WILEY | Hindawi

Research Article

Modeling Investor Behavior Using Machine Learning: Mean-Reversion and Momentum Trading Strategies

Thiago Christiano Silva [ID],[1,2] **Benjamin Miranda Tabak** [ID],[3] **and Idamar Magalhães Ferreira**[3]

[1]*Universidade Católica de Brasília, Distrito Federal, Brazil*
[2]*Department of Computing and Mathematics, Faculty of Philosophy, Sciences, and Literatures in Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil*
[3]*FGV/EPPG Escola de Políticas Públicas e Governo, Fundação Getúlio Vargas, School of Public Policy and Government, Getulio Vargas Foundation, Distrito Federal, Brazil*

Correspondence should be addressed to Benjamin Miranda Tabak; benjaminm.tabak@gmail.com

We model investor behavior by training machine learning techniques with financial data comprising more than 13,000 investors of a large bank in Brazil over 2016 to 2018. We take high-frequency data on every sell or buy operation of these investors on a daily basis, allowing us to fully track these investment decisions over time. We then analyze whether these investment changes correlate with the IBOVESPA index. We find that investors decide their investment strategies using recent past price changes. There is some degree of heterogeneity in investment decisions. Overall, we find evidence of mean-reverting investment strategies. We also find evidence that female investors and higher academic degree have a less pronounced mean-reverting strategy behavior comparatively to male investors and those with lower academic degree. Finally, this paper provides a general methodological approach to mitigate potential biases arising from *ad-hoc* design decisions of discarding or introducing variables in empirical econometrics. For that, we use feature selection techniques from machine learning to identify relevant variables in an objective and concise way.

1. Introduction

This paper studies the determinants of investors' behavior in the stock market using transaction-level data on buy and sell operations of investors. Our data contains detailed information of the investor's identity and her socioeconomic characteristics, the investment value, and variation due to the buy or sell operation over time. The data is confidential and comes from a large and representative Brazilian bank. With this rich dataset, we are able to study how investors respond to changes in the Brazilian stock market due to variations of its market index, called IBOVESPA. We use historical variations of the IBOVESPA index with different horizons (window length) to test which one better predicts the investors' behavior.

To mitigate potential concerns due to subjective decisions by the analyst—and also to prevent discarding a potentially relevant predictor—we opt to use an objective approach to identify those horizons that best explain investors' buy or sell operations. For that, we use a robust feature selection technique borrowed from the machine learning literature called elastic net. The great advantage of the elastic net comes by the simplicity of its loss function (just like a regression) and also the robustness in preventing overfitting by optimally using a convex combination of the Lasso and Ridge regularization methods. Overfitting can occur as the algorithm may learn the dynamics of the variable of interest and fit very well the training dataset but with poor predictability in other datasets. Evaluating the potential for overfitting is essential for researchers as it may undermine the model. We understand that our method seeks to avoid, to some extent, the perils of overfitting. The Ridge and the Lasso algorithms impose penalties for large weights in the model [1]. In this way, they tend to reduce the

model's complexity and hence are able to minimize concerns about overfitting.

Investors tend to trade using different strategies, such as buy-and-hold (passive strategy) or an active strategy in which they seek to outperform a benchmark, for example, a market index. If investors trade using active strategies, they may use two different and well-known approaches, a mean-reverting or a momentum strategy. See [2–4]; for seminal contributions in these two strategies.

In the first case, they react to market swings by betting that the market will mean-revert. Therefore, they assume the trend will change and therefore will sell after substantial upward changes and buy after downward changes. In the momentum strategy, they will bet that the trend will persist. Thus, they will increase investments in the stock market after an increase in the market index.

While we understand that they may be other trading strategies, we focus on the mean-reversion and momentum strategies because they are well established in the literature and serve as building blocks of many other strategies. There is a large body of the literature that discusses their use in different contexts [5–11]. In addition, they are easily testable in empirical specifications. Therefore, we seek to understand if investors decide to hold their stocks or sell them after negative/positive shocks.

The issue of how investors will behave on average is empirical. There is a large body of the literature that discusses predictability for the stock market [12–16]. In addition, there is another strand focusing on cognitive biases and excess trading on equity and other financial assets [17–23]. Our data containing transaction-level operations of buy and sell operation permit us to follow each investors' decisions over time and therefore test whether they use mean-reverting or momentum strategy after changes in the stock market index.

It is essential to notice that, if traders use such strategies, they may induce higher volatility in the market with their actions. In theory, market changes should occur as new information arrives, which is economically relevant to estimate future profits and dividend distribution. However, price substantially changes over time and volatility is higher than we would expect in a rational market. Therefore, we assume that the traders' decision to trade excessively will induce higher volatility in the market. Investors' decisions that follow different trading strategies may generate complex patterns in prices and volatility. They may induce long-range correlation, short-term predictability, and chaotic dynamics in prices over time. There is a large body of the literature that attempts to explain complex macrobehavior of systems using a composition of local rules. For that, agent-based modelling has been extensively used to explain price and volatility using artificial markets [24]. Using agent-based modelling, LeBaron [25] explores structural (macro) features that emerge in a market where participants adapt and evolve over time, while Bertella et al. [26] study the effect of investor's behavioral bias in prices. Understanding how investors behave and perform trading strategies is the first step for better understanding the complexity that is intrinsic to financial markets. Our paper also contributes to this matter.

To identify the most relevant predictors that explain investors' behavior, we depart from using traditional panel-data econometric techniques and goodness-of-fit measures and instead employ more robust methodologies borrowed from the machine learning literature. Contrasting to usual econometrics techniques that summarize relationships using linear regression analysis, machine learning offers a set of tools that can potentially capture nonlinear relationships between the data. According to Varian [27], bridging the gap between machine learning and econometrics is a natural tendency mainly because of the presence of large amount of data and the rising complexity—potentially highly nonlinear—between data relationships. Our work contributes to this endeavour by providing a real case study of a financial dataset using machine learning techniques.

Comparatively to econometrics, machine learning techniques have strong model selection techniques, mainly through the use of cross-validation techniques, which are a type of repeated resampling in random subsets of the dataset. Initially, the cross-validation procedure divides the data set into two disjoint and complete subsets: the training set and the test set. All the model's parameters are tuned using only the training set. After the model is selected (tuned) using the training data, we run it against the test set to check its accuracy or some other performance metric. The rationale is that, by training the model with some data and testing against another subset, we are estimating the model's out-of-sample prediction power and not simply learning the data. The test set therefore would be a simulation of real (production) data and the model's performance on this dataset would represent a rough estimate of actual performance of the model in real unseen data.

Since our data set comprises more than 350,000 observations representing individual investor's movements with respect to their investments over 2016 to 2018, we apply regularization techniques to prevent model's overfitting during the feature selection procedure with training data. For that, we apply an elastic net procedure [1] to control for the model's complexity. Elastic net is a generalization of the Ridge (-norm) and Lasso (L_1-norm) and hence is more robust. It uses an optimal convex combination of both types of regularization. Lasso tends to shrink the majority of the nonrelevant regressors to zero while keeping only the most important regressors as nonzero. In contrast, Ridge tends to output nonzero coefficients for almost all regressors. By using both regularization schemes, we are able to enjoy the positive characteristics of both schemes.

Regularization is an important issue in large data sets because it prevents methods with high variance and low bias from overfitting [28, 29]. This is the well-known bias-variance trade off in the machine learning literature [30]. While low bias prevents overfitting, it can generate underfitting of the data set. In contrast, high-variance methods can learn noise from the data and let go the true relationships of the data set. Low bias favors low model complexity at the cost of a potential overfitting. High variance tends to successfully capture smoothly nonlinear relationships between the data at the expense of a potential overfitting. Examples of low-bias algorithms are the linear regression or neural networks

with a single layer. Examples of high-variance algorithms are decision trees and multilayer neural networks. It is important to first set the rationale behind the regularization process from the viewpoint of our financial data set of buy and sell operations. On the one hand, a strand in the economics literature advocates that the agents' decisions are completely rational, in that decisions are taken by considering all information from the market (complete information) [31]. On the other hand, another body of the literature argues that investors cannot potentially consider every single information from the market when taking their decisions because (i) the agent does not have complete information and (ii) even if the agent did have complete information, she would be unable to perform all required calculations. In this way, they would naturally focus on the most relevant variables. In this case, we say that investors have a bounded rationality, term first coined by Simon [32]. We can frame these two theories into the two types of regularization frameworks used in this paper. Investors with unbounded rationality, i.e., that consider all potential variables, would better be modeled by a Ridge regularization procedure because it does not tend to place zero importance on any variables. In contrast, investors with bounded rationality would be better modeled by a Lasso regularization because it would choose a few (and more relevant) variables and set the remainder as zero. By using a weighted convex combination of both Ridge and Lasso regularization procedures, we are effectively considering both cases in our estimation process.

While Brazil does not have well-developed stock markets as advanced economies, it is an important emerging country that, due to its size and relative importance to its peers, deserves to be studied. In addition, capital markets have been increasing in the last years (according to the BM&FBovespa, which is the Brazilian stock exchange, the number of investors increased almost 20% from 2017 to 2018.), which strengthen the relevance of our work. Our main results suggest that investors use a mean-reverting trading strategy. Therefore, they reduce their investments after positive changes in the IBOVESPA and increase it after negative changes.

We also test whether investors' biological and socioeconomic characteristics explain their trading behavior. In terms of schooling, educated investors, in theory, should behave in more rational ways and trade less frequently when there is no new information arriving continuously in the market, at least those that are not relevant regarding potential for future profits. Therefore, we would expect these investors to have a smaller reaction to price fluctuations. We also test dissimilarities in investment decision making arising from the gender. Neyse et al. [33] and Lundeberg et al. [34] partly attribute investment differences among males and females due to systematic changes in overconfidence. Excessive overconfidence is associated with higher levels of testosterone, which is more pronounced in males. Overconfidence may induce investors to take on higher risks, leading them to look for higher returns in the short term. In this way, we would expect a less sensitive behavior of females to changes of past IBOVESPA variations as they would value more fundamentals and look for yields in the longer term. Our empirical analyses corroborate these views.

Several papers have studied investor behavior. Onishchenko and Ülkü [35] show a change in foreign investors, which have become more sophisticated. They find that foreign investors in Korea do not chase returns as the previous literature normally reports. Their results suggest a transition from positive to negative feedback trading over time. Abreu [36] finds that investors that buy warrants have specific characteristics, such as young age and less educated, or investors with gambling attitudes (and overconfidence) (see also [37–41]). To the best of our knowledge, our paper is one of the first that uses machine-learning techniques to unveil what are the characteristics that matter the most for explaining investor behavior at the disaggregated level. We study the reaction of investors to market changes and test whether they employ momentum or mean-reverting strategies.

2. Data

We collect and match several unique proprietary and public datasets. Our sample consists of public information from the IBOVESPA index, investor-specific information, and a proprietary customer database from a large Brazilian bank with investor-specific matched daily transactions on buy and sell operations in the IBOVESPA stock exchange market. The last two datasets are confidential.

The first source is the IBOVESPA index of the Brazilian stock exchange (BM & FBovespa). This index is considered the stock market benchmark for Brazil. We have 747 days in our sample spanning over the years of 2016 to 2018.

The second source is the investor registration information, such as their profession, degree of education, and equity. Information is from the database of the home broker and customer relationship management (CRM) solution. Our data set is comprehensive and encompasses 13,634 investors.

The last source provides each transaction made by each investor, on BM & FBOVESPA and on each of the days between January 2, 2016, and December 31, 2018, as well as their daily holdings. We observe their daily trading activities on investment decisions. This rich data set enables us to keep track of investors' buy and sell operations over time and therefore permits us to test whether they use the mean-reverting trading strategy or the momentum trading strategy as a response to IBOVESPA index changes. These are two common trading strategies that have been discussed in the literature [42, 43]. Other strategies exist, which may be more complex in nature and difficult to model, and they are not the object of our analysis. One such example would be rational traders that employ fundamental analysis and forecast future profits of traded companies to estimate their potential to distribute dividends and can value these stocks. The sample has 1,099,985 trading decisions (change in the investment volume). We also have 358,176 customer holdings over time.

Table 1 reports summary statistics of our data on investor's daily decisions on their investments. We can see that

TABLE 1: Summary statistics of our panel data on investors' daily trading decisions over the period of 2016 to 2018.

Statistic	N	Mean	St. Dev.	Min	Pctl (25)	Median	Pctl (75)	Max
Investment variation (%)	356,172	9.267	68.389	−100.000	−6.680	0.510	8.770	499.879
1-day IBOVESPA variation	356,172	0.145	1.528	−4.870	−0.740	0.110	1.010	6.600
2-day IBOVESPA variation	355,796	0.323	2.160	−6.550	−1.140	0.290	1.600	9.130
3-day IBOVESPA variation	355,419	0.472	2.598	−7.950	−1.140	0.540	2.020	10.880
5-day IBOVESPA variation	354,588	0.781	3.274	−8.250	−1.230	0.770	2.750	16.870
30-day IBOVESPA variation	343,592	4.863	8.282	−19.060	−0.740	5.240	10.500	28.770
IBOVESPA index	356,176	0.145	1.528	−4.870	−0.740	0.110	1.010	6.600

there is a large range of daily investment variations, going from −100% to almost 500%. On average, we see a positive investment variation (9.267%). We also show the IBOVESPA index level and its variations in the last 1, 2, 3, 5, and 30 days. We will use these IBOVESPA index changes to check how they correlate with the investment variations variable. One underlying hypothesis is that investors look at the IBOVESPA index to decide on their trading decisions.

Figures 1(a) and 1(b) portray weekday heatmaps showing the average daily investment changes in 2016 broken down by investor's gender and education level (see [44]). First, we observe the richness of our data set in which there is a large heterogeneity of investor's decisions on their investment on a daily basis. Second, though there is a similarity on how investors decide on their investments for males and females and for those with higher and lower education, we observe some discrepancies in some occasions, suggesting that these are two important features that we should study in our empirical analysis. Besides this subjective analysis, our feature selection procedure will corroborate such vision using an objective and quantitative approach. For instance, we observe that, on average, investors mostly buy by the beginning or end of the week while they sell on Wednesdays. There is evidence of behavioral changes of investors over weekdays in the stock market. For instance, Pena [45] studies the effect of reform on the Spanish stock exchange market. They find that, before such reform, there were positive abnormal excess returns on Mondays, effect of which disappeared following that reform.

Figure 2 shows how investments are split across Brazilian states over time. As we can see, there is also some heterogeneity across investors residing in different states, which suggests that we may have to control for state origin of investors. For instance, there are some large investment variations in the Northern region of Brazil.

Figure 3 depicts the distribution of investment variation across different Brazilian states broken down by investor's gender (female or male). Each distribution conditioned to the state and gender integrates to one. Interestingly, most of the distributions have three persist modals that occur not only across different states but also for different genders. The modals are centered at the zero (no investment variation) and at ± 30% investment variation marks. While, in most cases, the profiles of investment changes of both male and female largely coincide, there are some notable exceptions. For instance, in less developed regions—such as the North and Northeast—the distributions of investment decisions of

males and female significantly differ at some dates. Overall, males tend to change more their investment positions relatively to females. However, such feature is even more pronounced in less developed regions.

Figure 4 displays the same distribution of investment variation across different Brazilian states but now broken down by the investor's academic education. We consider investors with higher education and those with high school or below. Again, the three-modal distribution found when we broke down the distribution by investor's gender also show up when we look at their educational levels. In more developed regions—such as the Southeast and South—investors' decision are roughly the same regardless of their academic educational levels. Such similarity reinforces the divergence of academic degree and the level of financial literacy, especially in trading. In contrast, we observe a large heterogeneity in the North region; less educated investors tend to vary their investment positions more than investors those with higher education.

3. Feature Selection Using Machine Learning

In this section, we analyze the predictive power of our attributes in explaining investor responses to changes in the Brazilian stock market index. We use different time aggregations of changes in the IBOVESPA index, which is the financial index that is carefully looked by investors when deciding their investment strategies in Brazil. We use 2-, 3-, and 5-day IBOVESPA index variations, as well as 3- and 5-day IBOVESPA index averages. This analysis sheds light on how investors look at IBOVESPA changes when deciding their trading strategies in the stock market. It is an empirical open question to test whether investors take the very short-term changes, such as 2- or 3-day, or a more prolonged window, such as 5-day changes.

To test the predictive power, we use data-driven machine learning methods to identify the most relevant attributes [46–48]. Since we have data from 13,247 investors from January 1, 2016, to December 31, 2018, on a daily basis, we need first to purge out any macroeconomic factor that could be affecting all investors *in the same manner* over this time frame. This becomes even more important due to the fact that Brazil was facing a recession from 2014Q4 to 2016Q4 and therefore our sample contains part of that period. We perform this preprocessing to homogenize the data distribution, since machine learning methods best perform on cross-sectional data [30, 49].

To remove time factors homogeneously faced by investors in a period, we use a static panel-data specification
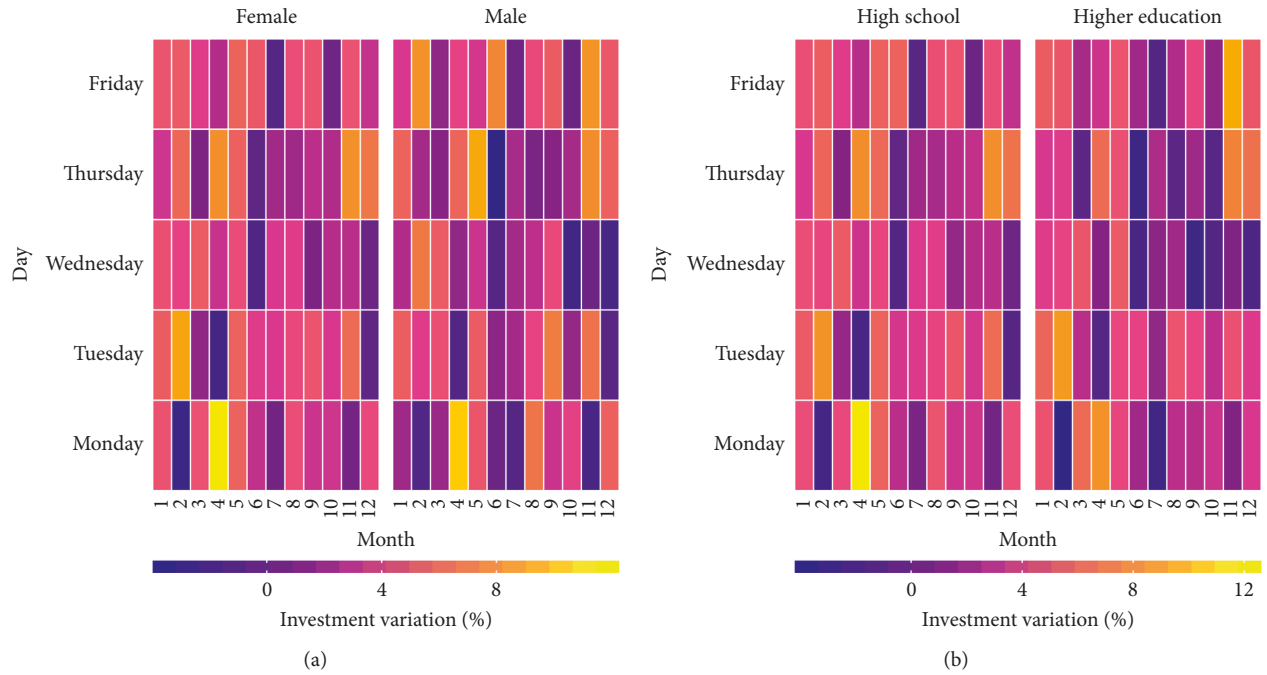
Figure 1: Heatmap showing average daily investment changes (%) in 2016 broken down by investor's gender (a) and education degree (b). We take average values of investment decisions that took place in every weekday within the month. Due to the existence of some large absolute values and to improve readability, we winsorize the investment variation distribution by 5% at each side before taking the average.
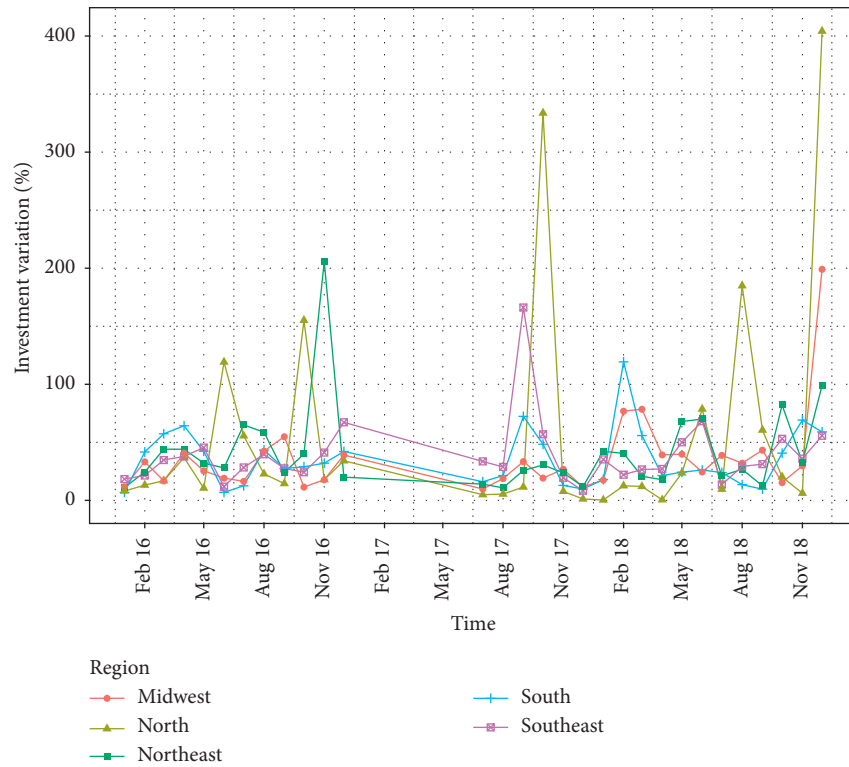

Figure 2: In this figure, we aggregate investment made by all investors by states to which they belong. As we can we have investors from all parts of Brazil in our database.

FIGURE 3: Distribution of investment variation across different Brazilian states (Midwest, North, Northeast, South, and Southeast) broken down by investor's gender (female or male). Due to the existence of some large absolute values and to improve readability, we winsorize the investment variation distribution by 5% at each side.
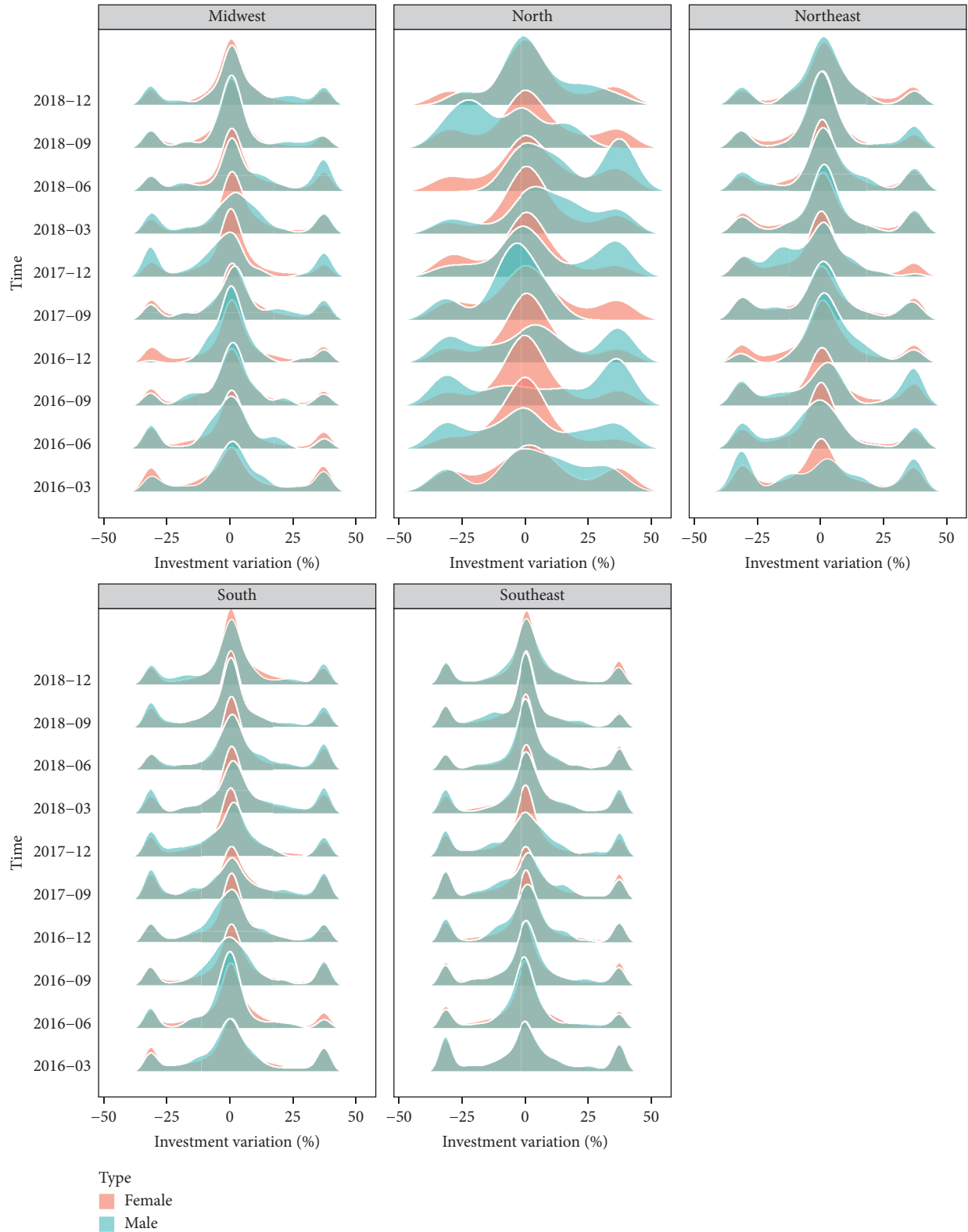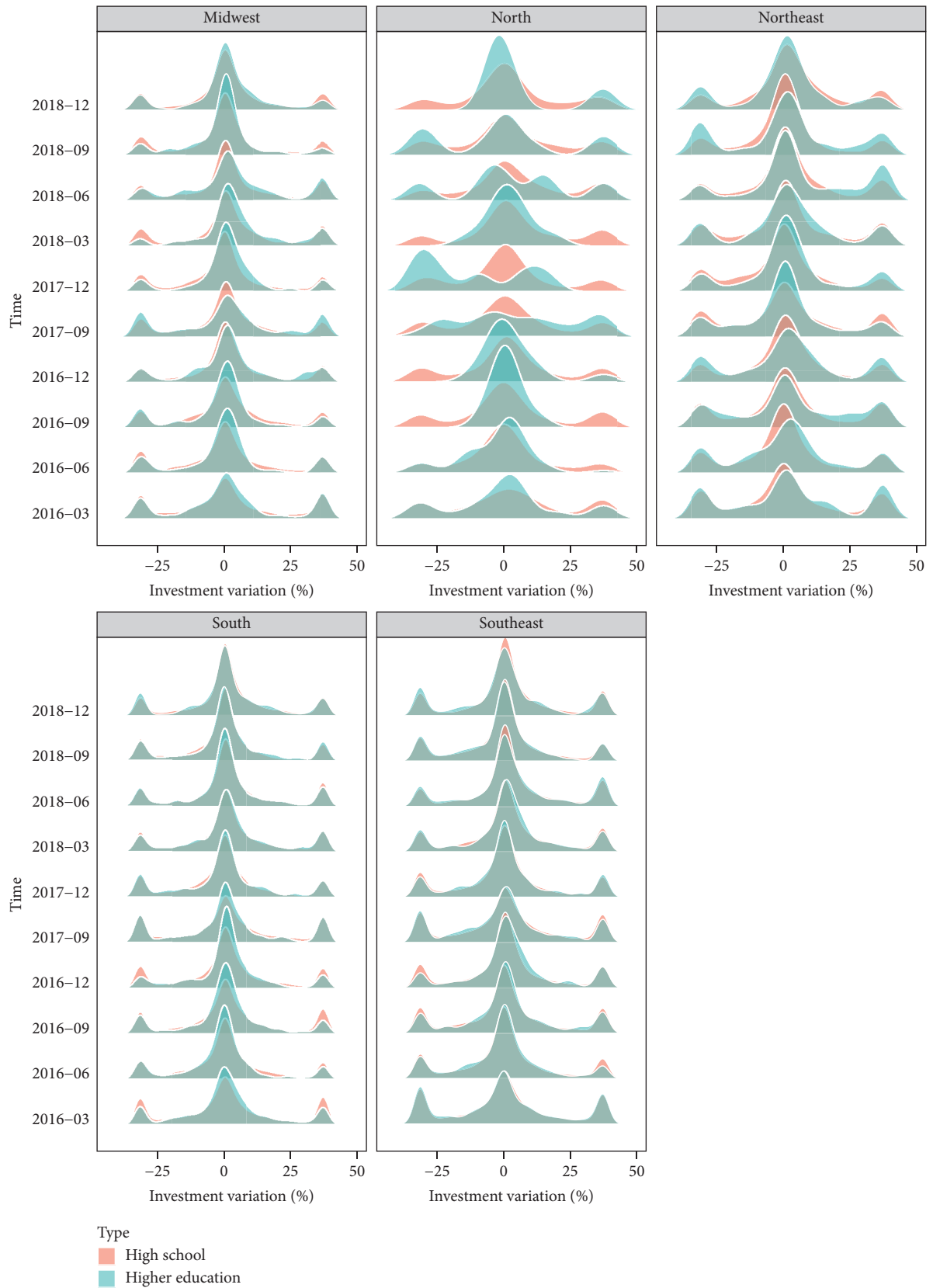
FIGURE 4: Distribution of investment variation across different Brazilian states (Midwest, North, Northeast, South, and Southeast) broken down by investor's education (higher education or high school). Due to the existence of some large absolute values and to improve readability, we winsorize the investment variation distribution by 5% at each side.

with time fixed effects to purge out macroeconomic components as follows:

$$\Delta y_{it} = \alpha_t + \epsilon_{it}, \tag{1}$$

in which Δy_{it} denotes the portfolio volume variation in the stock market of investor i at time t, α_t represents time fixed effects, and ϵ_{it} is the residual. In this specification, we interpret the residual ϵ_{it} as any variation of investor i's portfolio volume at time t that is not due to any time common factor, such as the underlying macroeconomic scenario. By using ϵ_{it} instead of Δy_{it}, we can effectively treat the data as a large cross-sectional unit. Hence, we are able to fully use machine learning methods at their best setup, which we discuss further.

We choose an elastic net regression to estimate the importance of each attribute in the model. Such regression optimally combines L_2-norm (Ridge) and L_1-norm (Lasso) regularization. Therefore, we are able to prevent any overfitting in our empirical model. Moreover, we use a convex combination of L_1-norm, which tends to shrinks the majority of the nonrelevant regressors to zero and keep the most important nonzero, and L_2-norm, which tends to output nonzero and approximate coefficients for all the similar regressors. By using both regularization schemes, we are able to enjoy the positive characteristics of both schemes.

To select the most important attributes, we use the residual ϵ_{it}, the investment volume variation of investor i at time t not due to common time factors, as dependent variable and different IBOVESPA index time aggregations and investors' biological and education characteristics as independent variables as follows:

$$\epsilon_{it} = \boldsymbol{\beta}^T \cdot \mathbf{X}_{it} + \text{error}_{it}, \tag{2}$$

in which $\mathbf{X}_{it}$ is a vector composed of past IBOVESPA changes with different windows (1-, 2-, 3-, 5-, 10-, 20-, and 30-day IBOVESPA changes) and investors' characteristics (state of residency, gender, and level of schooling). The term error_{it} is the standard error. According to the elastic net procedure, we select $\boldsymbol{\beta}$ that minimizes the following loss function $L(\beta)$:

$$L(\beta) = \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\epsilon_{it} - \sum_{j=1}^{p}\beta_j x_{it}^{(j)}\right)^2 + \lambda\left[(1-\alpha)\frac{||\boldsymbol{\beta}||_2^2}{2} + \alpha||\boldsymbol{\beta}||_1^2\right], \tag{3}$$

in which $t \in \{1, \ldots, N\}$ index times on a daily basis (from January 1, 2016, to December 31, 2018,) and i index investors. The term $x_{it}^{(j)}$ indexes the jth regressor of investor i at time t. The operators $||.||_1$ and $||.||_2$ indicate L_1- and L_2-norms taken over the vector input.

The first expression in (3) denotes the traditional data fitting error (residuals), while the second represents the regularization term. Parameter λ modulates the importance of the traditional and regularization terms. The term α controls the convex mixture of L_1 and L_2 regularization. The regularization works by penalizing large β coefficients. Therefore, it shrinks the estimated coefficients and the overall fit function becomes smoother over the data distribution.

In the elastic net regression, a takes values in between 0 and 1. We optimally tune a and λ using a nested cross-validation procedure with $k = 10$ folds and 100 independent repeats for statistical robustness [29, 49]. In this procedure, we use $k - 1 = 9$ folds for training and the remaining fold for testing. This procedure is cycled k times such that each fold appears exactly once for testing. Such methodology enables us to tune the regularization parameters while preventing overfitting of the model. We optimize a over the grid search space $\{0, 0.05, 0.10, \ldots, 1\}$ and λ over $\{0, 0.1, 0.2, \ldots, 5\}$. As standard practice, we preprocess all regressors by applying a Z-score standardization over all the data points using predetermined values extracted only from the training data (so as to prevent data leakage from the test set).

Figure 5 shows our results for the importance of different time aggregations of the IBOVESPA index in explaining investors' behavior. The optimal regularization parameters were $\lambda = 0.1$ and $\alpha = 0.35$. We normalize the coefficients in terms of the most important attribute. The attribute "1-day IBOVESPA variation" is the most powerful predictor for explaining investors' behavior, followed by "2-day IBOVESPA variation" and "5-day IBOVESPA variation." This suggests that investors prefer to base their investment decisions using short-term variations of the stock market index. Even though more prolonged periods of IBOVESPA index changes are important—such as 10-, 20-, and 30-day variations—they are much less important than the short-term variations. In addition, we find that investors' gender and schooling level are also important characteristics for explaining buy and sell operations over the Brazilian stock exchange market in the period from 2016 to 2018. We also observe that some regional variables are important, such as Santa Catarina, Rio de Janeiro, Distrito Federal, Minas Gerais, and Paraná and São Paulo. This may suggest a different mass of investors' composition across different states.

Our feature selection procedure gives us an objective way of identifying potentially important variables that should be accounted for in our econometric exercise. Such tool taken together with the analyst's expertise to assess their validity in terms of relationship with the analyzed measure is an important step in producing econometric methods in a more reliable manner. Our results point that we should control for investors' characteristics (gender and schooling level) and also past IBOVESPA variations. The investor's state is not important because we will employ a panel-data analysis with fixed effects at the investor level. Therefore, the investor's state is collinear with the investor fixed effect and would be dropped during the estimations.

4. Econometric Analysis with Selected Variables

In the previous section, we have found that short-term variations of the IBOVESPA index are better predictors for buy or sell operations in the Brazilian stock exchange market than long-term variations. The feature selection procedure is a transparent way of choosing relevant variables in an objective way. However, such method does not provide an
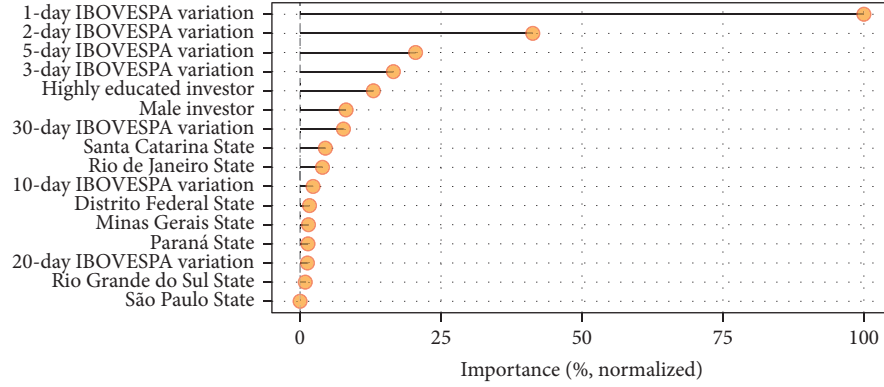
FIGURE 5: Feature selection results using an elastic net procedure with L_2 and L_1 regularization. Coefficients are normalized in terms of the most important attribute ("1-day IBOVESPA variation").

answer as to whether each variable contributes positively or negatively to the target variable, i.e., the investment decision of the investor (buy or sell). In this section, we look at such direction and estimate the magnitude of the most relevant variables found by our feature selection technique.

In Section 4.1, we first test whether investment decisions of Brazilian investors better fit to a mean-reversal or momentum strategy. For that, we regress total investment variations of investors against past variations of the IBO-VESPA index. For robustness, we use 1-, 2-, 3-, 5-, and 30-day variations of the IBOVESPA index. Our regressions are at the investor level, which enables us to control for unobserved time-invariant characteristics of each Brazilian investor, which would otherwise be impractical in case we had aggregate data like most existing studies. Therein, we find that the mean-reversal technique better explains buy or sell operations in the Brazilian stock market during the period from 2016 to 2018. Our results corroborate the findings of our feature selection technique: short-term variations explain more buy or sell operations than long-term variations.

In Sections 4.2 and 4.3, we study the determinants that either soften or exacerbate the mean-reversal behavior of Brazilian investors by looking at the role of gender and level of schooling, respectively, of investors. These exercises connect with the existing literature on the influence of socioeconomic and biological features in shaping the behavior of economic agents.

4.1. Do Investors Use a Mean-Reversion or a Momentum Strategy in Their Buy and Sell Operations? To answer how investors respond to changes in the IBOVESPA index, we run the following econometric specification:

$$\Delta y_{it} = \alpha_i + \alpha_{t^*} + \beta \Delta \text{IBOVESPA}_t + \eta_{it}, \qquad (4)$$

in which Δy_{it} is the portfolio volume variation of investor i at time t. There is a positive variation ($\Delta y_{it} > 0$) when investor i buys more stocks at time t and a negative variation ($\Delta y_{it} < 0$) when she sells. Alternatively, when the investor holds her investment over time, then ($\Delta y_{it} = 0$). The factor η_{it} is the standard error term.

Our coefficient of interest is β, which captures investors' responses to variations of the IBOVESPA index, denoted as $\Delta \text{IBOVESPA}_t$. We test whether investors use a mean-reversal or momentum strategy as follows (we discard the hypothesis that investors' buy and sell decisions are unrelated to variations of the IBOVESPA index because our feature selection technique identified past variations of the IBOVESPA index as the most relevant predictors of investor-specific investment variations):

(i) If investors use a mean-reversal strategy, then increases in the IBOVESPA index—i.e., $\Delta \text{IBOVESPA}_t > 0$—are followed by sell operations in such a way that the investment volume of investors, on average, decreases ($\Delta y_{it} < 0$). Therefore, a mean-reversal strategy is translated by a negative β coefficient ($\beta < 0$).

(ii) If investors use a momentum strategy, then increases in the IBOVESPA index—i.e., $\Delta \text{IBOVESPA}_t > 0$—are followed by buy operations in such a way that the investment volume of investors, on average, increases ($\Delta y_{it} > 0$). Therefore, a momentum strategy is translated by a positive β coefficient ($\beta > 0$).

As there is persistence of the past IBOVESPA index variations by construction, we test how investors' investment volume respond to 1-, 2-, 3-, 5-, and 30-day variations of the IBOVESPA index in an *independent manner*. This empirical design strategy prevents standard errors to get overly inflated due to high pairwise correlation of these regressors.

The term α_i represents investor fixed effects and absorbs any nonobserved time-invariant characteristic of each investor in the sample. This mitigates potential omitted variables that could bias our results, such as investors' skill, which is hard to measure. We should note that any omitted variable that is time variant would not be absorbed by the investor fixed effect. Therefore, while the introduction of such fixed effect mitigates omitted variable bias, it does not completely avoid it. For instance, if investors' skill significantly increases over time, then we would have an omitted variable bias. Since our panel spans a relatively small time period—2016 to 2018—it is fair to assume that investors' skill remains roughly constant. The

Table 2: Output from Regression (4). We ask how investors respond to changes in the IBOVESPA index. We only use changes rather than past averages because the former has greater prediction power as reported by our feature selection procedure. The dependent variable is the variation of portfolio investment volume of investor i at time t in the Brazilian stock market from the beginning of 2016 to the end of 2018. Regressors are 1- (1), 2- (2), 3- (3), 5- (4), and 30-day (5) IBOVESPA index variations. The panel is on a daily frequency basis. Following Petersen [50], we double-cluster standard errors at the investor and time levels. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Dependent variable	Investor portfolio volume variation (Δy_{it})				
	(1)	(2)	(3)	(4)	(5)
Regressor ΔIBOVESPA$_t$ with					
1-day variation	−9.693***				
	(1.580)				
2-day variation		−4.656***			
		(1.160)			
3-day variation			−2.400**		
			(0.964)		
5-day variation				−2.265***	
				(0.852)	
30-day variation					0.058
					(0.680)
Fixed effects					
Investor	Yes	Yes	Yes	Yes	Yes
Month-year	Yes	Yes	Yes	Yes	Yes
Observations	356,172	355,796	355,419	354,588	343,592
R^2	0.037	0.036	0.036	0.035	0.033
Error clustering	Investor	Investor	Investor	Investor	Investor
	Time	Time	Time	Time	Time

term α_{t^*} connotes time-fixed effects at the year-month level, which absorbs any homogeneous time-variant effect, such as the Brazilian recession or month-wise exchange rate fluctuations. Since our panel frequency is on a daily basis, we cannot add a time fixed effect at the same frequency because our coefficient of interest—β—would get absorbed by the time fixed effects as it only varies across time. To prevent such problem, we use a less granular time fixed effects, namely, month-year.

Our data set contains 13,247 investors in a large and representative bank in Brazil and 610 time points. Due to this configuration, we follow Petersen [50] and double-cluster standard errors at the investor and time levels. This is a robust strategy that is important for panels with a large number of individuals and time points because it mitigates heteroscedasticity and serial correlation. Finally, our data is in percentage terms.

Table 2 reports our estimates of Regression (4). We observe that a 1 percent increase of the IBOVESPA index associates with an average decrease of 9.693% of the investor portfolio volume when we look at the 1-day IBOVESPA variation. The results remain with a statistically significant coefficient across different lengths of past IBOVESPA variations (2-, 3-, and 5-day variations), except for 30-day variations, in which the statistical significance vanishes. Moreover, the magnitude of the coefficient reduces as we use less recent past variations of the IBOVESPA index, which is consistent with the view that investors in our sample are more concerned with short-term rather long-term variations of the IBOVESPA index. The negative and statistically significant sign corroborates the hypothesis that investors use mean-reverting trading strategies, in which they tend to

sell after substantial upward changes of the IBOVESPA index, and tend to buy after downward changes.

4.2. Does Gender Impact Investors Responsiveness to IBOVESPA Index Changes? We have showed empirical evidence that investors' strategy, on average, better fit to a mean-reverting behavior in the Brazilian stock market. That is, they tend to sell after positive changes of the IBOVESPA index and buy after negative changes. In this section, we ask whether the sensitiveness of investors to the IBOVESPA index depends on their biological characteristics, in special their gender. Biological factors—especially gender—have been extensively explored in investment decision-making. Notable works relating biological factors, including gender, are Hira and Loibl [51]; Lundeberg et al. [34]; Neyse et al. [33]; and Sunden and Surette [52]. This paper provides further evidence of the existence of such gender gap in investment decisions using a microdata on investor-matched buy and sell operations.

In this line of research, Neyse et al. [33] and Lundeberg et al. [34] partly attribute behavioral differences among males and females due to systematic changes in overconfidence. Excessive overconfidence is associated with higher levels of testosterone, which is more pronounced in males. Overconfidence may induce investors to take on higher risks, leading them to look for higher returns in the short term. In this way, we would expect that females be less sensitive to changes of past IBOVESPA variations as they would value more fundamentals and look for yields in the longer term. Therefore, short-term variations of the IBOVESPA indices would explain less their buy or sell operations comparatively

TABLE 3: Output from Regression (5). We ask whether female investors have different sensitiveness with respect to their investment portfolio to IBOVESPA index changes. We only use changes rather than past averages because the former has greater prediction power as reported by our feature selection procedure. The dependent variable is the variation of portfolio investment volume of investor i at time t in the Brazilian stock market from the beginning of 2016 to the end of 2018. Regressors are 1- (1), 2- (2), 3- (3), 5- (4), and 30-day (5) IBOVESPA index variations, as well as their interaction with the investor's gender. The panel is on a daily frequency basis. Following Petersen [50], we double-cluster standard errors at the investor and time levels. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Dependent variable	Investor portfolio volume variation (Δy_{it})				
	(1)	(2)	(3)	(4)	(5)
Regressor ΔIBOVESPA$_t$ with					
1-day variation	−10.345***				
	(1.754)				
2-day variation		−5.019***			
		(1.275)			
3-day variation			−2.650**		
			(1.055)		
5-day variation				−2.315***	
				(0.886)	
30-day variation					−0.001
					(0.693)
Interactions of ΔIBOVESPA$_t$ with gender					
1-day variation · Female	6.543***				
	(2.392)				
2-day variation · Female		3.708***			
		(1.264)			
3-day variation · Female			2.585**		
			(1.212)		
5-day variation · Female				0.517	
				(1.538)	
30-day variation · Female					0.572
					(0.831)
Fixed effects					
Investor	Yes	Yes	Yes	Yes	Yes
Month-year	Yes	Yes	Yes	Yes	Yes
Observations	356,172	355,796	355,419	354,588	343,592
R^2	0.039	0.037	0.036	0.035	0.034
Error clustering	Investor	Investor	Investor	Investor	Investor
	Time	Time	Time	Time	Time

to males. To empirically answer this question, we construct the following econometric specification:

$$\Delta y_{it} = \alpha_i + \alpha_{t*} + \beta_1 \Delta \text{IBOVESPA}_t + \beta_2 \Delta \text{IBOVESPA}_t \times \text{Female}_i + \eta_{it}, \quad (5)$$

in which Female$_i$ is a dummy variable that takes the value of 1 when investor i is female and 0 otherwise. We do not add the investor's gender alone in (5) because it would be absorbed by the investor fixed effects α_i. Our coefficient of interest is β_2, which captures any behavioral deviation of females to changes of the IBOVESPA index with respect to the average of the entire sample (male and female). If $\beta_2 > 0$, then the mean-reversal strategy is less pronounced to females, while $\beta_2 < 0$ indicates a more accentuated behavior towards the mean-reversal strategy. In the case $\beta_2 = 0$, then females and males respond, on average, equivalently to changes of the IBOVESPA index. Following the discussion on overconfidence and its influence on short-term decisions over males and females, our hypothesis is that $\beta_2 > 0$.

Table 3 reports our estimates of Regression (5). Our previous results relating the mean-reversal strategy of investors in the Brazilian stock market remain the same. We observe that the interaction of changes in the IBOVESPA index and the dummy female is positive and statistically significant. This empirical finding corroborates the view that female investors have a less pronounced mean-reversal strategy than males as they look at longer-term returns and are less attentive to short-term variations of the IBOVESPA index, which could arise due to noisy information. For instance, looking at Specification (1), a 1 percent positive change in the IBOVESPA index associates with a decrease of $-10.345 + 6.543 = -3.802\%$ of the invested volume of female investors. In contrast, the entire sample (males and females) decreases their portfolio volume, on average, by -10.345% for a 1 percent positive change in the IBOVESPA index. Interestingly, even though statistically insignificant, 30-day variations of the IBOVESPA index are positively associated with investment volumes for females, suggesting traits of a momentum strategy. This is also suggestive evidence that

TABLE 4: This table reports output from Regression (6). We ask whether investors with higher academic degree have different sensitiveness with respect to their investment portfolio to IBOVESPA index changes. We only use changes rather than past averages because the former has greater prediction power as reported by our feature selection procedure. The dependent variable is the variation of portfolio investment volume of investor i at time t in the Brazilian stock market from the beginning of 2016 to the end of 2018. Regressors are 1- (1), 2- (2), 3- (3), 5- (4), and 30-day (5) IBOVESPA index variations, as well as their interaction with the investor's academic degree. The panel is on a daily frequency basis. Following Petersen [50]; we double-cluster standard errors at the investor and time levels. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Dependent variable	Investor portfolio volume variation (Δy_{it})				
	(1)	(2)	(3)	(4)	(5)
Regressor $\Delta\Delta$IBOVESPA$_t$ with					
1-day variation	−10.347***				
	(1.795)				
2-day variation		−5.136***			
		(1.299)			
3-day variation			−2.750**		
			(1.076)		
5-day variation				−2.565***	
				(0.931)	
30-day variation					0.085
					(0.695)
Interactions of ΔIBOVESPA$_t$ with academic degree					
1-day variation · Higher education	5.347***				
	(1.520)				
2-day variation · Higher education		3.915***			
		(1.040)			
3-day variation · Higher education			2.864*		
			(1.573)		
5-day variation · Higher education				2.471*	
				(1.398)	
30-day variation · Higher education					−0.237
					(0.647)
Fixed effects					
Investor	Yes	Yes	Yes	Yes	Yes
Month-year	Yes	Yes	Yes	Yes	Yes
Observations	356,172	355,796	355,419	354,588	343,592
R^2	0.038	0.036	0.036	0.035	0.035
Error clustering	Investor	Investor	Investor	Investor	Investor
	Time	Time	Time	Time	Time

females tend to look at longer horizons when taking investment decisions.

4.3. Does Formal Education Impact Investors Responsiveness to IBOVESPA Index Changes?

In this section, we look at how formal education (academic degree or level of schooling) can influence investors' sensitiveness to IBOVESPA index variations. There are several works in the behavioral finance literature that attempt to establish a connection between level of schooling and investors' awareness of stock markets and their decision-making determinants. We highlight the research studies of Grinblatt et al. [53] and Guiso and Jappelli [54]. In theory, educated investors should behave in more rational ways and trade less frequently when there is no new relevant information arriving in the market but noises. Therefore, we would expect these investors to have a smaller reaction to price fluctuations as they are able to better identify information from noise. To empirically test this behavior, we run the following specification:

$$\Delta y_{it} = \alpha_i + \alpha_{t^*} + \beta_1 \Delta\text{IBOVESPA}_t + \beta_2 \Delta\text{IBOVESPA}_t \times \text{Higher Education}_i + \eta_{it}, \tag{6}$$

in which Higher Education$_i$ is a dummy variable that takes the value of 1 when investor i has a higher education (at least college degree) and 0 otherwise (high school or a lower degree). Our coefficient of interest is β_2, which captures any behavioral deviation of investors with higher formal education to changes of the IBOVESPA index with respect to the average of the entire sample. The hypothesis is that $\beta_2 < 0$, in which more educated investors tend to better discern information from noise out of variations of the IBOVESPA index and therefore the mean-reversal strategy would be less pronounced.

Table 4 reports our estimates of Regression (6). On average, the mean-reversal strategy remains. We note that the interaction of changes of the IBOVESPA index and the dummy higher education is positive and statistically significant. This suggests that investors with higher academic

degree have a less pronounced mean-reversal strategy than less educated investors, which corroborates our hypothesis. Looking at Specification (3), we observe a positive, though marginally significant, relationship between IBOVESPA changes and investment volume $(-2.750 + 2.864 = 0.114)$ for more educated investors, suggesting traits of a momentum strategy.

5. Conclusions

We employ machine learning techniques together with econometrics techniques to model investor behavior using a unique dataset for investors that focus on stock market investments. We propose a methodological approach to link machine learning methods widely used in computer science to standard econometric techniques commonly employed in social sciences.

Using the unique data set with high-frequency daily investment decision of a broad set of investors in Brazil, we provide evidence that investors look at past performance of a benchmark stock index in order to decide their investment decisions. Investors seem to prefer mean-reverting strategies in the short-run, rather than momentum. This may be associated with the disposition effect - investors prefer to sell the winners and buy the losers [55, 56]. Furthermore, research could exploit alternative explanations for this behavior.

In addition, we study the determinants that either soften or exacerbate the mean-reversal behavior of Brazilian investors by looking at the role of gender and level of schooling. We find that females and more educated investors are less sensitive to changes of past IBOVESPA variations, which is consistent with the literature on behavioral finance.

This paper highlights the importance of using non-traditional methods in econometric analysis. The use of machine learning methods permits us to automate the often subjective process of choosing which variables are important in any econometric analysis. By using a feature selection scheme—such as the elastic net in this paper—we are able to identify those attributes that best describe how investors decide to buy or sell their positions in an objective and statistically correct manner. In addition to that, the business specialist can always assess these variables pointed out as most important to analyze their economic meaning.

Data Availability

The data is confidential.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[2] E. F. Fama and K. R. French, "Permanent and temporary components of stock prices," *Journal of Political Economy*, vol. 96, no. 2, pp. 246–273, 1988.

[3] A. W. Lo and A. C. MacKinlay, "Stock market prices do not follow random walks: evidence from a simple specification test," *Review of Financial Studies*, vol. 1, no. 1, pp. 41–66, 1988.

[4] J. M. Poterba and L. H. Summers, "Mean reversion in stock prices," *Journal of Financial Economics*, vol. 22, no. 1, pp. 27–59, 1988.

[5] J. Conrad and G. Kaul, "An anatomy of trading strategies," *Review of Financial Studies*, vol. 11, no. 3, pp. 489–519, 1998.

[6] C. Engel and J. D. Hamilton, "Long swings in the dollar: are they in the data and do markets know it?," *The American Economic Review*, vol. 80, pp. 689–713, 1990.

[7] N. Jegadeesh and S. Titman, "Momentum," *Annual Review of Financial Economics*, vol. 3, no. 1, pp. 493–509, 2011.

[8] R. S. J. Koijen, J. C. Rodríguez, and A. Sbuelz, "Momentum and mean reversion in strategic asset allocation," *Management Science*, vol. 55, no. 7, pp. 1199–1213, 2009.

[9] M. Morrin, J. Jacoby, G. V. Johar, X. He, A. Kuss, and D. Mazursky, "Taking stock of stockbrokers: exploring momentum versus contrarian investor strategies and profiles," *Journal of Consumer Research*, vol. 29, no. 2, pp. 188–198, 2002.

[10] J. Okunev and D. White, "Do momentum-based strategies still work in foreign currency markets?," *The Journal of Financial and Quantitative Analysis*, vol. 38, no. 2, pp. 425–447, 2003.

[11] D. Schiereck, W. De Bondt, and M. Weber, "Contrarian and momentum strategies in Germany," *Financial Analysts Journal*, vol. 55, no. 6, pp. 104–116, 1999.

[12] D. O. Cajueiro and B. M. Tabak, "Testing for predictability in equity returns for European transition markets," *Economic Systems*, vol. 30, no. 1, pp. 56–78, 2006.

[13] D. O. Cajueiro and B. M. Tabak, "Testing for time-varying long-range dependence in real state equity returns," *Chaos, Solitons & Fractals*, vol. 38, no. 1, pp. 293–307, 2008.

[14] E. J. Chang, E. J. A. Lima, and B. M. Tabak, "Testing for predictability in emerging equity markets," *Emerging Markets Review*, vol. 5, no. 3, pp. 295–316, 2004.

[15] A. Sensoy, K. Ozturk, E. Hacihasanoglu, and B. M. Tabak, "Not all emerging markets are the same: a classification approach with correlation based networks," *Journal of Financial Stability*, vol. 33, pp. 163–186, 2017.

[16] B. M. Tabak and E. J. A. Lima, "Market efficiency of Brazilian exchange rate: evidence from variance ratio statistics and technical trading rules," *European Journal of Operational Research*, vol. 194, no. 3, pp. 814–820, 2009.

[17] C. M. Boya, "From efficient markets to adaptive markets: evidence from the French stock exchange," *Research in International Business and Finance*, vol. 49, pp. 156–165, 2019.

[18] R. Ding and P. Cheng, "Speculative trading, price pressure and overvaluation," *Journal of International Financial Markets, Institutions and Money*, vol. 21, no. 3, pp. 419–442, 2011.

[19] E. Lee and N. Piqueira, "Behavioral biases of informed traders: evidence from insider trading on the 52-week high," *Journal of Empirical Finance*, vol. 52, pp. 56–75, 2019.

[20] T.-Y. Pak and P. Babiarz, "Does cognitive aging affect portfolio choice?," *Journal of Economic Psychology*, vol. 66, pp. 1–12, 2018.

[21] T. Suzuki and Y. Ohkura, "Financial technical indicator based on chaotic bagging predictors for adaptive stock selection in Japanese and American markets," *Physica A: Statistical Mechanics and its Applications*, vol. 442, pp. 50–66, 2016.

[22] A. Urquhart and F. McGroarty, "Are stock markets really efficient? evidence of the adaptive market hypothesis," *International Review of Financial Analysis*, vol. 47, pp. 39–49, 2016.

[23] X. Xiong, Y. Meng, X. Li, and D. Shen, "An empirical analysis of the adaptive market hypothesis with calendar effects: evidence from China," *Finance Research Letters*, vol. 31, 2019.

[24] H. Takahashi and T. Terano, "Analyzing the influence of overconfident investors on financial markets through agent-based model," in *Intelligent Data Engineering and Automated Learning-IDEAL 2007*, H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao, Eds., Springer, Berlin, Heidelberg, Germany, 2007.

[25] B. LeBaron, "Empirical regularities from interacting long- and short-memory investors in an agent-based stock market," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 5, pp. 442–455, 2001.

[26] M. A. Bertella, F. R. Pires, L. Feng, and H. E. Stanley, "Confidence and the stock market: an agent-based approach," *PLoS One*, vol. 9, no. 1, Article ID e83488, 2014.

[27] H. R. Varian, "Big data: new tricks for econometrics," *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014.

[28] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, Springer Publishing Company, Incorporated, Switzerland, 2014.

[29] T. C. Silva and L. Zhao, *Machine Learning in Complex Networks*, Springer Publishing Company, Incorporated, Switzerland, 1st edition, 2016.

[30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, Berlin, Heidelberg, Germany, 2nd edition, 2009.

[31] K. J. Arrow and G. Debreu, "Existence of an equilibrium for a competitive economy," *Econometrica*, vol. 22, no. 3, pp. 265–290, 1954.

[32] H. A. Simon, *Models of Man: Social and Rational- Mathematical Essays on Rational Human Behavior in a Social Setting*, Wiley, Hoboken, NJ, USA, 1957.

[33] L. Neyse, S. Bosworth, P. Ring, and U. Schmidt, "Overconfidence, incentives and digit ratio," *Scientific Reports*, vol. 6, no. 1, 2016.

[34] M. A. Lundeberg, P. W. Fox, and J. Pun coha, "Highly confident but wrong: gender differences and similarities in confidence judgments," *Journal of Educational Psychology*, vol. 86, no. 1, pp. 114–121, 1994.

[35] O. Onishchenko and N. Ülkü, "Foreign investor trading behavior has evolved," *Journal of Multinational Financial Management*, vol. 51, pp. 98–115, 2019.

[36] M. Abreu, "How biased is the behavior of the individual investor in warrants?," *Research in International Business and Finance*, vol. 47, pp. 139–149, 2019.

[37] J.-C. Li, Y.-X. Li, N.-S. Tang, and D.-C. Mei, "The roles of mean residence time on herd behavior in a financial market," *Physica A: Statistical Mechanics and Its Applications*, vol. 462, pp. 350–357, 2016.

[38] C. Liu and X. Li, "Media coverage and investor scare behavior diffusion," *Physica A: Statistical Mechanics and Its Applications*, vol. 527, p. 121398, 2019.

[39] K. W. Park, S. H. Jeong, and J. Y. J. Oh, "Foreigners at the gate? foreign investor trading and the disposition effect of domestic individual investors," *The North American Journal of Economics and Finance*, vol. 49, pp. 165–180, 2019.

[40] Y. Shi, Y.-r. Tang, and W. Long, "Sentiment contagion analysis of interacting investors: evidence from China's stock forum," *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 246–259, 2019.

[41] J. R. Wei, J. P. Huang, and P. M. Hui, "An agent-based model of stock markets incorporating momentum investors," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 12, pp. 2728–2735, 2013.

[42] R. J. Balvers and Y. Wu, "Momentum and mean reversion across national equity markets," *Journal of Empirical Finance*, vol. 13, no. 1, pp. 24–48, 2006.

[43] D. B. Chaves and V. Viswanathan, "Momentum and mean-reversion in commodity spot and futures markets," *Journal of Commodity Markets*, vol. 3, no. 1, pp. 39–53, 2016.

[44] B. M. Barber and T. Odean, "Boys will be boys: gender, overconfidence, and common stock investment," *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 261–292, 2001.

[45] J. I. PeŇa, "Daily seasonalities and stock market reforms in Spain," *Applied Financial Economics*, vol. 5, no. 6, pp. 419–423, 1995.

[46] T. C. Silva and L. Liang Zhao, "Network-based high level data classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 954–970, 2012.

[47] T. C. Silva and L. Liang Zhao, "Network-based stochastic semisupervised learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 451–466, 2012.

[48] T. C. Silva and L. Liang Zhao, "Stochastic competitive learning in complex networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 385–398, 2012.

[49] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, Germany, 2006.

[50] M. A. Petersen, "Estimating standard errors in finance panel data sets: comparing approaches," *Review of Financial Studies*, vol. 22, no. 1, pp. 435–480, 2009.

[51] T. K. Hira and C. Loibl, "Gender differences in investment behavior," in *Handbook of Consumer Finance Research*, J. J. Xiao, Ed., Springer, New York, NY, USA, 2008.

[52] A. E. Sunden and B. J. Surette, "Gender differences in the allocation of assets in retirement savings plans," *The American Economic Review*, vol. 88, pp. 207–211, 1998.

[53] M. Grinblatt, M. Keloharju, and J. Linnainmaa, "IQ and stock market participation," *The Journal of Finance*, vol. 66, no. 6, pp. 2121–2164, 2011.

[54] L. Guiso and T. Jappelli, "Awareness and stock market participation," *Review of Finance*, vol. 9, no. 4, pp. 537–567, 2005.

[55] N. Barberis and W. Xiong, "What drives the disposition effect? an analysis of a long-standing preference-based explanation," *The Journal of Finance*, vol. 64, no. 2, pp. 751–784, 2009.

[56] H. Shefrin and M. Statman, "The disposition to sell winners too early and ride losers too long: theory and evidence," *The Journal of Finance*, vol. 40, no. 3, pp. 777–790, 1985.