Dissecting retail, institutional, and foreign buy-sell imbalances

Pedro Piccoli[1*]

ABSTRACT

Using a unique dataset containing the daily amount traded by retail, domestic institutional and foreigner investors in the spot market of the Brazilian Stock Exchange between 2018 and 2021, I run a battery of analysis related to buy-sell imbalances of these investor's types. My main results are as follows: a) imbalances of sophisticated investors are autocorrelated, whereas retail imbalances are not; b) retail investors are not liquidity providers for sophisticated investors; c) foreign investors are trend-seekers, whereas individuals are contrarians; d) past returns are especially important with regard to forecasting imbalances of sophisticated investors; and e) retail imbalances can predict future market returns, while sophisticated imbalances cannot. Overall, these results are difficult to reconcile with the notion that retail investors are noise-traders while sophisticated investors are arbitrageurs.

Keywords: Market Microstructure, Buy-sell imbalances, Retail investors, Market Efficiency

JEL Classification: G12, G14

[1]Corresponding author. Rua Imaculada Conceição 1155, Curitiba, PR, Brazil, 80215 901.Tel.: +55 41 99262-6661. *E-mail address:* pedro.guilherme@pucpr.br

*Associate Professor of Finance, Pontifícia Universidade Católica do Paraná, Brazil

1. Introduction

The literature on market microstructure has devoted considerable effort to investigating the determinants and consequences of order imbalances. Examining its determinants sheds light on the motivation behind the investment decision, while analyzing the consequences provides insights into why prices move. These are, obviously, two central issues in finance. Since the motivation to trade and its consequences variate among investor types, studying imbalances of different investor groups would be highly recommended in order to gain a more accurate understanding of these core issues. Nevertheless, the literature on the imbalances of different investor types is extremely scarce, given the lack of data on the trades of retail investors. Even though a considerable number of studies have attempted to address this by examining the data from specific brokerage houses (Barber and Odean, 2000; Barber et al. 2009; Barber et al., 2022; Ozik et al., 2022), the generalization power of their results is limited due to clientele biases (Kingsley et al., 2014; Bohemer et al., 2021), which could be an important hurdle in light of the magnitude of the aforementioned central issues. Motivated by this lack of evidence, in this paper, I access the daily amount traded in the spot market, at the aggregated level, by retail, domestic institutional and foreign investors on the Brazilian Stock Exchange from 2018 to 2021 and run a battery of analyses regarding buy-sell imbalances (BS). My main results are as follows.

There is abundant evidence that order imbalances are highly persistent (Chordia et al., 2002; Chordia and Subrahmanyam, 2004; Tóth et al., 2015). While some papers claim that this behavior is explained by herding (LeBaron and Yamamoto, 2007) others attribute this to order splitting (Chordia et al., 2004; Lee et al., 2004) from informed investors who aim to gradually profit from private information (Kyle, 1985). To provide new insights into this debate, in the first empirical exercise, I examine the autocorrelation properties of BS of the three categories of investors previously mentioned. If the persistence is stronger in retail imbalances, then the herding explanation would prevail, since individual investors are believed to be more exposed to herding behavior (Barber et al., 2009). On the contrary, if the persistence is more observable in sophisticated trades, then the order splitting hypothesis would be more plausible, since these investors are better informed and, therefore, more inclined to gradually profit from their information. The results demonstrate that sophisticated BS are highly autocorrelated, whereas retail BS are not. This autocorrelation pattern is especially significant in the case of foreign investors, who are the main economic force in the Brazilian Stock Market, since their trades represent around 48% of the daily amount traded in this market. Consequently, these results suggest that the order splitting hypothesis is the most consistent explanation for the order flow persistence phenomenon.

In the second empirical exercise, I run a VAR analysis among BS of these three different investor types. The rationale behind this investigation is that since noise traders are believed to trade based on rumors, they are natural liquidity providers for sophisticated investors who will act as arbitrageurs, aiming to profit from betting against noise traders (Friedman, 1953; Fama, 1965). With this framework, one would expect a negative influence of retail BS on sophisticated BS, indicating that the latter are trading in the opposite direction of the former, aiming to profit from price inefficiency. However, I do not find support for this conjecture. On the contrary, the VAR analysis reveals that domestic institutional imbalances are positively dependent on past retail imbalances, indicating that these informed investors trade in the same direction as retail investors. Moreover, I also document a negative influence of foreign imbalances on future domestic institutional imbalances, suggesting that the liquidity supply in the Brazilian Market occurs among sophisticated investors.

Several studies on imbalances have investigated how price dynamics and order flows are related (Chordia et al., 2002; Lee et al., 2014; Zhang et al., 2021). In this context, I also analyze whether this relation variates depending on investor types. Surprisingly, the results show that sophisticated investors depend much more heavily on past returns for their trading decision than retail investors do. More specifically, the findings indicate that domestic institutional investors are contrarians, while foreign investors are trend-seekers, suggesting that sophisticated imbalances are inefficient in the weak form (Fama, 1970), since they are strongly dependent on past returns. In an additional investigation, I examine whether different investor type imbalances are influenced by positive market trends, here operationalized as the proportion of days with positive market returns in a 5-day moving window. The estimates demonstrate that retail imbalances are negatively related to positive trends, whereas foreign imbalances exhibit a positive association among these constructs. Since these results suggest that retail investors are contrarians, they contradict the common belief that retail investors are driving forces of market bubbles (Hott, 2009, Shiller, 2014). Counterintuitively, this role seems to be more pertinent to sophisticated investors.

Regarding the consequences of imbalances, I investigate whether BS from each investor group can predict future market returns. To this end, I run two analyses: first, a VAR approach between market returns and each type of investor's imbalances; and second, an OLS regression of returns on past imbalances. In both cases, I document that imbalances of retail investors can forecast future market returns, whereas sophisticated imbalances cannot. My explanation for this counterintuitive finding is that the information emerging from the imbalance innovations of sophisticated investors is shadowed by the autocorrelation of their orders, since they are designed to reveal information smoothly. In the case of retail imbalances, however, since their orders are not gradual, their imbalances are more informative and, consequently, more likely to influence future prices. This explanation is conceptually in keeping with the study of Li and Li (2021), who find that household belief dispersion is more informative than professional disagreement, which can be explained by the greater heterogeneity of the consumer population, which leads to richer and more diverse information. These results are also consistent with recent evidence that individual investors improve market efficiency through informativeness (Wang and Zhang, 2015; Kelley and Tetlock, 2017; Farrel et al., 2022).

The last empirical analysis made in this paper refers to the relation between imbalances of different investors and calendar effects. Several papers (e.g., Chordia et al., 2002; Lee et al., 2004; Zhang et al., 2019) have already documented an association between imbalances and the day-of-the-week effect (DoW). I add to this literature by investigating whether this association variates among investor types and by examining the association between imbalances and the turn-of-the month (ToM) effect. The rationale for including this second effect emerges from the contradictory explanation provided by the literature regarding the ToM effect. While some studies advocate that the ToM effect is explained by salary payments being made on these days, leading to increasing trades on these occasions (Ogden, 1990), others argue that this effect is driven by window dressing activity from institutional investors (Barone, 1990). Investigating which investor type imbalance is more exposed to the ToM is a logical manner to enlighten this debate. The results indicate that the association between the ToM and imbalances is more observable in retail trades, which is consistent with the salary payment hypothesis (Ogden, 1990), since the positive relation identified suggests that these investors tend to buy more on these days. Finally, referring to DoW, I document that this effect is evidenced on Tuesdays in the Brazilian stock market and is more significant in the case of retail investors, two patterns that were also observed in the Taiwanese stock market (Lee et al., 2004).

The findings of this paper contribute to at least three streams of the literature. First, in the context of market microstructure, my findings suggest that the order flow persistence is better explained by the order splitting hypothesis (Chordia et al., 2004; Lee et al., 2004), since I document that only sophisticated imbalances are autocorrelated, which is consistent with the belief that these investors slice their orders, aiming to gradually profit from private information (Kyle, 1985). Another contribution to this field refers to the dynamics among different imbalances. To the best of my knowledge, this is the first paper to investigate whether imbalances among different investor types can influence each other, which is an important step towards understanding how information flows among investors. The evidence that retail imbalances positively influence domestic institutional imbalances, and that imbalances among sophisticated investors (i.e., domestic institutional and foreign traders) are negatively related, bring an interesting implication to this field, since it contradicts the notion that information from order imbalances flows from sophisticated to individual investors.

The positive influence of retail imbalances on future institutional imbalances also makes a novel contribution to the field of market efficiency, since the classical frameworks of this field (Friedman

1953; Fama, 1965) assume that informed investors bet against noise traders to profit from the inefficiency brought by the latter, who are believed to trade based on rumors. Also under the market efficiency umbrella, the paper provides additional evidence to explain the ToM anomaly, since I document that retail investors tend to be net buyers during the first and last day of the month, behavior that is consistent with the salary payments hypothesis (Ogden, 1990) that underlies this anomaly.

Finally, the evidence in the paper also contributes to the growing literature on retail investors. The case of the collective short-squeeze on Gamestop led by individual investors, together with the rise of financial products specifically designed for this type of trader, such as Robinhood brokerage or the Seeking Alpha platform, served as a catalyst for a series of studies on the relevance of retail investors to financial markets (e.g., Barber et al., 2022; Farrel et al., 2022; Long et al., 2022; Welch, 2022). This importance was reinforced during the COVID outbreak, since the literature reports that individual investors were relevant liquidity providers during that crisis (Ozik et al., 2021). In this context, some studies claim that retail investors trade based on noise and are more exposed to behavioral biases (Barber and Odean, 2000; Kumar and Lee 2006; Barber et al. 2009; Barber et al., 2022), while others report that these investors can take efficient investment decisions (Kelley and Tetlock 2013, 2017, Boehmer et al., 2020, Farrel et al., 2022; Welch, 2022) and improve market informativeness (Wang and Zhang, 2015; Kelley and Tetlock, 2017. Li and Li, 2021). My results are consistent with the latter group of papers, since I find that retail investors are less dependent on past returns, are contrarians with regard to positive market trends, and that their imbalances are more informative than the imbalances of sophisticated investors. Since my data are at the market level, the common critique of clientele effects biasing the results emerging from retail orders (Kingsley et al., 2014; Bohemer et al., 2021, Welch, 2022) does not apply to the findings reported herein.

The remainder of the paper is structured as follows. Section 2 describes the data and methodology. Section 3 analyses the order persistence of different imbalances. Section 4 investigates the relationship between imbalances and market returns, while Section 5 examines whether this relation is influenced by different market states. Section 6 analyzes whether different imbalances can predict future market returns. Section 7 studies the relationship between imbalances and calendar effects, and Section 8 concludes the work.

2. Data and Methodology

In cooperation with the Brazilian Stock Exchange (B3),[1] I obtained a report containing the total daily amount traded by retail investors, domestic institutional investors, and foreign investors in the

[1] This is a private data that can be acquired through the B3 data service (https://www.up2dataondemand.com.br), which forbids me from sharing the data. For academic studies, B3 makes these data available at lower prices.

spot market between 01.01.2018 and 12.31.2021. These daily reports inform separately the buy and sell amounts for each category of investors. Since the reports do not detail the assets traded, the information is at the market level. The buys/sells for each trader type i are scaled by the total buys/sells on the given day to generate a stationary dataset,[2] as expressed in Eq. (1) and (2). This procedure also provides a measure of the representativeness of the trader type with regard to the total trades of the market on day t.

$$\%Buys_{i,t} = \frac{Buys_{i,t}}{Total\ Buys_t} \tag{1}$$

$$\%Sells_{i,t} = \frac{Sells_{i,t}}{Total\ Sells_t} \tag{2}$$

Consistently, the buy-sell imbalance[3] (BS) for each trader type i is defined as follows:

$$BS_{i,t} = \%Buys_{i,t} - \%Sells_{i,t} \tag{3}$$

Figure 1 shows the time-series of the total daily amount (Buys and Sells) traded by each investor type during the sample period. As expected, foreign investors are the most representative traders, followed by domestic institutions and retail investors. Even though this hierarchy is stable over time, there are occasions on which retail trades become more significant, as during the first months of the Covid outbreak (March, 2020 to July, 2020), exhibiting a slight decrease afterwards. In this respect, it is worth mentioning that the Brazilian stock market exhibited a peculiar increase in the number of retail accounts during the sample period, escalating from 619 thousand to 5.0 million, while the daily amounted traded by these investors increased from 3.8 billion BRL (0.8 billion USD) to 9.1 (2.0 billion USD) billion BRL.[4] Despite this increase, the representativeness of these investors clearly did not follow the same path, indicating that this escalation was accompanied by the increase of trades from sophisticated investors.

(Figure 1)

[2] All imbalance variables (i.e., $\%Buys$, $\%Sells$, and BS for every investor type) are stationary at the 1% level based on the augmented Dickey-Fuller test.

[3] I also employ an alternative specification for imbalances as follows: $BS_i = (Buys_{i,t} - Sells_{i,t})/(Buys_{i,t} + Sells_{i,t})$ and find very consistent results, which can be made available upon request. The correlation between these two specifications are 0.98, 0.99 and 0.99 for retail, institutional and foreign imbalances, respectively. Even though it is used in many studies, my choice not to follow this specification is due to the fact that it does not inform the representativeness of each trader type for the total amount traded in the Brazilian market, which is in line with one of the main contributions of the paper: that it examines the dynamic of the imbalances between different trader types.

[4] These amounts are averages for the first and last month of the sample period.

Table 1 shows the descriptive statistics of the main variables. On a regular day, retail investors represent around 20% of the amount traded in the Brazilian Market, what is very similar to the participation of these agents on the US market,[5] while foreign investors account for almost half of the amount traded. This representativeness is almost stable during the sample period, given the moderate size of the standard deviations. Some papers (e.g., Ozik et al., 2022; Welch, 2022) document a sharp increase in the amount traded by retail investors in the US market during the Covid-19 lockdown. This was also the case for the Brazilian stock market. In the Covid-19 lockdown window (March, 2020 to May, 2020), the average daily amount traded by these investors in the local market was more than twice the average daily amount traded before the lockdown (January, 2018 until February, 2020). Nevertheless, the participation of these investors in the market was virtually untouched, since during the lockdown window their participation in total buys (sells) was of 21% (20%). This evidence opposes the argument that the participation of retail investors in recent years has been increasing (Ozik et al., 2022). Even though in absolute terms the amount traded by retail investors has indeed risen, in relative terms its representativeness has remained stable.

Overall, the series do not exhibit fat tails, while the distributions for trades from domestic and foreign investors are more volatile (higher standard deviations and kurtosis) than for individual investors, which can be plausibly explained by the greater economic power of sophisticated investors to enter massive trading orders in the market (Farmer et al., 2013). Even though the distributions are not fat-tailed, the mean-variance relation suggests a substantial dispersion of the buy-sell imbalance distributions for all type of investors, which could potentially limit the generalization power of results emerging from parametric models. I address this concern by running a quantile regression approach and find that the main results remain robust. This alternative approach is detailed in Section 4. Finally, it is also interesting to note that, on average, retail and domestic institutional investors were net-buyers (positive mean BS) while foreign investors were net-sellers (negative mean BS) during the period under analysis. This pattern suggests that the trades occur between different investor types, a feature that will be examined more closely in the following section.

3. Autocorrelation and causation between imbalances

My first empirical analysis begins with the investigation of the autoregressive properties of the imbalances among different investor types. The literature documents that imbalances are highly persistent, but it is yet unclear whether this persistence is observable for all types of investors. In

[5] https://www.bnnbloomberg.ca/citadel-securities-says-retail-is-25-of-the-market-during-peaks-1.1463021

general, the literature has provided two main explanations (Chordia et al., 2002, 2005) for this behavior: herding (LeBaron and Yamamoto, 2007); and order splitting (Chordia et al., 2004; Lee et al., 2004; Tóth et al., 2015). Since a consensus has yet to be reached, examining the autocorrelation properties of imbalances for different investor types can shed more light on this discussion. Given that retail investors are less informed, if persistence emerges from herding, it is plausible to expect retail imbalances to be more autocorrelated vis-à-vis sophisticated (i.e., institutional and foreign) imbalances. Conversely, if endurance is due to order splitting, then we would expect higher autocorrelation from institutional and foreign orders, since these investors are more prone to segment their orders to soften the impact and gradually profit from private information (Kyle, 1985). To examine this, I run an autocorrelation test of imbalances for each investor type. Figure 2 displays the corresponding correlograms.

(Figure 2)

Even though retail imbalances exhibit a short-term autocorrelation, its persistence is clearly less pronounced than for sophisticated imbalances, suggesting that the persistence of imbalances are better explained by the order splitting hypothesis, as documented by Tóth et al. (2015). The fact that the magnitude of the autocorrelations for foreign imbalances are much higher than for domestic institutional imbalances provide further support in this respect, since foreign investors have greater economic power in the Brazilian stock market, as demonstrated in Section 2, and, consequently, would have greater incentives to trade more gradually.

The second empirical exercise of this section is a VAR analysis aiming to investigate the magnitude and direction of the influence among different imbalance types. The rationale behind running this analysis is as follows. Given that individual investors are believed to be less informed, it is plausible to expect that they will provide liquidity for sophisticated investors, who will trade to profit from private information under the axiom that for one buyer there is one seller. Furthermore, despite being widely assumed by the literature, evidence of the relation between sophisticated and individual trades is extremely scarce, given the difficulty involved in accessing trading data for different investor types at the market level. Additionally, it is also interesting to investigate the trading dynamics among distinct sophisticated investors groups (i.e., domestic and foreign), as it would provide insight into the information flow among these agents. Consequently, a VAR analysis between imbalances of different investor types represent a consistent method for examining this relation. This investigation is conducted using the following generic models, where the number of lags is determined using the AIC:

$$BS_{i,t} = \delta_0 + \delta_1 BS_{i,t-1} + \cdots + \delta_n BS_{i,t-n} + \gamma_1 BS_{k,t-1} + \cdots + \gamma_n BS_{k,t-n} + \varepsilon_t \quad (4)$$

$$BS_{k,t} = \delta_0 + \delta_1 BS_{k,t-1} + \cdots + \delta_n BS_{k,t-n} + \gamma_1 BS_{i,t-1} + \cdots + \gamma_n BS_{i,t-n} + \varepsilon_t \quad (5)$$

where the subscripts k and i refer to the pair of investor type imbalances under analysis. Since there are three groups, this investigation leads to six different models. The estimates for the parameters are presented in Table 2.

(Table 2)

In the case of retail imbalances (first two columns), the results indicate that trades from sophisticated investors exert a marginal influence on retail trades. Nonetheless, imbalances of domestic institutional investors (two middle columns), seem to be positively dependent on past imbalances of retail investors. This influence is economically meaningful since a one standard deviation of retail imbalance results in a change of 85 basis points (0.344*0.025) on next-day institutional imbalances, which is 5 times higher than the mean institutional imbalance. This result is somewhat surprising since it indicates that institutional investors trade in the same direction as retail investors, contradicting the stylized fact that individuals are liquidity providers for informed investors. The results also demonstrate that domestic institutional investors are negatively affected by past trades from foreign investors, indicating that these agents trade in opposite directions. Finally, regarding foreign imbalances (last two right columns), the results demonstrate that they are not influenced either by past retail or by past institutional imbalances. It is also worth noting that autocorrelation is more significant for both institutional and foreign imbalances, which is consistent with the previously documented order flow persistence pattern.

Overall, the results suggest that domestic institution trades depend on past imbalances from both individual and foreign investors, whereas trades by sophisticated agents exert a negligible influence on retail imbalances. The next section will delve deeper into the influence of systematic factors on imbalances.

4. Imbalances and market returns

Several studies (e.g., Chordia et al., 2002; Lee et al., 2014; Zhang et al., 2021) have analyzed the association between contemporaneous and lagged returns on buy-sell imbalances. In this section, I run a similar analysis, but focusing separately on each type of investor's imbalances. In this endeavor, my baseline model is as follows:

$$BS_{i,t} = a_0 + \sum_{n=0}^{4} b_n \cdot R_{t-n} + \sum_{n=0}^{4} c_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} d_n \cdot BS_{i,t-n} + \varepsilon_t \quad (6)$$

Where R is the market return and $IVOL$ is the volatility index for the Brazilian market based on the paper by Carr and Wu (2006) and made available by the Nefin datacenter.[6] The intuition to control for volatility emerges from the evidence that retail investors are attracted by skewness (Kumar, 2009; Barberis and Xiong, 2012), which could potentially bias the return-imbalance relation especially for this type of investor. Furthermore, albeit pertinent, the possible influence of volatility on imbalances of different investor types has been paid little attention in the literature, which is another contribution of the paper. Finally, I also run different specifications of the baseline model without controlling for volatility and excluding the contemporaneous independent variables (i.e., Return and Volatility). These alternative approaches are intended to achieve comparable results with previous studies as well as provide a sensitivity analysis among the models, minimizing the chance that the results are driven by the model specification. The results are shown in Table 3.

(Table 3)

The baseline model (middle columns of each investor type panel) indicates a negative contemporaneous relation between market return and retail investor trades, whereas the association is positive in the case of sophisticated investors. The same pattern is observable in the alternative model in which volatility and lagged imbalances are omitted (Columns (1)). This behavior reinforces the evidence that sophisticated investors are more capable of timing the market than retail investors. An interesting finding refers to the importance of past return to imbalances, since the recent performance of the market seems to be much more important to the investment decision of sophisticated investors than for individual investors. This evidence is unexpected, as it is well known that past returns make poor predictions of future performance. In this respect, it is also worth noting that institutional investors exhibit contrarian behavior, given the negative association between past returns and future imbalances, whereas foreign investors are trend seekers, based on the positive sign of this relation.

Regarding volatility,[7] the results of the baseline model suggest that past volatility is important when explaining future imbalances of retail and foreign investors, whereas they do not influence

[6] The NEFIN is the Brazilian Center for Research on Financial Economics of the University of São Paulo and it makes data sets and variables available for free download, such as Brazilian risk factors and stock portfolios. The data are updated on a regular basis. For more details, visit: www.nefin.com.br.

[7] To facilitate the interpretation, the estimates for IVOL are multiplied by 100.

future trades of domestic institutional traders. The dynamics of this influence are unclear, however, since the sign of the relation changes according to the lag under analysis. Furthermore, when the contemporaneous variables are omitted (last column of each investor type panel), past volatility ceases to be significant at the 5% level in all cases, which is in line with the seminal findings of Chordia et al. (2002), although, in their paper, the authors do not segment the imbalances by investor type due to data constraints. Overall, these results suggest that volatility plays a marginal role in the dynamics of imbalances.

As previously commented, the standard deviations of imbalances are much higher than the corresponding means, which could suggest that inferences emerging from parametric models may not satisfactorily correspond to the regular behavior of the relations under analysis. To address this, I run a quantile regression using Equation (6). To avoid heteroskedastic and autocorrelation issues, the standard errors of the coefficients were obtained using the block-bootstrap procedure. The results are summarized in Table 4, where the first row informs the quantile under analysis. For brevity, I only report the coefficients of the returns.

(Table 4)

The results indicate that under regular circumstances (i.e., 50% quantile) the contemporaneous previously documented imbalances-return relation remains observable: a negative (positive) association in the case of retail (institutional and foreign) investors that is also present in the remaining quartiles. The importance of past return to future imbalances is pervasive in the case of domestic institutional investors, which remain contrarian, whereas this relevance for foreign trades is restricted to more extreme occasions. This last pattern suggests that massive sells (buys) from foreigner investors are influenced by recent market falls (rises). Finally, the market's past performance seems to be less important for the trading decision of retail investors.

5. Imbalances and market states

This section analyses whether imbalances for different investor type are influenced by market states. This analysis is addressed in three different specifications. In the first, I investigate how positive and negative returns are associated with future and contemporaneous imbalances as specified in Equation (7):

$$BS_{i,t} = a_0 + \sum_{n=0}^{4} b_n Max(0, R_{t-n}) + \sum_{n=0}^{4} c_n Min(0, R_{t-n}) + \sum_{n=0}^{4} d_n IVOL_{t-n} + \sum_{n=1}^{4} e_n BS_{i,t-n} + \varepsilon_t \tag{7}$$

The results are presented in Table 5 (Columns (2)). For comparison purposes, I also run two alternative specifications of the above baseline model: first, without the control variables (Columns (1)) and second, without the contemporaneous positive/negative returns (Columns (3)).

(Table 5)

The estimations of the baseline parameters demonstrate a negative (positive) relation between retail (sophisticated) imbalances and contemporaneous positive returns. This behavior reinforces the view that individual investors have a poor capacity to time the market in bull markets, since they tend to sell more under these circumstances, whereas sophisticated investors are net buyers. Nevertheless, in negative markets, retail investors display a better capacity for market timing, due to their tendency to sell on these occasions, while sophisticated investors appear to be net buyers. Curiously, the magnitude of the contemporaneous return-imbalances relation is more significant for retail investors than for sophisticated investors on both occasions. Since this association is contemporaneous, one cannot attest to the direction of the influence. This point is addressed in the following section. Finally, without controlling for contemporaneous returns (Columns (3)), institutional investors seem to be more likely to trade based on past positive and negative returns than individual or foreign investors.

In the second analysis of this section, I investigate how positive and negative returns relate to the buying and selling decisions of each investor type. In this respect, I replace the buy-sell imbalance as the dependent variable in Equation (6) with the buys or sells specifications shown in Equation (1) and (2), respectively. The results are shown in Table 6.

(Table 6)

During positive markets, retail investors tend to buy less and sell more, reaffirming their poor capacity to time the market on such occasions, while during down markets they are more likely to sell stocks. In turn, sophisticated investors exhibit opposite behaviors during bull markets: while institutional domestic investors are inclined to buy, foreign investors tend to sell on these occasions. As previously documented, past returns are particularly important for the investment decision of institutional investors, but this influence is more pronounced towards the buying decision, and independent of the sign of the past returns. Regarding individual investors, they exhibit a tendency to trade more after market declines, which is in keeping with recent papers focusing on the behavior of

Robinhood users during the Covid-19 outbreak (Ozik et al., 2022; Welch, 2022). Overall, the patterns exhibited on Tables 5 and 6 show that individual traders have a poor ability to time positive markets and that domestic investors' trades are more influenced by past returns, especially with regard to their buying decisions.

The last empirical exercise of this section aims to investigate whether investors' trades are related to recent market trends. Even though the previous baseline models controlled for past returns, these variables do not necessarily mean market trends, since the previous returns could exhibit different signs. This is, however, an important question since it could indicate which investor types, if any, are more influenced by market trends, a debate that remains open, since the literature has reported mixed results. Whereas some studies have found that individual investors trade based on recent performance (Barber et al., 2009; Bauer et al., 2009) others have documented that they are contrarians, while institutional investors are trend-seekers (Choi et al., 2015; Chau et al., 2016; Chen et al., 2016).

To examine this, I create a variable that captures recent positive trends based on the sign of the market returns of a 5-day moving window. More concretely, I introduce a dummy variable that equals 1 when the market exhibits a positive return. The trend variable is given by the average of the dummy variables of the moving window, as demonstrated in Equation (8)

$$TREND_t = \frac{\sum_{n=0}^{4} \mathrm{D}(R_{t-n} > 0)}{5} \tag{8}$$

To illustrate, if 4 out of 5 days of the moving window exhibited a positive return, the TREND variable equals 0.8 on the given day. This operationalization means that this variable captures positive trends, which is particularly interesting when it comes to investigating the influence of each investor type on market bubbles.[8] I then regress the imbalances on market trends using the following model:[9]

$$BS_{i,t} = a_0 + b \cdot TREND_t + \sum_{n=0}^{4} c_n \cdot R_{t-n} + \sum_{n=0}^{4} d_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} e_n \cdot BS_{i,t-n} + \varepsilon_t \tag{9}$$

The estimates are shown in Table 7, which also includes an alternative specification of Equation (9) without controlling for market returns to provide a sensitivity analysis (Columns (1)). The results indicate the strong contrarian behavior of individual investors towards market trends, while foreign imbalances are positively dependent on recent trends. The negative relation between institutional

[8] By the way that the variable is constructed, it indirectly captures the influence of negative trends on imbalances. In this case, the negative trends (NEG) would be given by $NEG_t = 1 - TREND_t$. Consequently, the relation between imbalances and negative trends are the coefficients of TREND with opposite signs.

[9] The time series of TREND measured by the augmented Dickey-Fuller test are stationary at the 1% level (Z(t) = -11.88).

imbalances and market trends ceases to be observable when controlled for past returns, suggesting that these investors are less influenced by trends. Taken together, the results indicate that foreign investors are trend-seekers, while retail traders are contrarians, contradicting the stylized fact that individual investors are a driving force of market bubbles. The results actually suggest that this role is assumed by foreign investors.

(Table 7)

As previously mentioned, the first two analyses of this section indicate a strong association between imbalances and contemporaneous returns for all investor types. Consequently, it is unclear whether imbalances can predict future returns. This point is addressed in the following section.

6. Predictability of market returns

In the previous sections I document that past returns are important drivers for the futures trades of some investor types. Consequently, it is reasonable to investigate the imbalance-returns relation in both directions. To this end, I employ the following VAR models:

$$R_t = \delta_0 + \delta_1 BS_{i,t-1} + \cdots + \delta_n BS_{i,t-n} + \gamma_1 R_{t-1} + \cdots + \gamma_n R_{t-n} + \varepsilon_t \quad (10)$$

$$BS_{i,t} = \delta_0 + \delta_1 BS_{i,t-1} + \cdots + \delta_n BS_{i,t-n} + \gamma_1 R_{t-1} + \cdots + \gamma_n R_{t-n} + \varepsilon_t \quad (11)$$

Where the subscript i refers to the different investor's types. The results are presented in Table 8. I begin by examining the predictability of returns. The estimates demonstrate that imbalances from retail investors have the most significant impact on future market returns. In this case, it is curious to note that this influence exhibits opposite signs depending on the lag under analysis. The negative influence of retail imbalances on next-day returns reinforces the evidence that these investors have poor market timing skills. In the case of institutional investors, I document a negative influence of imbalances on future market returns that is restricted to the second lag, whereas no impact stems from foreign imbalances. In their model, Chordia et al. (2004) predict a positive relation between imbalances and future returns that is caused by the autocorrelation of imbalances, since this autocorrelation is due to order splitting. These results do not support this hypothesis, since I do not document a positive influence of lagged imbalances on price in the case of sophisticated investors, whose imbalances are autocorrelated.

(Table 8)

Regarding the dependence of imbalances, consistent with the behaviors documented in the previous sections, retail trades are not dependent on previous returns, whereas institutional (foreign) investors exhibit contrarian (trend-seeker) behavior.

In the second empirical exercise, I analyze the influence of past imbalances on future returns by employing the following OLS model:

$$R_{i,t} = a_0 + \sum_{n=1}^{4} b_n \cdot BS_{i,t-n} + \sum_{n=1}^{4} c_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} d_n \cdot R_{t-n} + \varepsilon_t \tag{12}$$

The estimates are presented in Table 9. The results demonstrate that past imbalances from retail investors can predict future returns even when controlling for market volatility. This capacity, however, is not observable in the case of sophisticated investors. In their seminal work, Chordia et al. (2002) find no evidence that past imbalances can predict future returns. Since sophisticated investors dominate the market, our results are partially consistent with their evidence. However, when segmenting the imbalances by investor type, I document that retail imbalances can forecast future returns. My explanation for this puzzling finding is as follows. Since sophisticated imbalances are autocorrelated due to order splitting, the price pressure created by innovation in these investors' trades is negligible, since a portion of the information released from changes in imbalances is shadowed by this correlation pattern. However, since retail investors do not split their orders, their imbalance innovations are more informational, exerting a larger influence on future prices. These results are consistent with the recent findings of Farrel et al. (2022), who document that aggregated retail order imbalances predict future returns, which suggests that retail trades reveal fundamental information. This indicates that retail investors are important drivers of market efficiency, which is in line with a recent stream of the literature, advocating that individual traders improve efficiency (e.g., Kelley and Tetlock, 2013; Vozlyublennaia, 2014; Wang and Zhang, 2015; Kelley and Tetlock, 2017).

(Table 9)

7. Imbalance types and calendar effects

In this last empirical exercise, I examine the relation between imbalances and calendar effects, aiming to investigate which investor type (if any) is more sensitive to this kind of anomaly. The anomalies examined here are the day-of-the-week and the turn-of-the-month effects. To this end, for the first anomaly, I introduce into Equation (6) a dummy variable that equals 1 when the given trading day falls on the day of the week in question, whereas for the second anomaly, the dummy equals 1 for the first and the last trading days of the given month. The results are shown in Table 10. For brevity, I only report the estimates for the calendar variables.

(Table 10)

In the table, the first panel (Rt) shows the estimation of the model, using the market return as the dependent variable to investigate the existence of both calendar effects in the Brazilian stock market. My main focus, however, is on the remaining columns. The results indicate a day-of-the-week effect on the Brazilian market occurring on Tuesdays. This effect is driven by retail investors, who tend to buy on this day, given the positive relation between the dummy and the imbalances of these investors. For sophisticated traders, however, this relation is not significant on Tuesdays. It is also worth noting that foreign investors exhibit a tendency to buy on Mondays. However, this behavior does not appear to be influential, since the overall market does not exhibit a day-of-the-week effect on Mondays, which can be attributed to the low level of significance of the relation (t-stat = 1.66). These results are consistent with the study of Lee et al. (2004), who also document a day-of-the-week effect in imbalances of the Taiwanese Stock Market occurring on Tuesdays, which is restricted to individual investors' trades. However, unlike the present study, they document a negative sign of the effect, indicating that local retail investors tend to sell on this day.

Regarding the turn-of-the-month effect, individual investors are inclined to buy on both the first and the last day of the month, what is in line with the hypothesis that this effect can be linked to salaries being paid on these days (Ogden, 1990). The fact that domestic institutional investors do not exhibit any trading pattern on these occasions is additional evidence in this respect, as it is plausible to expect that the salary payment explanation would only be observable in the case of individual investors. Finally, foreign investors tend to sell on the last trading day of the month, which could suggest that these investors tend to close their positions on these occasions.

Under the lens of market efficiency, the results emerging from the calendar effect analysis suggest that these anomalies can be largely explained by the influence of individual investors. Nevertheless, foreign investors also play an important role in this phenomenon, mainly in the case of the turn-of-the-month effect.

8. Conclusion

In this paper, I access a unique dataset made available by the Brazilian Stock Exchange that informs the daily amount traded by retail, domestic institutional and foreign investors at the aggregated spot market level from 01.01.2018 to 12.31.2021. Based on these data, I run a battery of empirical tests regarding the buy-sell imbalances of these different investor types. The main results are as follows. First, the imbalances of sophisticated investors are autocorrelated, whereas retail imbalances are not, suggesting that the capital flow persistence documented by the market microstructure literature is explained by order splitting by sophisticated investors who aim to smoothly profit from private information. I also document that domestic institutional imbalances are positively dependent on past retail imbalances, contradicting the notion that the former commonly bet against the latter, as expected in classical arbitrage models. Moreover, the paper documents that past returns are important determinants of the future trades of sophisticated investors, but less important determinants of retail imbalances. In the specific case of positive market trends, the results indicate that foreign investors are trend-seekers, whereas individuals are contrarians. Taken together, these results are difficult to reconcile with the notion that retail investors trade based on noise (i.e., past returns, price trends or herding towards informed investors).

I also investigate whether different imbalances can predict future returns, and find that this predictability property can be attributed exclusively to retail imbalances, which is in line with a recent stream of papers claiming that individual investors are important agents of market efficiency. Finally, I document that the turn-of-the-month effect can be plausibly explained by salary payments that are clustered on these days, since I report that retail investors tend to buy on these occasions.

For practitioners and regulators, the important message that emerges from these results is the relevance of individual investors to stock markets. Given how important they are to market informativeness, the launch and improvement of financial tools that facilitate their participation in the market through lower transaction costs and that democratize their investment research should be strongly encouraged.

References

Barber, B. M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55(2), 773–806

Barber, B. M., Odean, T., Zhu, N., 2009. Do retail traders move the market? *Review of Financial Studies*, 22(1), 151-186

Barber, B., Huang, X., Odean, T., Schwarz, C., 2022. Attention-Induced Trading and Returns: Evidence from Robinhood Users. *Journal of Finance*, 77(6), 3141-3190

Barberis, N., Xiong, W., 2012. Realization utility. *Journal of Financial Economics*, 104, 251-271

Barone, E., 1990. The Italian stock market: efficiency and calendar anomalies. *Journal of Banking and Finance*, 14(2), 483–510.

Bauer, R., Cosemans, M., Eichholtz, P., 2009. Option trading and individual investor performance. *Journal of Banking and Finance,* 33. 731-746

Bohemer, E., Jones, C., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance*, 76(5), 2249-2305

Carr, P., Wu, L., 2006. A Tale of Two Indices. *Journal of Derivatives*, 13(3), 13-29

Chau F., Deesomsak, R., Koutmos, D, 2016. Does investor sentiment really matter? *International Review of Financial Analysis*, 48, 221-232

Chen, H., Cow, E., Shiu, C., 2016. The informational role of individual investors in stock pricing: Evidence from large individual and small retail investors. *Pacific-Basin Finance Journal*, 31, 36-56

Choi, J., Kedar-Levy, H., Yoo, S., 2015. Are individual or institutional investors the agents of bubbles? *Journal of International Money and Finance*, 59, 1-22

Chordia, T., 2004. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 65, 111-130

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, 76, 271-292

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, 76, 271-292

Fama, E., 1965. The Behavior of Stock Market Prices. *Journal of Business* 38(1), 34-105

Fama. E., 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25(2), 383-417

Farmer, J., Gerig, A., Lillo, F., Waelbroeck, H., 2013. How efficiency shapes market impact. *Quantitative Finance,* 13(11), 1743-1758

Kingsley., Y., Gallagher, D., Lee, A., 2014. Individual investors and broker types. *Journal of Financial and Quantitative Analysis*, 49(2), 431-451

Farrell, M., Green, T., Jame, R., Markov, S., 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, 145, 616-641

Friedman, M., 1953. The Case for Flexible Exchange Rates. *In Essays in Positive Economics*. University of Chicago Press: Chicago

Hott, C., 2009. Herding behavior in asset markets. *Journal of Financial Stability*, 5, 35-56

Kelley, E., Tetlock, P., 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance*, 68(3), 1229-1265

Kelley, E., Tetlock, P., 2017. Retail short selling and stock prices. *Review of Financial Studies*, 30(3), 801-834

Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance*, 64(4), 1889-1933

Kumar, A., Lee, C. M. C., 2006. Retail investor sentiment and return Comovements. *Journal of Finance*, 61(5), 2451-2486

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

LeBaron, B., Yamamoto, R., 2007. Long-memory in an order-driven market. *Physica A: Statistical Mechanics and its Applications*, 383(1), 85–89.

Lee Y., Liu, Y., Roll, R., Subrahmanyam, A., 2004. Order Imbalances and Market Efficiency: Evidence from the Taiwan Stock Exchange. *Journal of Financial and Quantitative Finance*, 39(2), 327-341

Li, D., Li, G., 2021. Whose Disagreement Matters? Household Belief Dispersion and Stock Trading Volume. *Review of Finance*, 25(6), 1859-1900

Long, S., Lucey, B., Xie, Y., Yarovaya, L., 2022, I Just Like the Stock": The Role of Reddit Sentiment in the GameStop Share Rally. *Financial Review.* In Press.

Ogden, J., 1990. Turn-of-month evaluations of liquid profits and stock returns: a common explanation for the monthly and January effects. *Journal of Finance* 45(4), 1259–1272

Ozik, G., Sadka, R., Shen, S., 2021. Flattening the Illiquidity Curve: Retail Trading during the COVID-19 Lockdown. *Journal of Financial and Quantitative Analysis*, 56(7), 2356-2388

Shiller, R., 2014. Speculative asset prices, *American Economic Review*, 104(6), 1486-1517

Tóth, B., Palit, I., Lillo, F., Farmer, J., 2015. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51, 218-239

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. *Journal of Banking and Finance*, 41, 17-35

Welch, I., 2022. The Wisdom of the Robinhood Crowd. *Journal of Finance,* 78(3), 1489-1527

Wang, Q., Zhang, J., 2015. Does individual investor trading impact firm valuation? *Journal of Corporate Finance*, 35, 120-135

Zhang, T., Gu, G., Zhou, W., 2019. Order imbalances and market efficiency: New evidence from the Chinese stock market. *Emerging Markets Review*, 38, 458-467

Zhang, T., Jiang, G., Zhou, W., 2021. Order imbalance and stock returns: New evidence from the Chinese stock market. *Accounting and Finance*, 61, 2809-2836

Table 1: Descriptive statistics

The table shows the summary statistics of the buys, sells and buy-sell imbalances for individual, domestic institutional and foreign investors in the spot market of the Brazilian stock market. %Buys (%Sells) are defined as the daily amount bought (sold) by the given investor type, divided by the total buys (sells) on the given day. The imbalances (BS) are the differences between %Buys and %Sells. The data are daily, from 01.01.2018 to 12.31.2021.

	%Buys			%Sells			BS		
	Individual	Institutional	Foreign	Individual	Institutional	Foreign	Individual	Institutional	Foreign
Mean	0.201	0.317	0.473	0.197	0.315	0.479	0.004	0.002	-0.005
Median	0.199	0.309	0.476	0.195	0.308	0.477	0.005	0.001	-0.005
Stdev	0.041	0.054	0.047	0.040	0.050	0.049	0.025	0.031	0.036
Kurtosis	1.060	1.321	1.584	1.150	2.584	1.993	0.173	0.264	0.455
Asymmetry	-0.176	0.848	0.037	-0.227	1.088	0.099	-0.017	0.168	-0.115

Table 2: VAR analysis of imbalances of different investor types

This table shows the estimates for the VAR models between the imbalances of individual (BSRet), institutional (BSIns) and Foreign (BSFor) investors. The number of lags is based on the AIC. See Section 2 for a detailed description of the construction of these variables. Standard errors are reported below the coefficients. The R^2 values are expressed as percentages. The data are daily from 01.01.2018 to 12.31.2021. N = 976

	$BSRet_t$		$BSIns_t$		$BSFor_t$	
$BSRet_{t-1}$	0.000	-0.043	0.344***			-0.019
	(0.034)	(0.038)	(0.038)			(0.047)
$BSRet_{t-2}$	0.060	0.052	0.119***			-0.042
	(0.035)	(0.038)	(0.040)			(0.047)
$BSRet_{t-3}$	0.077**	0.028	0.051			-0.016
	(0.036)	(0.038)	(0.040)			(0.046)
$BSRet_{t-4}$	0.001*	0.019	0.058			0.008
	(0.036)	(0.037)	(0.040)			(0.046)
$BSIns_{t-1}$	0.024		0.279***	0.000	0.042	
	(0.030)		(0.034)	(0.038)	(0.041)	
$BSIns_{t-2}$	-0.001		0.102***	0.033	0.025	
	(0.031)		(0.035)	(0.038)	(0.041)	
$BSIns_{t-3}$	0.058*		0.057	0.022	-0.011	
	(0.031)		(0.035)	(0.038)	(0.041)	
$BSIns_{t-4}$	-0.025		0.096***	0.060	-0.005	
	(0.029)		(0.033)	(0.037)	(0.041)	
$BSFor_{t-1}$		-0.048*		-0.306***	0.375***	0.345***
		(0.030)		(0.035)	(0.038)	(0.037)
$BSFor_{t-2}$		0.000		-0.065*	0.130***	0.096**
		(0.032)		(0.037)	(0.040)	(0.039)
$BSFor_{t-3}$		-0.033		-0.038	0.111***	0.111***
		(0.032)		(0.037)	(0.040)	(0.039)
$BSFor_{t-4}$		0.011		-0.024	0.060	0.070
		(0.029)		(0.036)	(0.040)	(0.036)
Intercept	0.004***	0.004***	-0.002*	-0.001	-0.002**	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R^2	1.5	2.2	22.1	21.5	26.1	27.0

***Significant at the 1%, **5%, and *10% levels.

Table 3: Imbalances and market returns

This table shows the estimates for the regressions between imbalances of different investor types (BS_i) and market returns (R) using the following OLS model:

$$BS_{i,t} = a_0 + \sum_{n=0}^{4} b_n \cdot R_{t-n} + \sum_{n=0}^{4} c_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} d_n \cdot BS_{i,t-n} + \varepsilon_t \tag{6}$$

where $IVOL$ is the idiosyncratic volatility of the Brazilian Stock Market and was obtained from the NEFIN (Brazilian Center for Research in Financial Economics of the University of São Paulo) database. To facilitate the interpretation, the estimates for $IVOL$ are multiplied by 100. The table also contains two alternative specifications of the above model to provide a sensitivity analysis. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=976.

	Retail			Institutional			Foreign		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
R_t	-0.967***	-1.059***		0.492***	0.500***		0.560***	0.583***	
	(-6.64)	(-8.08)		(7.83)	(7.22)		(4.54)	(4.81)	
R_{t-1}	-0.060	-0.146	0.132*	-0.293***	-0.447***	-0.550***	0.399***	0.268***	0.124*
	(-0.93)	(-1.51)	(1.71)	(-3.80)	(-5.43)	(-7.32)	(5.13)	(3.77)	(1.71)
R_{t-2}	0.111*	0.306***	0.083	-0.278***	-0.289***	-0.222***	0.158	0.011	0.138**
	(1.83)	(3.64)	(1.10)	(-4.43)	(-5.09)	(-2.96)	(1.64)	(0.14)	(2.28)
R_{t-3}	-0.049	0.020	0.157**	-0.186***	-0.127**	-0.174***	0.259***	0.105*	0.075
	(-1.24)	(0.32)	(2.46)	(-2.87)	(-2.12)	(-2.82)	(3.56)	(1.86)	(1.22)
R_{t-4}	0.033	0.058	0.041	-0.139**	-0.114**	-0.081	0.131*	-0.038	-0.002
	(0.89)	(1.15)	(0.66)	(-2.43)	(-2.12)	(-1.35)	(1.85)	(-0.75)	(-0.04)
$IVOL_t$		-0.032			0.001			0.000	
		(-1.37)			(0.03)			(1.23)	
$IVOL_{t-1}$		-0.033	0.030*		-0.023	-0.062***		0.050**	0.021
		(-1.38)	(1.77)		(-1.01)	(-3.29)		(2.14)	(1.25)
$IVOL_{t-2}$		0.061***	-0.002		-0.020	0.009		-0.045***	-0.001
		(3.46)	(-0.09)		(-0.93)	(0.50)		(-2.59)	(-0.06)
$IVOL_{t-3}$		0.006	0.003		0.014	0.009		0.009	0.006
		(0.36)	(0.13)		(0.80)	(0.50)		(0.43)	(0.3)
$IVOL_{t-4}$		0.012	-0.021		0.010	0.028		-0.034	-0.012
		(0.91)	(-1.1)		(0.46)	(1.53)		(-1.55)	(-0.46)
BS_{t-1}		-0.040	0.041		0.271***	0.286***		0.318***	-0.197***
		(-0.74)	(0.87)		(7.49)	(7.70)		(9.10)	(-3.54)

BS_{t-2}		0.211***	0.086*		0.109***	0.090**		0.125***	0.210***
		(4.87)	(1.90)		(2.94)	(2.32)		(3.46)	(5.58)
BS_{t-3}		0.042	0.118***		0.072**	0.084**		0.098***	0.148***
		(1.24)	(2.83)		(2.15)	(2.39)		(3.04)	(4.58)
BS_{t-4}		0.031	0.035		0.112***	0.095***		0.066**	0.112***
		(0.87)	(0.78)		(3.59)	(2.89)		(2.27)	(3.44)
Intercept	0.005***	0.000	0.001	0.002	0.006**	0.005**	-0.006***	-0.004*	-0.006*
	(6.86)	(-0.01)	(0.28)	(1.46)	(2.32)	(2.07)	(-3.76)	(-1.69)	(-1.81)
adj-R^2	46.0	50.1	1.2	13.8	29.2	22.1	12.4	32.8	19.1

***Significant at the 1%, **5%, and *10% levels

Table 4: Quantile regression of imbalances on returns

This table shows the estimates for the quantile regressions between the imbalances of different investor types (BS_i) and market returns (R) using the following model:

$$BS_{i,t} = a_0 + \sum_{n=0}^{4} b_n \cdot R_{t-n} + \sum_{n=0}^{4} c_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} d_n \cdot BS_{i,t-n} + \varepsilon_t \qquad (6)$$

where $IVOL$ is the idiosyncratic volatility of the Brazilian Stock Market and was obtained from the NEFIN (Brazilian Center for Research in Financial Economics of the University of São Paulo) database. For brevity, the table reports only the estimates for market returns. The corresponding quantiles are shown in the first row. The numbers in parentheses are blocked bootstrap *t*-statistics. The sample period is from 01.01.2018 to 12.31.2021. The R^2 values are expressed as percentages. N=976.

	q05	q10	q25	q50	q75	q90	q95
Retail Buy-Sell Imbalance							
R_t	-0.994***	-1.146***	-1.238***	-1.306***	-1.304***	-1.341***	-1.344***
	(-4.43)	(-11.53)	(-16.03)	(-18.74)	(-17.31)	(-11.16)	(-12.80)
R_{t-1}	0.038	-0.194*	-0.188**	-0.032	-0.030	-0.142	-0.108
	(0.21)	(-1.82)	(-2.05)	(-0.39)	(-0.27)	(-0.82)	(-0.56)
R_{t-2}	0.474*	0.417***	0.259*	0.183***	0.139*	0.209	0.206
	(1.67)	(3.79)	(1.73)	(2.69)	(1.75)	(1.56)	(1.24)
R_{t-3}	0.109	0.191	0.057	0.054	-0.068	-0.225	-0.435***
	(0.59)	(1.28)	(0.67)	(0.52)	(-1.15)	(-1.06)	(-2.42)
R_{t-4}	0.004	-0.005	0.153	0.062	0.057	0.070	0.186*
	(0.02)	(-0.05)	(1.25)	(0.99)	(0.66)	(0.74)	(1.67)
Intercept	-0.022***	-0.017***	-0.007*	0.000	0.007**	0.009**	0.009
	(-3.94)	(-4.94)	(-1.79)	(0.15)	(2.09)	(2.34)	(0.02)
R2	43.3	48.3	49.6	48.8	48.5	46.8	42.2
Controls	YES	YES	YES	YES	YES	YES	YES
Institutional Buy-Sell Imbalance							
R_t	0.364***	0.489***	0.506***	0.568***	0.580***	0.511***	0.540***
	(5.10)	(3.41)	(7.46)	(9.02)	(7.62)	(7.31)	(7.83)
R_{t-1}	-0.348***	-0.408***	-0.469***	-0.466***	-0.560***	-0.663***	-0.669***
	(-3.60)	(-2.81)	(-4.72)	(-5.87)	(-5.76)	(-6.32)	(-5.4)
R_{t-2}	-0.427***	-0.367**	-0.369***	-0.362***	-0.301***	-0.247**	-0.173*
	(-4.16)	(-2.49)	(-4.78)	(-5.34)	(-2.88)	(-2.09)	(-1.87)
R_{t-3}	-0.070	-0.320**	-0.209***	-0.137*	-0.226**	-0.137**	-0.127
	(-0.63)	(-2.51)	(-3.86)	(-1.78)	(-2.33)	(-2.10)	(-0.76)
R_{t-4}	-0.176*	-0.086	-0.074	-0.028	-0.152	-0.282***	-0.440***
	(-1.99)	(-0.64)	(-0.85)	(-0.39)	(-1.25)	(-3.28)	(-3.40)
Intercept	-0.040***	-0.031***	-0.012***	0.005	0.022***	0.038***	0.047***
	(-7.36)	(-4.80)	(-3.76)	(1.27)	(5.73)	(13.7)	(10.76)
R2	26.1	28	29.4	29.8	29.7	28.4	27.7
Controls	YES	YES	YES	YES	YES	YES	YES
Foreign Buy-Sell Imbalance							
R_t	0.724***	0.707***	0.652***	0.632***	0.581***	0.567***	0.536***
	(6.47)	(8.49)	(6.39)	(5.12)	(5.12)	(7.42)	(5.32)
R_{t-1}	0.410***	0.406***	0.250***	0.183**	0.247	0.402***	0.403***
	(4.39)	(4.75)	(2.65)	(2.01)	(1.64)	(3.97)	(3.42)
R_{t-2}	0.028	0.121	0.060	0.092	0.172	-0.060	-0.166
	(0.26)	(1.08)	(0.85)	(0.81)	(1.46)	(-0.45)	(-1.2)

R_{t-3}	0.200**	0.258***	0.151	0.094	0.221	0.084	0.023
	(2.29)	(3.10)	(1.18)	(0.94)	(1.62)	(0.71)	(0.25)
R_{t-4}	0.076	0.021	-0.075	-0.066	-0.031	0.118	0.172
	(1.12)	(0.26)	(-1.38)	(-0.98)	(-0.3)	(1.37)	(1.33)
Intercept	-0.058***	-0.045***	-0.021***	0.001	0.016**	0.035***	0.043***
	(-15.17)	(-11.14)	(-5.82)	(0.24)	(2.23)	(9.69)	(7.94)
R2	31.5	32.3	33.4	32.9	32.2	30.3	25.4
Controls	YES	YES	YES	YES	YES	YES	YES

***Significant at the 1%, **5%, and *10% levels

Table 5: Imbalances and market states

This table shows the estimates for the regression of imbalances on positive ($Max(0,R)$) and negative ($Min(0,R)$) market returns using the following model:

$$BS_{i,t} = a_0 + \sum_{n=0}^{4} b_n Max(0, R_{t-n}) + \sum_{n=0}^{4} c_n Min(0, R_{t-n}) + \sum_{n=0}^{4} d_n IVOL_{t-n} + \sum_{n=1}^{4} e_n BS_{i,t-n} + \varepsilon_t \quad (7)$$

where $IVOL$ is the idiosyncratic volatility of the Brazilian Stock Market and was obtained from the NEFIN (Brazilian Center for Research in Financial Economics of the University of São Paulo) database. For brevity, the table reports only the estimates for market returns. The table also contains two alternative specifications of the above model to provide a sensitivity analysis. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=976.

	Retail			Institutional			Foreign		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$Max(0,R_t)$	-1.328***	-1.487***		0.635***	0.749***		0.784***	0.845***	
	(-9.47)	(-12.80)		(5.05)	(5.89)		(6.08)	(7.66)	
$Max(0,R_{t-1})$	0.250**	0.141	0.156***	-0.493***	-0.666***	-0.642***	0.248	0.076	0.093
	(2.15)	(1.52)	(3.17)	(-4.78)	(-6.00)	(-5.17)	(1.45)	(0.57)	(0.78)
$Max(0,R_{t-2})$	-0.146	-0.011	0.131	-0.109	-0.016	-0.156	0.253*	0.163	0.026
	(-1.34)	(-0.10)	(1.00)	(-0.91)	(-0.15)	(-1.40)	(1.66)	(1.32)	(0.23)
$Max(0,R_{t-3})$	0.075	0.061	0.355	-0.287**	-0.238*	-0.333***	0.217	0.034	-0.047
	(0.75)	(0.65)	(1.46)	(-2.32)	(-1.92)	(-2.97)	(1.52)	(0.29)	(-0.40)
$Max(0,R_{t-4})$	0.021	-0.025	0.106	-0.125	0.009	-0.057	0.127	-0.030	-0.100
	(0.26)	(-0.29)	(0.27)	(-1.20)	(0.08)	(-0.50)	(1.14)	(-0.28)	(-0.89)
$Min(0,R_t)$	-0.759***	-0.790***		0.427***	0.375***		0.417***	0.429***	
	(-5.02)	(-5.74)		(4.75)	(4.40)		(2.32)	(2.83)	
$Min(0,R_{t-1})$	-0.299***	-0.311**	0.155	-0.155	-0.293***	-0.517***	0.528***	0.423***	0.150*
	(-3.66)	(-2.42)	(0.08)	(-1.57)	(-2.70)	(-4.63)	(5.46)	(4.07)	(1.92)
$Min(0,R_{t-2})$	0.271**	0.427***	0.025	-0.402***	-0.441***	-0.249*	0.112	-0.071	0.111
	(2.05)	(3.25)	(0.24)	(-5.09)	(-5.42)	(-1.93)	(0.69)	(-0.48)	(1.29)
$Min(0,R_{t-3})$	-0.166	-0.058	0.008	-0.110	-0.054	-0.063	0.304***	0.179	0.143
	(-1.36)	(-0.46)	(0.90)	(-1.10)	(-0.46)	(-0.69)	(2.65)	(1.61)	(1.51)
$Min(0,R_{t-4})$	-0.028	0.021	0.020	-0.124	-0.140	-0.106	0.174	0.000	0.021
	(-0.38)	(0.26)	(-0.2)	(-1.28)	(-1.63)	(-1.15)	(1.56)	(0.00)	(0.22)
Intercept	0.006***	-0.001	0.002	0.002	0.007***	0.004*	-0.007**	0.000	-0.006**
	(3.31)	(-0.57)	(1.01)	(0.92)	(2.74)	(1.81)	(-2.05)	(1.09)	(-2.25)

adj-R2	49.1	53.3	1.5	14.1	30.1	22.1	12.5	33.2	25.5
Controls	NO	YES	YES	NO	YES	YES	NO	YES	YES

***Significant at the 1%, **5%, and *10% levels

Table 6: Buys and sells of different investor types and market states

This table shows the estimates for the regression of buys/sells on positive ($Max(0, R)$) and negative ($Min(0, R)$) using Equation (7) but replacing the dependent variable with investors' buys (sells), defined as the daily amount bought (sold) by the given investor type, divided by the total buys (sells) on the given day. For brevity, the table reports only the estimates for market returns. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=976.

	Retail		Institutional		Foreign	
	Buys	Sells	Buys	Sells	Buys	Sells
$Max(0,R_t)$	-0.718***	0.764***	0.747***	0.021	0.002	-0.850***
	(-4.05)	(5.73)	(4.28)	(0.13)	(0.02)	(-5.18)
$Max(0,R_{t-1})$	0.314**	-0.092	-0.506***	0.142	0.199*	0.048
	(2.43)	(-0.65)	(-3.67)	(0.95)	(1.54)	(0.28)
$Max(0,R_{t-2})$	0.112	0.135	-0.272*	-0.173	0.174	-0.027
	(0.84)	(1.11)	(-1.70)	(-1.19)	(1.13)	(-0.15)
$Max(0,R_{t-3})$	0.205	-0.041	-0.311*	0.086	-0.005	-0.004
	(1.48)	(-0.29)	(-1.73)	(0.44)	(-0.03)	(-0.02)
$Max(0,R_{t-4})$	-0.089	-0.154	0.261	0.234	-0.175	-0.155
	(-0.67)	(-1.25)	(1.64)	(1.40)	(-1.04)	(-0.85)
$Min(0,R_t)$	-0.161	0.606***	-0.039	-0.441***	0.250*	-0.170
	(-1.34)	(3.31)	(-0.30)	(-3.36)	(1.78)	(-0.73)
$Min(0,R_{t-1})$	0.264***	0.450***	0.196	0.375***	-0.450***	-0.888***
	(2.12)	(4.80)	(1.51)	(2.69)	(-3.76)	(-6.34)
$Min(0,R_{t-2})$	0.609***	0.115	-0.507***	-0.030	-0.042	-0.025
	(4.49)	(0.64)	(-3.03)	(-0.18)	(-0.23)	(-0.10)
$Min(0,R_{t-3})$	0.143	0.184	-0.255	-0.166	0.114	-0.050
	(0.93)	(1.79)	(-1.64)	(-1.19)	(0.94)	(-0.34)
$Min(0,R_{t-4})$	0.116	0.025	-0.218	-0.199	0.088	0.089
	(1.07)	(0.22)	(-1.43)	(-1.34)	(0.69)	(0.64)
Intercept	0.059***	0.064***	0.106***	0.128***	0.152***	0.214***
	(7.74)	(8.37)	(8.89)	(9.32)	(7.80)	(10.03)
adj-R2	33.0	34.7	38.3	27.0	26.4	21.5
Controls	YES	YES	YES	YES	YES	YES

***Significant at the 1%, **5%, and *10% levels

Table 7: Imbalances and market trends

This table shows the estimates for the regression of imbalances on market trends using the following model:

$$BS_{i,t} = a_0 + b \cdot TREND_t + \sum_{n=0}^{4} c_n \cdot R_{t-n} + \sum_{n=0}^{4} d_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} e_n \cdot BS_{i,t-n} + \varepsilon_t \quad (9)$$

where $TREND$ is the proportion of days with positive market returns in a 5-days moving window. For brevity, the table reports only the estimates for the variable of interest and for market returns. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N = 976.

	Retail		Institutional		Foreign	
	(1)	(2)	(1)	(2)	(1)	(2)
$TREND_t$	-0.042***	-0.015***	-0.010**	-0.003	0.029***	0.014**
	(-9.69)	(-2.72)	(-2.06)	(-0.59)	(5.91)	(2.01)
R_t		-0.997***		0.514***		0.526***
		(-7.13)		(7.49)		(4.05)
R_{t-1}		-0.084		-0.432***		0.206**
		(-0.90)		(-4.89)		(2.84)
R_{t-2}		0.344***		-0.277***		-0.037
		(4.10)		(-4.53)		(-0.47)
R_{t-3}		0.063		-0.115*		0.054
		(1.00)		(-1.81)		(0.87)
R_{t-4}		0.094*		-0.104*		-0.084
		(1.95)		(-1.91)		(-1.64)
Intercept	0.024***	0.007***	0.008**	0.007**	-0.020***	-0.011***
	(7.30)	(2.35)	(2.38)	(2.09)	(-4.90)	(-2.57)
adj-R2	10.8	50.8	13.2	29.2	28.4	33.1
Controls	YES	YES	YES	YES	YES	YES

***Significant at the 1%, **5%, and *10% levels

Table 8: VAR analysis between imbalances and market returns

This table shows the estimates for the VAR models between imbalances (BS) and market returns (R). The number of lags is based on the AIC. Standard errors are reported below the coefficients. The data are daily from 01.01.2018 to 12.31.2021. The R^2 values are expressed as percentages. N = 976

	Retail		Institutional		Foreign	
	BS_t	R_t	BS_t	R_t	BS_t	R_t
BS_{t-1}	0.037	-0.058**	0.290***	0.027	0.334***	0.023
	(0.044)	(0.031)	(0.033)	(0.021)	(0.033)	(0.019)
BS_{t-2}	0.076*	0.139***	0.089**	-0.041**	0.107***	-0.039
	(0.044)	(0.031)	(0.035)	(0.022)	(0.035)	(0.020)
BS_{t-3}	0.121***	-0.081	0.084**	0.023	0.113***	0.030
	(0.045)	(0.031)	(0.034)	(0.021)	(0.035)	(0.020)
BS_{t-4}	0.041	-0.010	0.098***	-0.030	0.073**	0.012
	(0.045)	(0.031)	(0.033)	(0.020)	(0.033)	(0.019)
R_{t-1}	0.100	-0.170***	-0.495***	-0.147***	0.121**	-0.148***
	(0.064)	(0.044)	(0.054)	(0.034)	(0.059)	(0.034)
R_{t-2}	0.035	0.289***	-0.155***	0.186***	0.052	0.172***
	(0.064)	(0.045)	(0.056)	(0.035)	(0.060)	(0.034)
R_{t-3}	0.137**	-0.131***	-0.141**	-0.083**	0.040	-0.073**
	(0.064)	(0.045)	(0.057)	(0.035)	(0.060)	(0.034)
R_{t-4}	0.041	0.000	-0.074	0.050	-0.031	0.016
	(0.064)	(0.044)	(0.056)	(0.035)	(0.058)	(0.033)
Intercept	0.003***	0.000	0.001	0.000	-0.002**	0.000
	(0.001)	(0.550)	(1.230)	(0.580)	(0.001)	(0.780)
R2	1.9	10.2	22.1	7.7	27.2	5.6

***Significant at the 1%, **5%, and *10% levels

Table 9: Influence of imbalances on future returns

This table shows the estimates for the regressions of market returns on past imbalances of different investor types (BS_i) using the following OLS model:

$$R_{i,t} = a_0 + \sum_{n=1}^{4} b_n \cdot BS_{i,t-n} + \sum_{n=1}^{4} c_n \cdot IVOL_{t-n} + \sum_{n=1}^{4} d_n \cdot R_{t-n} + \varepsilon_t \qquad (12)$$

where $IVOL$ is the idiosyncratic volatility of the Brazilian Stock Market and was obtained from the NEFIN (Brazilian Center for Research in Financial Economics of the University of São Paulo) database. To facilitate the interpretation, the estimates for $IVOL$ are multiplied by 100. The table also contains an alternative specification of the above model to provide a sensitivity analysis. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are t-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=976.

	Retail		Institutional		Foreign	
	(1)	(2)	(1)	(2)	(1)	(2)
BS_{t-1}	-0.058	-0.072	0.027	0.029	0.023	0.025
	(-1.02)	(-1.38)	(0.93)	(1.14)	(0.99)	(1.10)
BS_{t-2}	0.139*	0.120*	-0.041	-0.037	-0.039	-0.035
	(1.93)	(1.82)	(-1.46)	(-1.43)	(-1.28)	(-1.29)
BS_{t-3}	-0.081***	-0.076**	0.023	0.024	0.030	0.027
	(-2.34)	(-2.29)	(0.93)	(1.03)	(1.41)	(1.28)
BS_{t-4}	-0.010	-0.007	-0.030	-0.035	0.012	0.014
	(-0.24)	(-0.17)	(-1.28)	(-1.53)	(0.55)	(0.64)
$IVOL_{t-1}$		-0.076**		-0.079**		-0.079**
		(-2.44)		(-2.38)		(-2.34)
$IVOL_{t-2}$		0.057**		0.058**		0.057**
		(2.54)		(2.29)		(2.29)
$IVOL_{t-3}$		-0.004		-0.009		-0.007
		(-0.20)		(-0.51)		(-0.37)
$IVOL_{t-4}$		0.029		0.036		0.035
		(1.32)		(1.59)		(1.54)
R_{t-1}	-0.010*	-0.240**	-0.147*	-0.205**	-0.148*	-0.206**
	(-1.70)	(-2.32)	(-1.92)	(-2.58)	(-1.91)	(-2.48)
R_{t-2}	0.289*	0.225	0.186	0.135	0.172	0.120
	(1.76)	(1.57)	(1.45)	(1.29)	(1.39)	(1.18)
R_{t-3}	-0.131*	-0.134**	-0.083*	-0.094**	-0.073	-0.085**
	(-1.84)	(-2.14)	(-1.75)	(-2.32)	(-1.52)	(-2.03)
R_{t-4}	0.000	0.016	0.050	0.065	0.016	0.027
	(-0.01)	(0.23)	(0.75)	(1.11)	(0.29)	(0.52)
Intercept	0.000	-0.001	0.000	-0.001	0.000	-0.001
	(0.54)	(-0.49)	(0.55)	(-0.53)	(0.78)	(-0.46)
adj-R2	9.5	13.2	7.0	11.1	7.0	11.0

***Significant at the 1%, **5%, and *10% levels

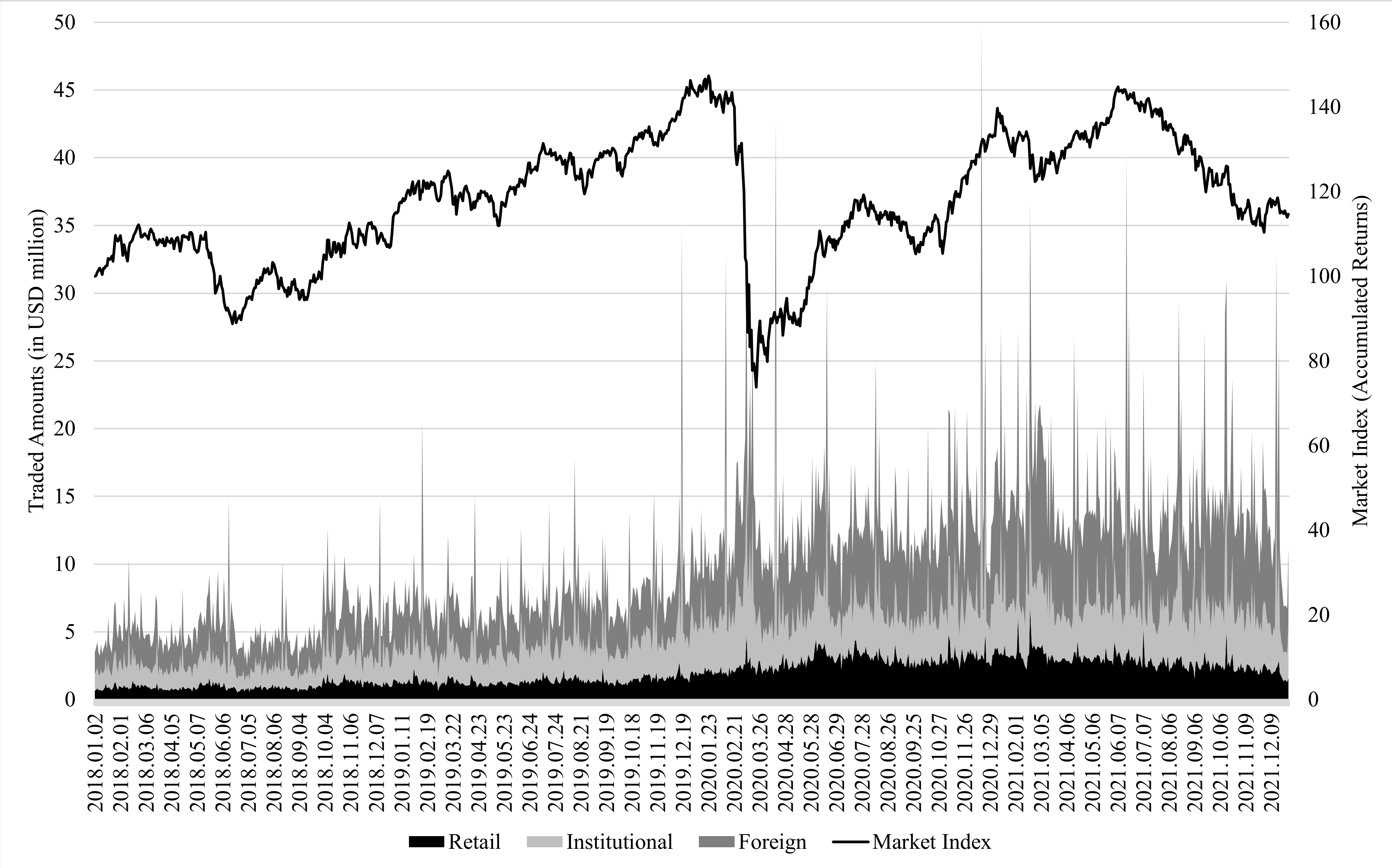
Table 10: Imbalances and calendar effects.

This table shows the estimates of the relation between imbalances of different investor types and two calendar effects: day-of-the-week effect (Model 1) and turn-of-the-month effect (Model 2). For the first, I include in Equation (6) a dummy variable that that equals 1 when the given trading day falls on the day of the week in question, whereas for the second, the dummy equals 1 for the first (month beginning) and the last (month end) trading days of the given month. In the table, the first left Panel (Rt) shows the estimation of the model employing the market return as the dependent variable to investigate the existence of both calendar effects in the Brazilian stock market. For brevity, I only report the estimates for the calendar variables. The sample period is from 01.01.2018 to 12.31.2021. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=976.

	R_t		Retail		Institutional		Foreign	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
MON	0.0001		-0.0025		-0.0015		0.0052*	
	(0.06)		(-1.30)		(-0.57)		(1.66)	
TUE	0.0032*		0.0062***		-0.0018		-0.0046	
	(1.80)		(4.02)		(-0.67)		(-1.51)	
WED	0.0012		0.0010		0.0005		0.0004	
	(0.68)		(0.55)		(0.22)		(0.15)	
THU	0.0014		0.0025		-0.0001		-0.0009	
	(0.71)		(1.25)		(-0.05)		(-0.31)	
Month Beginning		0.0025		0.005*		0.0029		-0.0054
		(1.19)		(1.73)		(0.72)		(-1.19)
Month End		-0.0042*		0.005**		0.0048		-0.0111**
		(-1.71)		(2.07)		(1.16)		(-2.43)
Intercept	-0.0002	0.0010	-0.0015	0.000	0.0064**	0.0055**	0.2616***	-0.0035
	(-0.10)	(0.64)	(-0.85)	(-0.30)	(2.09)	(2.19)	(3.77)	(-1.40)
adj-R2	19.0	18.9	51.2	50.2	29.0	29.2	33.3	33.2
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Lagged BS	YES	YES	YES	YES	YES	YES	YES	YES

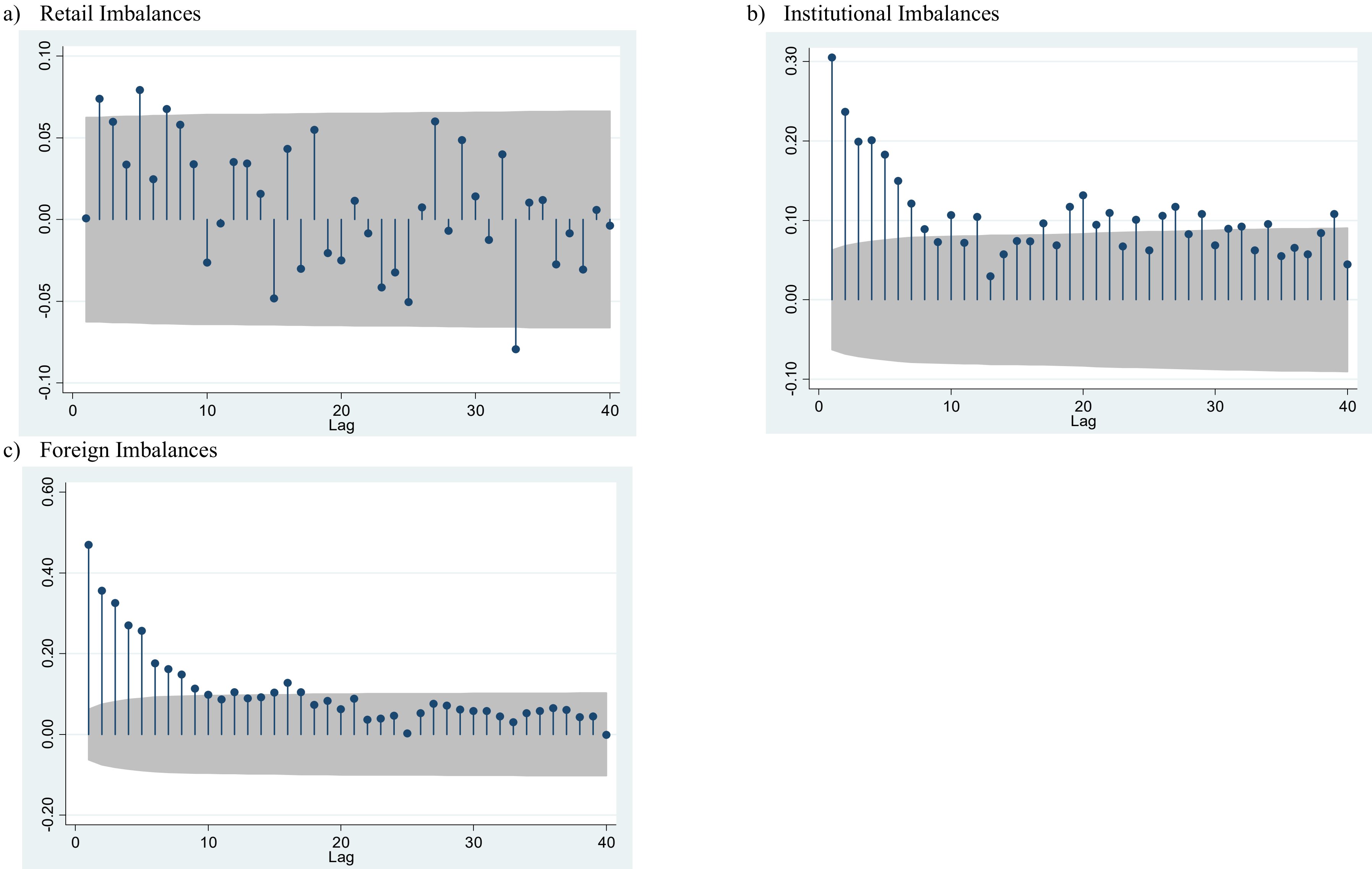
***Significant at the 1%, **5%, and *10% levels

Figure 1: Daily amount traded by different investor types and the cumulative returns of the Brazilian market index



Note: The figure shows the daily amount traded (buys plus sells) by retail, domestic institutional and foreign investors in the Brazilian Stock Market (left axis), together with the accumulated returns of the local market index (right axis, solid line). To facilitate interpretation, the amounts were transformed into USD using the average exchange rate for the sample period (i.e., 4.538 USD/BRL). The data are daily from 01.01.2018 to 12.31.2021.

Figure 2: Autocorrelation of imbalances for different investor types



The figure shows the correlograms of imbalances for retail, domestic institutional, and foreign investors. The Bartlett's confidence interval of 95% is represented by the gray area. The data are daily from 01.01.2018 to 12.31.2021.