18	2.94	4.46	1.52
5	15.91	18.48	2.58	2.63	3.91	1.28	2.20	2.50	0.30	1.67	2.49	0.82
1 − 5	2.82	4.93	2.10	1.26	2.87	1.61	1.70	2.59	0.89	1.27	1.97	0.70
	(2.51)	(3.36)	(1.32)	(1.35)	(2.75)	(1.45)	(1.40)	(2.20)	(0.79)	(1.20)	(1.55)	(0.54)

Note: The sample contains stocks listed on the NYSE, AMEX and NASDAQ from the end of 1980Q1 to the end of 2019Q1. At the end of each quarter, these stocks are first sorted into quintiles on the basis of $\text{InsIBS}^{r2}_{i,t}$. The stocks in each of these quintiles are then sorted into quintiles on the basis of $\text{ACC}_{i,t}$. Finally, portfolios of stocks are formed on the basis of each of these quintiles and the quintiles of $\text{InsIBS}^{r2}_{i,t}$, and the equal-weighted total raw returns over the next one to four quarters on these and the long-short portfolios are calculated. The presented returns (in percent) are the time-series means of these portfolio returns and the alphas from estimating the CAPM, FF3 and PS4 models with the time series of the portfolio returns. The *t*-statistics (presented in parentheses) are based on the time series of the portfolio returns and they are calculated using Newey-West standard errors, adjusted for heteroskedasticity and an autocorrelation with a maximum lag order of four quarters. The variable definitions are presented in the Appendix.

this explanation is correct and if our results truly reflect the underreaction to new information of the prices of stocks, then the size of their investor bases should not matter for this anomaly. Indeed, there is no reason why the degree of overreaction to accruals would vary with the size of stocks' investor bases. It is hard to imagine that the prices of neglected stocks would overreact more than the prices of well-known stocks. Likewise, in the presence of arbitrageurs, the prices of the latter should not overreact more than the prices of the former.

To test this prediction, we again repeat all of our analyzes as in Tables 3–5, but this time with the stocks' accruals (i.e., $\text{ACC}_{i,t}$) instead of $\text{E/P}_{i,t}$, $\text{ROA}_{i,t}$ and $\text{RET}_{i,t:t-4}$. Table 10 presents the results from these analyzes.

The accrual anomaly may seem to be more pronounced among the stocks in the top quintile of $\text{InsIBS}^{r2}_{i,t}$. In particular, among these stocks, except for the yearly PS4 alphas, all of the long-short portfolio returns are statistically significant at the 5% level and they are higher than those among the stocks in the bottom quintile of this variable.[10] In support of our prediction, however, none of the differentials between these returns are statistically meaningful at that level.

3.9.4. Investor base size from 10-K forms

As already noted, the Compustat variable CSHR does contain data on the number of investors in a given stock. These data, however, are not the best suited for addressing our research question, for at least two reasons. First, they are obtained from firms' 10-K forms, meaning that CSHR is unavailable at a quarterly frequency. Second, in those forms, firms do not report the number of investors who keep their shares in the accounts of their nominees, which means that the data are incomplete and CSHR could identify the size of stocks' investor bases with a large error.

Nonetheless, in this section, we use this variable instead of $\text{InsIBS}_{i,t}$. Specifically, we begin by converting it from a yearly to a quarterly frequency, where, for stock i at the end of quarter t, the new variable $\text{IBS-10K}_{i,t}$ takes the values from the most recent fiscal year-end. We then estimate a version of Model 2 in Table 2, as follows. First, we replace $\text{InsIBS}_{i,t}$ with $\text{IBS-10K}_{i,t}$ (note that the correlation of the two is 0.61). Second, we do not include the squared term of $\ln(\text{M}_{i,t})$. The reason for this is that, despite the correlation of 0.50 between $\ln(\text{M}_{i,t})$ and $\text{IBS-10K}_{i,t}$, the scatterplot of these two variables against each other, which is not presented for brevity, does not reveal any nonlinearities between them. Third, we also exclude $\text{InsOWN}_{i,t}$, since its correlation with $\text{IBS-10K}_{i,t}$ is only 0.04. Finally, we once again repeat our analyzes as in Tables 3 and 4, but now with the residuals from this model. The results from these analyzes, however, which are presented in Table OA.1 in the Online Appendix, do not cast doubt on our previous conclusions.

3.9.5. Controlling for share turnover

Despite all of the arguments in Section 3.5. that our results do not reflect an omitted variable bias related to transaction costs, we do perform additional analyzes in which we seek to control for such costs more explicitly. Particularly, our goal is to control for the share turnover of stocks, which has traditionally been thought of as being closely related to transaction costs (see, e.g., Datar et al., 1998). In this section, hence, we yet again repeat all of our analyzes as in Tables 3 and 4, but instead of using $\text{InsIBS}^{r2}_{i,t}$, we now use the residuals from Model 3 in Table 2, which includes $\text{TURN}_{i,t}$ as an independent variable. The variable containing these residuals is denoted $\text{InsIBS}^{r3}_{i,t}$. Table OA.2 in the Online Appendix presents the results from these analyzes. In sum, these results suggest that the stock return anomalies that appear among stocks with smaller investor bases continue to do so even after controlling for the share turnover of stocks. Therefore, although we cannot conclusively rule out that transaction costs contribute toward our results, it seems unlikely that they are driven by such costs.

[10]We thank the anonymous reviewer for pointing out that, in contrast to Sloan (1996), all of the alphas on the short side of the trading strategies (i.e., among the stocks in the top quintile of $\text{ACC}_{i,t}$) are positive. We trace this discrepancy to our extended sample period, which is relevant here due to the findings in Green et al. (2011) that the accrual anomaly has disappeared over time. Specifically, these authors find that the size-adjusted return to this anomaly has been 0.6% (t-statistic $= 4.5$) during the period 1970/5–1995/12, 0.5% (t-statistic $= 1.4$) over the period 1996/1–2003/12 and -0.02% (t-statistic $= -1.0$) for the period 2004/1–2010/3. Although statistically insignificant, note that the return over the latest period is negative, meaning that the short side earns higher return than the long side. Thus, to ensure that our alphas are not calculated incorrectly, we have recalculated them for the first half of our sample period (i.e., 1980Q1–1998Q4). All of the alphas on the short side are then negative, which is reassuring and helps us to reconcile the discrepancy with Sloan (1996).

3.9.6. Other robustness checks

Without any major implications for our inferences, in the Online Appendix, we also show that all our results continue to hold even after (1) including the stocks with the smallest market capitalizations (see Table OA.3), (2) including the stocks with negative book values of common equity (see Table OA.4), and (3) adding Carhart's (1997) momentum factor (MOM_t) to the FF3 and PS4 models in the trading strategies based on $E/P_{i,t}$ and $ROA_{i,t}$ (see Table OA.5).

4. Conclusion

In this paper, we provide empirical evidence that several well-documented underreaction-consistent stock return anomalies, such as the valuation anomaly of Basu (1977), the profitability anomaly of Haugen and Baker (1996) and the momentum anomaly of Jegadeesh and Titman (1993), arise and persist only among stocks with smaller (institutional) investor bases, which are presumably stocks that are neglected by investors. These results are driven by the short side of our long-short trading strategies (i.e., by the seemingly overpriced stocks in the bottom quintiles of the anomaly variables), they appear even after controlling for several stock characteristics (e.g., market capitalization, institutional ownership and share turnover) and potential risk factors, and they are considerably more pronounced during periods with more information and/or less technology. Altogether, therefore, these findings are consistent with the hypothesis that the incomplete dissemination of (negative) information across investors helps in explaining the occurrence and the persistence of the aforementioned anomalies.

Considering the failure of several recent studies, such as those of Linnainmaa and Roberts (2018) and Hou et al. (2020), to reproduce the vast majority of the stock return anomalies claimed in the prior literature, it is important to note that our findings are unlikely to be statistical artifacts. Indeed, the underreaction-consistent anomalies that we examine are chosen primarily because they are some of the very few such anomalies that have been reproduced in these studies, even with very recent data. Moreover, most of our results are statistically reliable at the 5% level even when using the critical t-value suggested by Harvey et al. (2016) of ± 2.78, which is adjusted for multiple testing. In fact, in some of our most restrictive tests, in Table 4, all of the PS4 alphas of the long-short portfolios among the stocks in the bottom quintile of $InsIBS^{r2}_{i,t}$ and almost all of the differentials between these alphas and those of the long-short portfolios among the stocks in the top quintile of this variable have t-statistics of a greater magnitude. One explanation for the cogency of these results could be that the choice of the main explanatory variable in this paper—the size of stocks' investor bases—is made on the basis of a fundamental theoretical model (i.e., the CAPM), not some empirical exercise.

Overall, apart from their implications for researchers (see Section 1), our findings have implications for both investors and firms. In particular, among stocks with smaller investor bases, there still seem to be some arbitrage opportunities for investors, in the sense that they could earn higher returns (at least before accounting for transaction costs) by taking long positions in potentially underpriced stocks and short positions in potentially overpriced stocks. Note that investors could do that at a presumably lower risk, since all of the raw quarterly long-short portfolio returns among the stocks in the bottom quintile of $InsIBS^{r2}_{i,t}$ are around 3–4% less volatile than those among the stocks in the top quintile of this variable (see the standard deviations in Panel C of Table 3). In contrast, firms could prevent a possible mispricing of their stocks by dedicating resources to the expansion of their investor bases. They could do so, for example, by advertising their stocks in the financial media and/or by listing them on multiple stock exchanges.

Acknowledgments

For their insightful comments and suggestions, we are grateful to Beatriz García Osma (editor), an anonymous reviewer, Jeffrey L. Callen, Jennifer Conrad, Adam Farago, Kristine W. Hankins, Jörgen Hellström, Robert W. Holthausen, Håkan Jankensgård, Juha-Pekka Kallunki, Michael Kisser, Ralph S. J. Koijen, John Christian Langli, Stefan Nagel, Rickard Olsson, Stephen H. Penman, Jeffrey Pontiff, Amedeo Pugliese, Kuntara Pukthuanthong, David Sraer, Tobias Svanström and Michael Ungeheuer, as well as to the seminar participants at the 2021 World Finance Conference, the 2018 Nordic Finance Network Young Scholars Workshop, BI Norwegian Business School, Stockholm School of Economics, Tilburg University, the 2018 European School of Management and Technology Berlin Annual Accounting Conference, the 2017 Swedish House of Finance Ph.D. Workshop and the 2017 University of Bolzano Emerging Researchers Consortium.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental data

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09638180.2023.2265975.

References

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*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, *32*(3), 663–682. https://doi.org/10.2307/2326304

Birru, J. (2018). Day of the week and the cross-section of returns. *Journal of Financial Economics*, *130*(1), 182–214. https://doi.org/10.1016/j.jfineco.2018.06.008

Bodnaruk, A., & Ostberg, P. (2009). Does investor recognition predict returns?. *Journal of Financial Economics*, *91*(2), 208–226. https://doi.org/10.1016/j.jfineco.2008.01.006

Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review*, *73*(3), 305–333. https://www.jstor.org/stable/248542

Calluzzo, P., Moneta, F., & Topaloglu, S. (2019). When anomalies are publicized broadly, do institutions trade accordingly?. *Management Science*, *65*(10), 4555–4574. https://doi.org/10.1287/mnsc.2018.3066

Cao, J., Han, B., & Wang, Q. (2017). Institutional investment constraints and stock prices. *Journal of Financial and Quantitative Analysis*, *52*(2), 465–489. https://doi.org/10.1017/S0022109017000102

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82. https://doi.org/10.2307/2329556

Chen, X., He, W., Tao, L., & Yu, J. (2023). Attention and underreaction-related anomalies. *Management Science*, *69*(1), 636–659. https://doi.org/10.1287/mnsc.2022.4332

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, *54*(6), 2045–2073. https://doi.org/10.1111/0022-1082.00181

Datar, V. T., Naik, N. Y., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219. https://doi.org/10.1016/S1386-4181(97)00004-9

D'Avolio, G. (2002). The market for borrowing stock. *Journal of Financial Economics*, *66*(2–3), 271–306. https://doi.org/10.1016/S0304-405X(02)00206-4

Dechow, P. M., Hutton, A. P., Meulbroek, L. K., & Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, *61*(1), 77–106. https://doi.org/10.1016/S0304-405X(01)00056-3

Duffie, D., Gârleanu, N., & Pedersen, L. H. (2002). Securities lending, shorting, and pricing. *Journal of Financial Economics*, *66*(2–3), 307–339. https://doi.org/10.1016/S0304-405X(02)00226-X

Edelen, R. M., Ince, O. S., & Kadlec, G. B. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics*, *119*(3), 472–488. https://doi.org/10.1016/j.jfineco.2016.01.002

Engelberg, J., McLean, D. R., & Pontiff, J. (2018). Anomalies and news. *The Journal of Finance*, *73*(5), 1971–2001. https://doi.org/10.1111/jofi.12718

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, *18*(3), 25–46. https://doi.org/10.1257/0895330042162430

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, *63*(4), 1653–1678. https://doi.org/10.1111/j.1540-6261.2008.01371.x

Fama, E. F., & French, K. R. (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, *29*(1), 69–103. https://doi.org/10.1093/rfs/hhv043

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636. https://doi.org/10.1086/260061

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*(3), 433–463. https://doi.org/10.1093/rof/rfn005

Gompers, P. A., & Metrick, A. (2001). Institutional investors and equity prices. *The Quarterly Journal of Economics*, *116*(1), 229–259. https://doi.org/10.1162/003355301556392

Green, J., Hand, J. R. M., & Soliman, M. T. (2011). Going, going, gone? The apparent demise of the accruals anomaly. *Management Science*, *57*(5), 797–816. https://doi.org/10.1287/mnsc.1110.1320

Griffin, J. M., & Lemmon, M. L. (2002). Book-to-market equity, distress risk, and stock returns. *The Journal of Finance*, *57*(5), 2317–2336. https://doi.org/10.1111/1540-6261.00497

Griffin, J. M., & Xu, J. (2009). How smart are the smart guys? A unique view from hedge fund stock holdings. *The Review of Financial Studies*, *22*(7), 2531–2570. https://doi.org/10.1093/rfs/hhp026

Guo, L., Li, F. W., & John Wei, K. C. (2020). Security analysts and capital market anomalies. *Journal of Financial Economics*, *137*(1), 204–230. https://doi.org/10.1016/j.jfineco.2020.01.002

Harvey, C. R., Liu, Y., & Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68. https://doi.org/10.1093/rfs/hhv059

Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, *41*(3), 401–439. https://doi.org/10.1016/0304-405X(95)00868-F

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, *55*(1), 265–295. https://doi.org/10.1111/0022-1082.00206

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, *21*(2), 109–128. https://doi.org/10.1257/jep.21.2.109

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705. https://doi.org/10.1093/rfs/hhu068

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133. https://doi.org/10.1093/rfs/hhy131

Huberman, G., & Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *The Journal of Finance*, *56*(1), 387–396. https://doi.org/10.1111/0022-1082.00330

Ivković, Z., & Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, *60*(1), 267–306. https://doi.org/10.1111/j.1540-6261.2005.00730.x

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91. https://doi.org/10.2307/2328882

Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, *12*(1), 13–32. https://doi.org/10.1016/0304-405X(83)90025-9

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2016). Return seasonalities. *The Journal of Finance*, *71*(4), 1557–1590. https://doi.org/10.1111/jofi.12398

Koski, J. L., & Pontiff, J. (1999). How are derivatives used? Evidence from the mutual fund industry. *The Journal of Finance*, *54*(2), 791–816. https://doi.org/10.1111/0022-1082.00126

Kothari, S. P., Shu, S., & Wysocki, P. D. (2009). Do managers withhold bad news?. *Journal of Accounting Research*, *47*(1), 241–276. https://doi.org/10.1111/j.1475-679X.2008.00318.x

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *The Review of Financial Studies*, *12*(5), 1113–1141. https://doi.org/10.1093/rfs/12.5.1113

Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics*, *102*(1), 62–80. https://doi.org/10.1016/j.jfineco.2011.05.012

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *The Review of Financial Studies*, *31*(7), 2606–2649. https://doi.org/10.1093/rfs/hhy030

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13–37. https://doi.org/10.2307/1924119

McLean, D. R., & Pontiff, J. (2016). Does academic research destroy stock return predictability?. *The Journal of Finance*, *71*(1), 5–32. https://doi.org/10.1111/jofi.12365

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, *42*(3), 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance*, *32*(4), 1151–1168. https://doi.org/10.2307/2326520

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*(4), 768–783. https://doi.org/10.2307/1910098

Nagel, S. (2005). Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics*, *78*(2), 277–309. https://doi.org/10.1016/j.jfineco.2004.08.008
Ng, L., & Wang, Q. (2004). Institutional trading and the turn-of-the-year effect. *Journal of Financial Economics*, *74*(2), 343–366. https://doi.org/10.1016/j.jfineco.2003.05.009
Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, *29*(1), 104–147. https://doi.org/10.1093/rfs/hhv063
Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685. https://doi.org/10.1086/374184
Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003
Perold, A. F. (2004). The capital asset pricing model. *Journal of Economic Perspectives*, *18*(3), 3–24. https://doi.org/10.1257/0895330042162340
Reinganum, M. R. (1983). The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects. *Journal of Financial Economics*, *12*(1), 89–104. https://doi.org/10.1016/0304-405X(83)90029-6
Richardson, S., Sloan, R. G., & You, H. (2012). What makes stock prices move? Fundamentals vs. investor recognition. *Financial Analysts Journal*, *68*(2), 30–50. https://doi.org/10.2469/faj.v68.n2.2
Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance*, *52*(1), 327–340. https://doi.org/10.1111/j.1540-6261.1997.tb03818.x
Shumway, T., & Warther, V. A. (1999). The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance*, *54*(6), 2361–2379. https://doi.org/10.1111/0022-1082.00192
Sias, R. W., Starks, L. T., & Titman, S. (2006). Changes in institutional ownership and stock returns: Assessment and methodology. *The Journal of Business*, *79*(6), 2869–2910. https://doi.org/10.1086/508002
Sikes, S. A. (2014). The turn-of-the-year effect and tax-loss-selling by institutional investors. *Journal of Accounting and Economics*, *57*(1), 22–42. https://doi.org/10.1016/j.jacceco.2013.12.002
Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?. *The Accounting Review*, *71*(3), 289–315. https://www.jstor.org/stable/248290
Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302. https://doi.org/10.1016/j.jfineco.2011.12.001
Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, *30*(4), 1270–1315. https://doi.org/10.1093/rfs/hhw107
Stoll, H. R., & Whaley, R. E. (1983). Transaction costs and the small firm effect. *Journal of Financial Economics*, *12*(1), 57–79. https://doi.org/10.1016/0304-405X(83)90027-2

Appendix: Variable Denotations and Definitions

For stock i at the end of quarter t, the variables are denoted and defined as follows:

(1) $\text{InsIBS}_{i,t}$: number of institutional investors with long positions (Thomson Reuters variable) at the end of quarter t. $\text{InsIBS-DED/TRA/QIX}_{i,t}$ are defined correspondingly, using the number of dedicated, transient and quasi-indexing institutional investors (Brian J. Bushee's variables).

(2) $\text{E/P}_{i,t}$: ratio of the earnings before extraordinary items and after preferred dividends (Compustat variable IBCOMQ) at the end of quarter t to $\text{M}_{i,t}$ (see item 6).

(3) $\text{ROA}_{i,t}$: ratio of the earnings before extraordinary items (Compustat variable IBQ) at the end of quarter t to the mean of the total assets (Compustat variable ATQ) at the end of quarters t and $t - 1$.

(4) $\text{RET}_{i,t:t-4}$: total return (CRSP variable RET) from the end of quarter $t-4$ to the end of quarter t. Similarly as in Jegadeesh and Titman (1993), the return over the last month in quarter t is excluded from the calculation and, in the case of a delisting, the delisting return (CRSP variable DLRET), adjusted as in Shumway (1997) and Shumway and Warther (1999), is included in it.

(5) $ACC_{i,t}$: ratio of the difference between the earnings before extraordinary items (Compustat variable IBQ) at the end of quarter t and the operating cash flows before extraordinary items (Compustat variable OANCFQ) at the end of quarter t to the mean of the total assets (Compustat variable ATQ) at the end of quarters t and $t-1$.[11]

(6) $M_{i,t}$: product of the price (CRSP variable PRC) at the end of quarter t and the number of outstanding shares (CRSP variable SHROUT) at the end of quarter t, both adjusted for splits with the cumulative adjustment factors (CRSP variables CFACPR and CFACSHR).

(7) $InsOWN_{i,t}$: ratio of the number of outstanding shares owned by institutional investors (Thomson Reuters variable) at the end of quarter t to the total number of outstanding shares (CRSP variable SHROUT) at the end of quarter t, both adjusted for splits with the cumulative adjustment factor (CRSP variable CFACSHR). $InsOWN\text{-}DED/TRA/QIX_{i,t}$ are defined correspondingly, using the number of outstanding shares owned by dedicated, transient and quasi-indexing institutional investors (Brian J. Bushee's variables).

(8) $TURN_{i,t}$: ratio of the trading volume (CRSP variable VOL, divided by two if the stock is listed on NASDAQ) in quarter t to the total number of outstanding shares (CRSP variable SHROUT) at the end of quarter t, both adjusted for splits with the cumulative adjustment factor (CRSP variable CFACSHR).

(9) $BETA_{i,t:t-4}$: coefficient on the market factor from estimating the CAPM with the monthly total returns on the stock (CRSP variable RET, adjusted as in item 4), the one-month US treasury bill and the market factor (Kenneth R. French's variables RF and MKTRF, respectively) from the end of quarter $t-4$ to the end of quarter t.

(10) $IVOL_{i,t:t-4}$: standard deviation of the residuals from estimating the CAPM as in item 9.

(11) $AGE_{i,t}$: number of calendar years from the starting date of the data from CRSP (CRSP variable BEGDAT) to the end of quarter t.

[11] The operating cash flows before extraordinary items are calculated as the difference between the change in the yearly cumulative operating cash flows (Compustat variable OANCFY) from the end of quarter $t-1$ to the end of quarter t and the extraordinary items (Compustat variable XIDOQ) at the end of quarter t. If missing, they are calculated as the sum of the earnings before extraordinary items (Compustat variable IBQ) at the end of quarter t, the depreciation and amortization (Compustat variable DPQ) at the end of quarter t and the change in the working capital (Compustat variable WCAPQ) from the end of quarter $t-1$ to the end of quarter t.