Lottery Stocks in Brazil: Analyzing Return on Betting

Gabriel Sifuentes Rocha

FEARP-USP

Abstract

Lotteries, contrary to expectations from traditional utility models under uncertainty, attract the interest of many human beings. In the North American financial market, this phenomenon could be observed in several contexts, with individuals exchanging – consciously or not – expected risk-adjusted return for expected positively skewness, even if this implied taking more risks and obtaining lesser average returns. In this study, the various theories formulated to explain this anomaly in human behavior and, consequently, an anomaly in the returns of traded financial assets, are condensed. In addition, an investigation is carried out regarding the premium paid by anti-lottery stocks over lottery stocks in Brazil following three different methodologies and, subsequently, a multivariate analysis of the lottery factor is also carried out in conjunction with the five-factor model developed by Fama and French (2014) controlling the portfolios analyzed by size and illiquidity.

JEL: C00, D84, G11

1. INTRODUCTION

To understand the determinants of stock returns, several studies focused on asset pricing models. One of the studies that started this path was the Capital Asset Pricing Model (CAPM), innovating by modeling the nature of investor preferences, classifying them as risk averse and creating a pricing model that considered the rationality of these individuals. Thus, the CAPM established the relationship between the risk and return of an asset considering its sensitivity to market fluctuations. Assets that are more (less) sensitive to economic conditions, that is, more (less) exposed to systemic risk, would tend to perform better (worse) in moments of high (low) market, while assets with a beta close to zero would converge to the risk-free interest rate (SHARPE, 1964). Furthermore, due to the risk aversion shown by investors, these assets that are more sensitive to market risk would tend to have a higher discount rate in their prices and, therefore, a higher expected return (due to the risk premium).

However, with the help of other researchers, such as Kraus and Litzenberger (1976), Fama and French (1992) and Bali, Brown, Murray and Tang (2017), it was possible to observe that the CAPM does not systematically explain a relevant part of the stock variance, which could be explained by other factors. One of them concerns the greater return of value stocks compared to growth stocks even after controlling for market beta (FAMA; FRENCH, 1992), a result not expected by the model proposed by Sharpe (1964), since such stocks should present, on average, gains (losses) proportional to their respective market betas. This difference between the expected result and the empirical result is due to the omission of two other risk factors capable of explaining the return on assets: the size of the company (measured by market capitalization) and its value (measured by the ratio between its price and book value or by the ratio between its price and earnings). Later, Fama and French (2014) added two more factors, investment and profitability, resulting in five risk factors to make the pricing model even more complete and accurate. Other researchers also brought advances to the theory with the study of new factors, such as Barberis and Huang (2004) with asymmetry and Skočir and Lončarski (2018), with momentum, default risk and liquidity *risk*.

However, although some of these premiums are easily explained by greater risk, which would support the efficient markets hypothesis (FAMA; FRENCH, 2004), others do not. This is the case of the momentum identified by Jegadeesh and Titman (1993) when noting that companies that performed well in the short term tended to continue their respective trajectory, as well as companies with unsatisfactory results, continued to lose market value. In these cases, behavioral explanations such as disposition-effect (investors tend to sell assets at a gain, to make a profit and hold on to falling assets, to avoid losses) are more adequate than explanations based on greater risk to justify the "abnormal" returns in question (PITTHAN, 2018). This is because investors are not purely rational, and there are costs and risks that limit the arbitrage opportunities that arise in the market, so that these distortions arising from behavioral biases can persist for extended periods and continue to exist even after the phenomenon is widely known.

Thus, one can see that, just like the momentum factor, many of the elements that seek to explain the expected return on stocks cannot be supported by theses that assume rational investors, perfectly competitive, efficient and frictionless markets (which would allow the arbitrage and correction of possibly distorted prices). These new theses are based on behavioral issues that assume a certain tendency/bias of investor behavior that leads them to distort the fair prices of assets. Such theses are also based on the existence of frictions and arbitrage difficulties in financial markets, either due to the risks involved (it takes time for the price of an asset to converge to its fundamentals), or due to transaction costs (which reduce the arbitrator's premium).

In this article, the behavioral bias that will be studied is the preference for lottery stocks, that is, stocks with a distribution of expected returns represented by an asymmetry to the right (positively skewed). One hypothesis that explains investors preference for lotteries is investor’s overconfidence, which leads them to believe that their investments will be rewarded with very high returns (BRUNNERMEIER; PARKER, 2005), which causes tail probabilities to be overestimated, leading them to pay an amount greater than the expected gain discounted by the risk premium.

Therefore, the main objective of this study is to identify this behavior in the returns of this type of asset in Brazil, focusing more specifically on the MAX factor and the illusion caused by the low nominal price (two lottery metrics that will be used in this article). It is expected that, equipped with more information, investors will have greater knowledge about the behavioral biases that can negatively affect their investment decisions, in addition to identifying more precisely what was the size of this impact on the expected return of the portfolio in the years studied.

In this sense, in the next section, Literature Review, the main ideas of the authors about lottery stocks and their impact on the United States's financial market will be presented, in addition to fundamentals that may explain the preference of individual investors for this type of asset. In section 3, Methodology and Data, it will be explained how the present work will be conducted, in addition to the data source used for the research; in section 4, the results obtained; in 5, the Conclusion of the work and, finally, the references.

2. LITERATURE REVIEW

Human beings's interest in gambling is definitely not a recent phenomenon: mentions of games for the distribution of wealth, clothing and land based on luck date back to biblical times in books such as Joshua (21:8) and Ezekiel (48:29). Currently, in addition to lotteries *per se*, other bets based on chance, such as raffles and roulettes, demonstrate this age-old human desire to submit to luck to obtain, or not, some return.

According to the more traditional view of Tversky and Kahneman (1979) on decision making under uncertainty, it is observed that the utility function for changes in wealth is concave for gains and convex (in addition to being more sensitive) for losses, which implies a decreasing tendency of the marginal value of the function according to its magnitude. Visualizing this hypothesis algebraically, being x the variation in the income of a given individual and f(x) its utility function, we have that:

$$f(x)'' < 0 \text{ for } x > 0 \quad (1)$$

$$f(x)'' > 0 \text{ for } x < 0 \quad (2)$$

This means that the greater the magnitude of a person's gains (losses), the less sensitive they will be to this increase (reduction) of wealth, that is:

> "[...] the difference in value between a gain of 100 and a gain of 200 appears to be greater than the difference between a gain of 1,100 and a gain of 1,200. Similarly, the difference between a loss of 100 and a loss of 200 appears greater than the difference between a loss of 1,100 and a loss of 1,200, unless the larger loss is intolerable." (TVERSKY; KAHNEMAN, 1979, p.17).

According to this theory, human beings would not be expected to be attracted by lotto games, since, in general, their utility functions have the characteristic of loss aversion and, in these games, the chance of losing is quite significant without there being a high expectation of return, on average, to compensate for the risk. So, being more sensitive to losses than gains, and the chances of losing being almost 100%, why do countless individuals subject themselves to these games? Moreover, if the marginal utility of the gains decreases as the values of the gains increase, it is difficult to understand this obstinacy for exorbitant prizes located in the right tail of the probability density function of this type of bet.

In this line of reasoning, it seems (almost) inexplicable that the preference for distributions of returns (payoffs) inconsistent with the characteristics of the utility functions theorized and measured by most of the traditional research. Motivated to justify this phenomenon, Garret and Sobel (1999) suggest that people, when participating in games of chance, are not looking for risk (violating the traditional hypothesis of loss aversion), but in search of an asymmetric distribution of returns (positively skewed or right-skewed distributions). This is in line with the empirical conclusion found by Golec and Tamarkin (1998) from the observation of horse races. Kraus and Litzenberger (1976) and Chiu (2010) propose an algebraic solution to this problem, defending the cubic format of the utility function, thus accommodating the increasing marginal utility in the right tail of the probability distribution of prizes in games of chance.

This demand for skewness can also be identified in the financial market - which will be the main object of this study, attracting mainly young people, single/divorced men and low-income people, with the phenomenon being aggravated during economic recessions, such as the 1929 crisis (KUMAR, 2009). Despite the different definitions used in the literature to describe what lottery stocks are, they tend to share a set of common characteristics, such as, for example, attracting mainly non-institutional investors. The work by Agarwal, Jiang and Wen (2022) complements this statement by demonstrating that most professional managers do not invest in *lottery stocks* , but when they do, they usually have the objective of generating a greater flow of investments - from individual investors - to their respective funds. The only exception to this is during the fourth quarter, when professional investors who have underperformed their peers year-to-date look for “lotteries” in a valiant attempt to end the year above the fund's benchmark.

But, to the misfortune of these gamblers, several studies show that this type of stock presents a significantly negative return on risk compared to investing in non-lottery stocks, with market portfolio or other benchmarks established by stock research. The explanation for why the average return is lower for this type of investment is simple: either because they value skewness, or because they assign weights greater than the real ones to the probability of occurrence of tail events, many investors are willing to buy these shares for values greater than expected. Thus, they remain consistently more expensive, which leads to a lower expectation of future returns. Basically, what you see is an exchange - conscious or not - of expectancy of skewness for expectancy of return. It is possible to observe this tradeoff in several studies and scenarios, which indicates that this effect is robust in the most diverse contexts and is part of a general human characteristic. For this reason, it is expected to find similar evidence in the Brazilian market. Some examples of studies that reinforce the existence of such anomalies will be presented below.

Barberis and Huang (2004), based on the Cumulative Prospect Theory (CPT) by Tversky and Kahneman (1992), add a new beta to the CAPM model, indicating the asymmetry of stocks, in order to predict their expected return. With this, they demonstrate that the introduction of a security with positive skewness in the economy leads to the overvaluation of the security and, therefore, to a lower average return. This behavior is commonly observed in Initial Public Offerings (IPOs), where there is a statistically significant distortion of returns to the right. Bearing in mind the tradeoff described above, the underperformance of IPOs is no longer an unexpected anomaly and is now explained by this "*lotto preference*" theory.

The same phenomenon can also be observed in stocks traded on the American over-the-counter (OTC) market (ERAKER; READY, 2015). In addition to their individual positive skewness, their (CAPM and Fama-French) alphas are significantly negative on average, around -2% per month, especially after considering transaction costs. Thus, one of the reasons why investors admit such results is the fact that this type of asset allows the construction of a portfolio whose returns have an asymmetric distribution to the right, similar to that of a lottery. Conrad, Kapadia and Xing (2014) also observed this phenomenon in companies with a high risk of bankruptcy. As they provide a tail probability with a high payoff in the unlikely recovery scenario, investors are attracted to buying these shares at prices higher than the expected value generation provided by the share, to generate a negative alpha when buying them.

Kumar (2009), to identify lotteries in the market, selected stocks that simultaneously met three criteria relative to other US stocks: nominal price below the kth percentile; asymmetry of returns greater than the kth percentile; and idiosyncratic volatility above the kth percentile. The tests were repeated for K=50 and K=33, and similar results were obtained in both applications. With this classification, the author, in line with other research, found that stocks with a lottery profile have low market value, high book value/price (expensive stocks), less analyst coverage, low institutional ownership (small participation of legal entities in the company's shareholder body), lower trading volume, shorter listing time on the stock exchange and paying few dividends. Besides, once again, they have (CAPM, Fama-French and others) negative alphas, with robust statistical significance. These findings are important because they explain, in part, the reason for this price distortion to remain persistent: short sale constraints (BALI; CAKICI; WHITELAW, 2011). As overpricing occurs mainly in illiquid stocks and low capitalization, it would be costly to sell the stock short in order to correct the security price due to transaction costs and difficulties in liquidating the position in times of need. Furthermore, it is a dangerous operation, especially when dealing with lottery stocks since the gain is limited by the value of the operation and the loss is potentially infinite.

Based on the above, one of the main identifying factors of lotteries is the distortion of their return to the right, so that one of the ways to objectively measure it is through asymmetry. The skewness (third moment in a probability density function) indicates the distance between the symmetry/regularity and the frequency distribution of a given event, being generally defined mathematically according to the equation below:

$$Skewness = \frac{1}{n}\sum\left(\frac{r_i - \bar{r}}{\sigma}\right)^3 \qquad (3)$$

Correcting the degree of freedom bias for small samples, we have:

$$Skewness = \frac{n}{(n-1)\times(n-2)}\sum\left(\frac{r_i - \bar{r}}{\sigma}\right)^3 \qquad (4)$$

Arditti (1967), when studying preferences for skewness, explores the relationship between total asymmetry and ex post returns, showing that, given a constant variance, investors incur lower expected returns since they provide an expected payoff with greater distortion to right. In a context of variable variance, in order to obtain lower volatility, it would be necessary to diversify between distorted securities, which would reduce the exposure of the portfolio as a whole to the expected asymmetry. The Barberis and Huang (2004) model demonstrates that, when choosing to diversify between stocks, the skewness of the portfolio reduces more abruptly than the standard deviation, which makes diversification unattractive for those individuals who seek asymmetry. Complementarily, Brunnermeier, Gollier and Parker (2007) found that investors choose to under-diversify their portfolios in order to obtain positively distorted returns. These conclusions corroborate the findings of Boyer, Mitton and Vorkink (2009) and Chiu (2010) and show that not only the systemic asymmetry of a role is valued by individuals, but mainly the idiosyncratic asymmetry, since they hardly invest in properly diversified portfolios (the median number of securities in the portfolio of US individuals is three) (KUMAR, 2009). For this reason, in this work applied to Brazil, idiosyncratic metrics will be used.

Furthermore, just like Kraus and Litzenberger (1976), Boyer, Mitton and Vorkink (2009) agree with the perspective that ex-ante (past) skewness does not necessarily imply expectations of future asymmetry on the part of investors in the market. For them, past asymmetry alone is not able to adequately predict future asymmetry (in the context of under-diversification of a portfolio). Instead, they found that a stock's lagged specific volatility is a strong predictor of its respective idiosyncratic asymmetry (characteristic of lotteries). For this reason, in this article, we will not use mere past idiosyncratic asymmetry as an indicator of expected skewness, but an idiosyncratic volatility metric. Using Fama and French's (2014) three-factor model, they found considerable pricing effects of predicted asymmetry - comparing bonds with the lowest and highest asymmetry, they found an alpha difference of about 1% per month, an inverse - and significant - relationship between earnings and expected idiosyncratic asymmetry. Thus, they concluded that investors end up paying more for a more volatile asset and, consequently, face lower expected returns because this oscillation indicates a lottery opportunity, that is, future exposure to skewness, the desired position.

A better candidate for a lottery-stocks indication might be MAX, as defined by Bali, Cakici and Whitelaw (2011), that constructs this metric as the maximum daily return of a stock during the previous month, propose that the greater this factor is, the greater the potential for extremely positive returns perceived by investors, which makes that they agree to pay more and earn lower average returns per type of asset. That is, compared to non-lottery assets, the return on lottery stocks (classified according to their MAX) is lower. These conclusions were obtained empirically by the authors through two analysis methodologies.

The first, a bivariate analysis, consisted of grouping portfolios of deciles ordered by their MAX. The authors found significant differences between decile 10 (highest MAX) and decile 1 (lowest MAX): for portfolios weighted by company value, the reduction in gross return found was -1.03%, an amount similar to the alpha of the model Fama-French-Carhart four-factor analysis of -1.18%. The results found in deciles 1 to 7 do not vary so much, however, from decile 7 to 10 (with the sharpest MAX), the average returns and alphas drop substantially. Similar conclusions were obtained using the average of the two, three, four and five highest daily returns in the month, with the MAX (5) results being even more robust.

Furthermore, this indicator was observed to be persistent over time: stocks in the tenth decile were 35% (68%) likely to be in the top tenth (tenth, ninth, or eighth) decile in the subsequent month. This implies a certain rationality on the part of investors, since, even though investing in stocks with a high MAX entails, on average, lower returns, the utility of these individuals does not consist only in the expectancy of return, but also in the expectation of asymmetry (which is very positive in this type of stock).

The low (high) average returns results for high (low) MAX stocks were robust to several other factors. This is important to prove that the MAX brings with it new information and is not merely a redefinition of an existing factor. Therefore, even controlling for book-to-market (value factor), momentum, short-term reversal (short-term mean reversion), illiquidity, market capitalization (size), idiosyncratic skewness, systemic skewness and idiosyncratic volatility, the MAX effect remained statistically significant, in other words, these other factors were unable to explain the low returns associated with higher deciles. Similar conclusions can be reached analyzing the alphas generated by the indicator, with some of them being even more significant than the difference in returns itself.

Although the bivariate analysis has advantages for being a non-parametric model, therefore, for not imposing a functional formula of the relationship between future returns and the MAX, its methods have some points of failure: i) the difficulty in controlling several factors concurrently, ii) the underutilization of information within the cross section; and iii) the clustering percentiles may not be granular enough to capture the difference in control variables. In this sense, to increase the robustness of the MAX factor, the tests were redone using another methodology, the Fama and MacBeth regression (1973):

$$R_{i,t+1} = \lambda_{\mathbf{0},t} + \lambda_{\mathbf{1},t}\, \boldsymbol{MAX}_{i,t} + \lambda_{\mathbf{2},t}\, \boldsymbol{BETA}_{i,t} + \lambda_{\mathbf{3},t}\, \boldsymbol{SIZE}_{i,t} + \lambda_{\mathbf{4},t}\, \boldsymbol{BM}_{i,t} + \lambda_{\mathbf{5},t}\, \boldsymbol{MOM}_{i,t} + \lambda_{\mathbf{6},t}\, \boldsymbol{REV}_{i,t} + \lambda_{\mathbf{7},t}\, \boldsymbol{ILLIQ}_{i,t} + \boldsymbol{\varepsilon}_{i,t+1} \quad (5)$$

In the equation above, $R_{i,t+1}$it represents the realized return by stock i in month t+1, and the regressions use the variables MAX, beta, company size, book-to-market ratio, reversal and illiquidity lagged by one month. Momentum is calculated over an 11-month period with *a* 2-month lag.

By using this other method of analysis, the conclusions also corroborate the results obtained previously and are robust to the exclusion of outliers, providing relevant evidence that there is a negative relationship between positively distorted returns and future returns, demonstrating the effect of idiosyncratic lottery payoffs on pricing of stocks.

Some critics may see the MAX not as a measure of the phenomenon of stocks with a right-skewed return distribution, but rather as a reformulation of the attention bias. Thus, they can try to explain the abnormal returns of stocks with high MAX not by preference for asymmetry, but based on behavioral biases of social interactions and attention: those stocks that are focused on the media, are subject of news and discussion agendas in investor social circles and are therefore more likely to be bought. Thus, stocks with a high MAX would tend to have greater social and media attention and, therefore, would have this abnormal return. However, Bali, Hirshleifer and Tang (2021) demonstrate that MAX is also robust to the attention bias effect (although possibly part of its effect is caused by it).

Furthermore, taking into account that the public attracted by *lotteries* are, in short, individual investors, *skewness* may not represent the most relevant indicator of demand for lottery shares, as it is difficult to identify. Complementarily, Barberis and Huang (2008), based on the *Cumulative Prospect Theory (CPT)* of Tversky and

Kahneman (1992), state that it is the extreme returns and low chance states that cause the pricing effects, not the asymmetry of the distribution itself. Brunnermeier, Gollier and Parker (2007) corroborate this belief. In view of this, the use of past *skewness* may no longer make sense. On the other hand, MAX - an idiosyncratic metric that suggests a large potential daily return in the short term (mainly in the scenario where investments are under-diversified) - does not suffer from these oppositions. Moreover, according to the study by Bali, Cakici and Whitelaw (201 1), the distortion of returns caused by the MAX appears to be more statistically and economically significant than the past asymmetry. For this reason, the MAX was chosen as one of the indicators to guide this study in Brazil.

Another indicator capable of classifying stocks in lottery or anti-lottery is the Nominal Price. Kummar (2009) states that investors tend to be attracted by low-priced assets due to the search for cheap bets, an idea similar to that developed by Green and Hwang (2009), who conclude that this price, perceived as very close to zero, to the eyes of an individual may appear to have a great deal of room to rise and little to fall. Birru and Wang (2012) found evidence of this behavior when measuring the expected skewness of several securities using options days before and days after a split and identified a sharp and significant increase in investor expectations. Birru and Wang (2016) also showed, in line with what is expected, a negative premium for companies with low nominal prices, that is, this type of asset has lower risk-adjusted returns than stocks with low nominal prices. high nominal prices.

Given this empirical evidence of the anomaly, many studies focused on understanding human preference for lotteries or, more specifically, understanding why people continue to invest in lottery-stocks even though they deliver a substantially higher risk-return ratio. lower than other stocks.

One of the reasons that explains this preference for asymmetry is the optimism bias (overconfidence). Brunnermeier and Parker (2005) show evidence that investors overestimate the return on investments in lottery stocks, confident in the belief that they will be the next “drawn by the lottery”. In addition, Kumar (2009) assumes that the greater volatility characteristic of the asset makes the investor infer with greater probability that extreme returns will occur in the future. Brunnermeier, Gollier, Parker highlight the strength of bias by adding a vicious cycle to the theory:

> "[...] there is a natural complementarity between believing a state more likely and purchasing more of the asset that pays off in that state. Once a state is perceived as more likely, one wants more consumption in that state, and once one has more consumption in that state, one wants even more to believe that state is more likely." (BRUNNERMEIER; GOLLIER; PARKER, 2007, p.2).

These behaviors show signs of a certain human subjectivity in the valuation of gains/losses and in the understanding of estimated probabilities. Kahneman (1992) supports this view by bringing advances to the *Prospect Theory* and developing the *Cumulative Prospect Theory (CPT)*, which adds cumulative weighting functions in the calculation of the utility of individuals, a function that weights the estimated probabilities in a non-linear and not well-behaved way, especially at the extreme ends of the distribution. As a result, individual investors make decisions not based on estimated probabilities, but based on probabilities transformed by the additivity function described by the author. Thus, small chances can be interpreted as more likely than they actually are, influencing the individual to take more risk than expected to obtain a lower return expectancy, an inversion of the typical "risk aversion", found in most part (well - behaved) of the CPT.

Other possible explanations developed by Brunnermeier, Gollier and Parker (2007) revolve around the discounted present value of the utility maximization of individual investors at each moment in time. They argue that the utility of individuals does not depend solely on current wealth, but also on expected future wealth, discounted to present value. This implies that, even if the alpha expectancy when investing in lottery-like assets is negative, investing in this type of asset allows individuals to believe that they will win, increasing the utility in the present moment. However, this increase in present utility, as demonstrated by empirical research, does not come without costs, in a way that leads to a suboptimal ex post decision, even resulting in a lower average future utility.

Although the real reasons behind this preference for lottery stocks are still not entirely clear, it is a fact that, regardless of this, its effect is robust and negatively affects average investment returns. Despite all the empirical evidence that corroborates these conclusions in the United States, Brazil lacks studies regarding this theme. Therefore, in the next section, the methodologies used to analyze lottery stocks in the Brazilian stock market will be explained.

3. METHODOLOGY AND DATA

Given the context set out in the previous section, this paper aims to introduce the study of lottery stocks in Brazil and, more precisely, identify whether these assets have a lower expected return than other stocks, as predicted by the literature.

For this, the MAX and Nominal Price factors will be used to identify stocks with a lottery profile, with the MAX defined as the highest daily return of a stock in the month prior to the portfolio establishment date multiplied by -1. The Nominal Price is defined as the average nominal price of the share also in the last month. For both indicators, the smaller, the more lottery-like the stock is. MAX was chosen because, as presented in the Literature Review, it tends to capture the effect that the study proposes to analyze (attraction to extreme returns of low probability) better than asymmetry. Some other points that make the MAX convenient for this investigation are: i) its idiosyncratic aspect, which allows observing the market in a context in which investors do not diversify their portfolios; ii) be more robust (statistically significant) than the past skewness*;* iii) be easily identifiable by investors; iv) its proven persistence over time, so that individuals who seek high MAX in their investments actually achieve this goal by investing in high MAX stocks. The Nominal Price, in turn, was selected because it indicates - in a misleading way - a greater margin for the appreciation than for the depreciation of an asset, in other words, an expected asymmetry to the right. Moreover, according to some authors, such as Bali, Cakici and Whitelaw (2011), the Nominal Price is robust in relation to several factors such as market capitalization, liquidity, short-term reversal, momentum and the price/book value ratio.

The stocks considered available for investment and, therefore, eligible to compose the portfolios in this analysis, will be the securities that, in the month prior to each portfolio rebalancing date, are in the composition of the IBRX-100 (Brazil 100 Index) or in the SMLL (Small Cap Index). The first index was selected due to its content: the 100 most tradable and representative assets in the Brazilian stock market; the second was chosen because it encompasses companies with smaller capitalization, since these, as previously presented, tend to be more likely to behave like lotteries. Thus, the idea of the database, which will be obtained from Economatica, is to contain both the largest and the smallest and most liquid stocks, covering a large part of the Brazilian stock market and excluding illiquid companies.

Daily adjusted price, nominal price and traded volume data will also be obtained from the Economatica database for the period from 08/01/2008 to 12/31/2021, totaling 161 months. The period of analysis was chosen in view of the availability of information on the Small Caps Index, existing from 2008 to the present moment in which this article is written. Selic yield data (risk-free rate for this research) will be collected from the Central Bank of Brazil website and data from Ibovespa returns (used to calculate the alpha of the CAPM model) will be collected from Investing.

Initially, the study will make a brief analysis of portfolios formed by percentiles of each of the indicators, in order to evaluate them separately. Percentiles below 10 will be chosen to identify lottery stocks (Short) and above 90 to identify anti-lottery stocks (Long). To analyze the Long and Short portfolios formed monthly based on the MAX and Nominal Price factors, a series of indicators will be calculated:

- Value (average): the average value of the indicators;
- Value (standard deviation): the standard deviation of the indicators value;
- No. of Shares (average): the average number of shares per portfolio;
- No. of Shares (standard deviation): the standard deviation of the number of shares per portfolio;
- Return (average): the average return on the portfolio;
- Return (Standard Deviation): The standard deviation of the portfolio returns;
- Difference Return (average): the difference between the average return between portfolios;
- Return Difference (Standard Deviation): the standard deviation of the portfolios' difference in return;
- Difference Return (t-value): the average difference in return of the portfolios divided by the standard deviation of this difference, multiplied by the square root of N, in this case, 161;
- Sharpe: portfolio return minus the Selic divided by the standard deviation of the return minus the monthly Selic;
- Asymmetry: the skewness of the portfolios' monthly returns as described by equation (4);

- Kurtosis: the kurtosis of the monthly returns of the portfolios according to the formula below:

$$K = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}}{s^4}$$

where s means the sample variance, x_i and $\bar{x}$ the observation return i and the average return, respectively, n the number of observations and K the kurtosis;

- Excess Market Return: difference between the market return, measured by the Ibovespa return, and the risk-free rate in Brazil, given by the Selic, according to the expression:

$$R_m = r_m - r_f$$

where r_m represents the market return, r_f the risk-free interest rate and R_m the excess market return.

- Excess Portfolio Return: difference between the portfolio return and the risk-free interest rate (Selic), according to the formula:

$$R = r - r_f$$

where r_f represents the risk-free interest rate, r means the portfolio return and R the excess of the portfolio return.

- Beta: sensitivity of the portfolio return to the market return, measured from the operation:

$$\beta = \frac{cov(R, R_m)}{\sigma^2(R_m)}$$

where R and R_m represent the excess return of the portfolio and the market, respectively, and β the sensitivity.

- Expected Return by the CAPM model, given by:

$$E(R_{CAPM}) = \beta * R_m + r_f$$

with β regard to the sensitivity of the portfolio's return to the market return, R_m to the excess market return and $E(R_{CAPM})$ the expected yield according to the CAPM hypotheses;

- Alpha CAPM (mean): the difference between the portfolio return (r) and the return expected by the CAPM model ($E(R_{CAPM})$):

$$\alpha = r - E(R_{CAPM})$$

- CAPM alpha (standard deviation): the standard deviation of the CAPM alpha (mean);
- CAPM alpha difference (mean): mean difference between the alpha of one portfolio and the other ($\bar{\alpha}$), according to the equation below, where α_i(α_j) represents the CAPM alpha given portfolio i (j), as shown below:

$$\bar{\alpha} = \alpha_i - \alpha_j$$

- CAPM Alpha Difference (Standard Deviation): The average deviation of monthly CAPM alpha differences across portfolios ($\bar{\sigma}$), as per the formula:

$$\bar{\sigma} = \sqrt{\frac{\sum((\alpha_i - \alpha_j) - (\bar{\alpha}_i - \bar{\alpha}_j))^2}{n}}$$

where α_i(α_j) represents the alpha measured by the CAPM model given portfolio i (j);

- CAPM alpha difference (t-value): the average CAPM alpha difference between portfolios divided by the standard deviation of these differences multiplied by the root of N, in this case, root of 161.

All indicators that involve price variations in their calculation will use the difference of the natural logarithm of prices in order to normalize the results and simplify the modeling.

Then, the study will consist of a joint analysis of the factors, also forming lottery and anti-lottery portfolios. For an asset to qualify as lottery (anti-lottery), it must be evidenced at the Kth percentile of the highest (lowest) MAX and the Kth percentile of the lowest (highest) nominal price, simultaneously in the month prior to the rebalancing. This process will be redone for each of the months in the sample. The focus of the investigation will consist in comparing the behavior of the monthly performance of the two portfolios with a K = 30. The same indicators will be calculated for these new portfolios.

Complementarily, a multivariate analysis will also be performed using the five-factor model created by Fama and French (2014) using data provided by NEFIN *(Núcleo De Pesquisa em Economia Financeira - USP).* The objective is to understand if the MAX, the Nominal Price and the factor that combines the two indicators (Comb) are able to explain the monthly returns of portfolios controlled by simultaneous quartiles of Size (1: Smallest; 2: Largest) and Illiquidity (1: Minor; 2: Major) (2x2, respectively). The shares eligible for the calculation of these additional factors and for composing the portfolios must be traded on Bovespa, must be the company's most liquid stock and, in the last year, must have been traded on at least 80% of the days with a volume greater than 500,000 reais per day. Portfolios are rebalanced monthly, assigning equal weights to all stocks.

4. RESULTS

When analyzing Table 1, it is possible to observe that the comparison between the first and last decile was able to separate the groups of stocks with statistically different MAX values. When analyzing the number of stocks, we clearly noticed a limitation of the study of factors when applied to the Brazilian reality. While the American studies cited in the bibliographic review had a universe of about three thousand stocks, this study in Brazil has an Investable Universe close to 150.

With a universe of assets of this size, the portfolios were left with, on average, around 14 stocks, which can distort the values of the indicators because there is not a complete and sufficient diversification of the idiosyncratic volatility. Thus, a single stock could stochastically distort the results found in order to generate a false positive or a false negative in the search. Making clear this limitation of the results found, it is possible to continue the analysis observing the return differential between the Long and Short portfolios. The difference is statistically significant when using a one-tailed t-test with a significance of 2.5%, with a monthly return differential greater than 1.2%

Table 1 - Monthly Results of the MAX factor

Indicator	Long	Shorts
Value (average)	-0.0200	-0.1097
Value (standard deviation)	0.0077	0.0359
No. of Shares (average)	13.9627	13.9627
No. of Shares (standard deviation)	1.5032	1.5032
Return (average)	0.0017	-0.0109
Return (standard deviation)	0.0561	0.1029
Difference Return (average)	0.0127	-0.0127
Difference Return (standard deviation)	0.0707	0.0707
Return Difference (t-value)	2.2715	-2.2715
Sharpe	-0.0963	-0.1751

Asymmetry	-2.3471	-0.7201
kurtosis	13,9994	4.4744
alpha CAPM (average)	-0.0031	-0.0137
alpha CAPM (standard deviation)	0.0317	0.0577
CAPM alpha difference (average)	0.0107	-0.0107
CAPM alpha difference (standard deviation)	0.0589	0.0589
CAPM alpha difference (t-value)	2.2981	-2.2981

Source: own elaboration based on data from Investing, Economatica and Banco Central

per month. Another interesting point is that the asymmetry of the Short portfolio is greater than that of the Long portfolio and the kurtosis of the Short portfolio is smaller than that of the Long portfolio, indicating, as expected, thicker tails in the distribution of lottery stocks and a relative distortion of returns to the right. The alpha difference between the portfolios measured by the CAPM model is also statistically significant when we use a one-tailed t-test at a level of 2.5%. As lottery stocks tend to have a higher beta, the t-value of the CAPM is, as expected, greater than the t-value of the difference in returns. There is also a greater Sharpe ratio observed in the Long portfolio relative to the Short and a volatility almost twice as high in the Short portfolio relative to the Long.

Table 2 - Monthly Results of the Nominal Price factor

Indicator	**Long**	**Shorts**
Value (average)	71.7290	3.4583
Value (standard deviation)	25.5101	1.3771
No. of Shares (average)	13.9627	13.9627
No. of Shares (standard deviation)	1.5032	1.5032
Return (average)	0.0049	-0.0165
Return (standard deviation)	0.0572	0.1151
Difference Return (average)	0.0215	-0.0215
Difference Return (standard deviation)	0.0871	0.0871
Return Difference (t-value)	3.1251	- 3.1251

Sharpe	-0.0386	-0.2053
Asymmetry	-2.0432	-0.1977
kurtosis	8.9003	3.3159
alpha CAPM (average)	0.0001	-0.0191
alpha CAPM (standard deviation)	0.0357	0.0725
CAPM alpha difference (average)	0.0192	-0.0192
CAPM alpha difference (standard deviation)	0.0748	0.0748
CAPM alpha difference (t-value)	3.2533	-3.2533

Source: own elaboration based on data from Investing, Economatica and Banco Central

When analyzing Table 2, it is possible to observe that the comparison between the first and last decile was able to separate the groups of shares with statistically different Nominal Price values. The difference in returns is statistically significant when using a one-tailed t-test with a significance of 0.5%, with a monthly return differential greater than 2% per month, much greater than the MAX indicator and the premiums found in the United States in the previously mentioned surveys. Another interesting point is that, once again, the asymmetry of the Short portfolio is greater than that of the Long portfolio and the kurtosis of the Short portfolio is smaller than that of the Long portfolio, indicating, as expected, thicker tails in the distribution of lottery stocks and a distortion relative right returns in the Short versus Long portfolio. The alpha difference between the portfolios, measured by the CAPM model, is also statistically significant when we use a one-tailed t-test at a level of 0.5%. As lottery stocks tend to have a higher beta, the t-value of the CAPM is, as expected, greater than the t-value of the difference in returns.

Table 3 - Monthly Results of the Combined Model

Indicator	**Long**	**Shorts**
MAX value (average)	-0.0266	-0.0903
MAX value (standard deviation)	0.0106	0.0314
Value Price (average)	49.1102	4.9756
Value Price (standard deviation)	6.8620	1.9919
No. of Shares (average)	12.3975	12.4099
No. of Shares (standard deviation)	1.4154	1.4250
Return (average)	0.0042	-0.0150

Return (standard deviation)	0.0561	0.1129
Difference Return (average)	0.0193	-0.0193
Difference Return (standard deviation)	0.0861	0.0861
Return Difference (t-value)	2.8387	- 2.8387
Sharpe	-0.0518	-0.1962
Asymmetry	-3.2971	-0.5307
kurtosis	22.6339	3.4370
alpha CAPM (average)	-0.0006	-0.0176
alpha CAPM (standard deviation)	0.0339	0.0673
CAPM alpha difference (average)	0.0169	- 0.0169
CAPM alpha difference (standard deviation)	0.0729	0.0729
CAPM alpha difference (t-value)	2.9478	- 2.9478

Source: own elaboration based on data from Investing, Economatica and Banco Central

There is also a greater Sharpe ratio observed in the Long portfolio relative to the Short and an almost 100% greater volatility in the Short portfolio relative to the Long.

When analyzing Table 3, it is possible to observe that the combination of the two factors was able to separate the groups of shares with statistically different MAX and Nominal Price values. The difference in returns is statistically significant when using a one-tailed t-test with a significance of 0.5%, with a monthly return differential of almost 2% per month, much higher than the premiums found in the United States in the previously mentioned surveys. Another interesting point is that, again, the asymmetry of the Short portfolio is greater than that of the Long portfolio and the kurtosis of the Short portfolio is smaller than that of the Long portfolio, indicating, as expected, thicker tails in the distribution of lottery stocks and a relative distortion of the returns to the right in the Short portfolio in relation to the Long. The alpha difference between the portfolios, measured by the CAPM model, is also statistically significant when we use a one-tailed t-test at a level of 0.5%. As lottery stocks tend to have a

Table 4 - Nominal Price Factor controlled by Size and Illiquidity (2x2)

Variable	**(1x1)**	**(1x2)**	**(2x1)**	**(2x2)**
constant	0.0116***	0.0047***	0.0006	0.0015
	(0.0040)	(0.0015)	(0.0013)	(0.0031)
MKT	0.9688***	0.8910***	0.9505***	0.9479***
	(0.0765)	(0.0284)	(0.0243)	(0.0602)
SMB	0.6462***	0.3762***	0.0754	-0.2418*

	(0.1795)	(0.0666)	(0.0569)	(0.1411)
HML	-0.2332**	-0.0883**	-0.0227	-0.0737
	(0.1115)	(0.0414)	(0.0354)	(0.0877)
IML	-0.0823	-0.0361	-0.0110	0.2580***
	(0.0877)	(0.0325)	(0.0278)	(0.0689)
WML	-0.3333*	0.4739***	-0.0012	0.7988***
	(0.1781)	(0.0661)	(0.0565)	(0.1401)
Pnom	-0.2958***	-0.0733**	0.0168	0.0580
	(0.0786)	(0.0292)	(0.0249)	(0.0618)

Source: own elaboration based on data from Investing, Economatica and Nefin
Significance Code: 0 '***' 0.001 '**' 0.01 '*'

higher beta, the t-value of the CAPM is, as expected, greater than the t-value of the difference in returns. There is also a greater Sharpe ratio observed in the Long portfolio relative to the Short and an almost 100% greater volatility in the Short portfolio relative to the Long. From what can be seen, the Nominal Price is more significant than the MAX and the combination of the two indicators reduces the anomaly of the results of the portfolios obtained from the Nominal Price.

4.1 FAMA-FRENCH RISK PREMIUM ANALYSIS

Extending the Fama-French pricing model and dividing portfolios based on market capitalization and companies' illiquidity, it is possible to obtain the results for each of the three indicators used to obtain the Lottery Factor premium, reported in Table 4 (Factor Nominal Price controlled by Size and Illiquidity (2x2)), Table 5 (MAX

Table 5 - MAX Factor controlled by Size and Illiquidity (2x2)

Variable	(1x1)	(1x2)	(2x1)	(2x2)
constant	0.0080**	0.0038**	0.0012	0.0032
	(0.0040)	(0.0015)	(0.0012)	(0.0030)
MKT	0.9841***	0.8963***	0.9394***	0.9187***
	(0.0813)	(0.0295)	(0.0246)	(0.0612)

SMB	0.7895***	0.4157***	0.0396	-0.3402**
	(0.1854)	(0.0674)	(0.0562)	(0.1396)
HML	-0.0837	-0.0492	-0.0453	-0.1390*
	(0.1080)	(0.0392)	(0.0327)	(0.0813)
IML	-0.2007**	-0.0658**	-0.0018	0.2875***
	(0.0846)	(0.0308)	(0.0257)	(0.0637)
WML	-0.2591	0.4905***	0.0066	0.8146***
	(0.1862)	(0.0677)	(0.0565)	(0.1403)
max	-0.0875	-0.0163	-0.0327	-0.0787
	(0.0725)	(0.0263)	(0.0220)	(0.0546)

Source: own elaboration based on data from Investing, Economatica and Nefin
Significance Code: 0 '***' 0.001 '**' 0.01 '*'

Factor controlled by Size and Illiquidity (2x2)) and Table 6 (Combined Factors controlled by Size and Illiquidity (2x2)).

As the Tables 5-7 show, the Nominal Price indicator presented negative coefficients with statistical significance lower than 0.001 for the larger portfolios and positive coefficients for the larger ones, which was expected, since companies with larger capitalization tend to be anti-lottery in their composition (KUMAR, 2009). For the combination of indicators, the same analysis could be performed, however, for MAX alone, no conclusion could be reached since the indicator presented negative coefficients for all four portfolios without statistical significance in any of them.

Table 6 - Combined Factors controlled by Size and Illiquidity (2x2)

Variable	(1x1)	(1x2)	(2x1)	(2x2)
constant	0.0089**	0.0043***	0.0012	0.0026
	(0.0040)	(0.0014)	(0.0012)	(0.0031)
MKT	0.9904***	0.8943***	0.9455***	0.9381***
	(0.0786)	(0.0285)	(0.0241)	(0.0601)

SMB	0.6977***	0.3701***	0.0391	-0.3015**
	(0.1920)	(0.0695)	(0.0589)	(0.1468)
HML	-0.1296	-0.0721*	-0.0454	-0.1191
	(0.1105)	(0.0400)	(0.0339)	(0.0845)
IML	-0.1447	-0.0441	0.0058	0.2899***
	(0.0891)	(0.0322)	(0.0273)	(0.0681)
WML	-0.2289	0.5067***	0.0053	0.7967***
	(0.1854)	(0.0671)	(0.0569)	(0.1417)
Combination	-0.1505**	-0.0555**	-0.0236	-0.0184
	(0.0738)	(0.0267)	(0.0227)	(0.0564)

Source: own elaboration based on data from Investing, Economatica and Nefin
Significance Code: 0 '***' 0.001 '**' 0.01 '*'

In addition to the limitation generated by the small number of shares available for trading in Brazil (which leads to idiosyncratic volatility in portfolios), there is another difficulty in finding the significance of the factors: the size of the trading history available. While in the United States surveys have more than 70 years of data available, the Brazilian stock market has developed recently, so this survey has about 14 years of history. With a smaller number of observations and a larger standard deviation due to more concentrated portfolios, the statistical significance of the indicators is reduced. Even so, it was possible to find significance in the tested indicators, especially the Nominal Price and the combination of this indicator with the MAX. Possibly, this anomaly will be persistent in Brazil due to the higher transaction costs present in the Brazilian stock market when compared to the US stock market and, therefore, greater difficulty in selling short lottery shares in order to profit by correcting their inflated price.

5. CONCLUSION

Lotteries, contrary to what would be expected by traditional utility models under uncertainty, arouse the interest of many human beings. In the North American financial market, this phenomenon could be observed in several contexts, with individuals exchanging – consciously or not – expectancy of risk-adjusted return on their investments for expectancy of asymmetrical returns to the right in their portfolios, even if this meant incurring more risks and lower average returns. In this study, an analysis of the expected return of *lottery-stocks* in the Brazilian stock market was tested for the first time.

It was shown that lottery stocks have significantly negative returns, volatility-adjusted returns and alphas measured by the CAPM model, in all lottery definitions used in this research. There is an alpha differential of about 2% per month between lottery and anti-lottery stocks, so this bias can cause significant damage to the portfolio of investors who love extreme and asymmetric to the right returns. Furthermore, starting from the five-factor model of Fama and French (2014) and adding to this, a sixth factor, the lottery prize in Brazil, it was possible to obtain significance with the Nominal Price indicators and their combination with the MAX for the portfolios of low market capitalization.

In the future, new studies could be carried out using other indicators jointly or separately as a lottery definition, for example, idiosyncratic volatility, asymmetry or kurtosis. It could also be studied how risk factors behave in universes of few available stocks and short history of time, in addition to how to deal with these cases of small samples. In this sense, another suggestion for investigation would be ways of optimizing the tradeoff between concentrating the portfolio by obtaining a more extreme average factor at each of the Long and Short ends (but with a higher standard deviation in returns) versus selecting broader percentiles, with greater diversification, but with the selected factor less extreme and, therefore, less emphasized, at each end of the portfolio.

6. REFERENCES

AGARWAL, Vikas; JIANG, Lei; WEN, Quan. Why Do Mutual Funds Hold Lottery Stocks? **Journal of Financial and Quantitative Analysis** , v.57, n.3, p.825-856, 2022.

AKBAS, Ferhat; GENE, Egemen. Do Mutual Fund Investors Overweight the Probability of Extreme Payoffs in the Return Distribution? **Journal of Financial and Quantitative Analysis** , v. 55, n.1, p.223-261, 2020.

BALI, Turan et al. A Lottery Demand-Based Explanation of the Beta Anomaly. **Journal of Financial and Quantitative Analysis** , v.52, n.6, p.2369-2397, 2017.

BALI , Turan et al. Attention, Social Interaction, and Investor Attraction. 2021.

BALI, Turan; CAKICI, Nusret; WHITELAW, Robert. Maxing out: Stocks as lotteries and the cross-section of expected returns. **Journal of Financial Economics** , v.99, n.2 p.427-446, 2011.

BARBERIS, Nicholas; HUANG, Ming. Stock as Lotteries: The Implications of Probability Weighting for Security Prices. 2004. **American Economic Review,** v.98, n.5, p.2066-2100, 2008.

BOYER, Brian et al. Expected Idiosyncratic Skewness. **The Review of Financial Studies** , v.23, n.1, 2010.

BRUNNERMEIER, Markus; GOLLIER, Christian; PARKER, Jonathan. Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns. **American Economic Review** , v.97, n.2, p.159-165, 2007.

BRUNNERMEIER, Markus; PARKER, Jonathan. Optimal Expectations. **American Economic Review,** v.95, n.4, p.1092-1118, 2005.

CHIU, Henry. Skewness Preference, Risk Taking and Expected Utility Maximisation. **The Geneva Risk and Insurance Review** , v.35, p.108-129, 2010.

CONRAD, Jennifer; KAPADIA, Nishad; XING, Yuhang. Death and jackpot: Why do individual investors hold overpriced stocks? **Journal of Financial Economics** , v.113, n.3, p.455-475, 2014.

ERAKER, Bjorn. READY, Mark. Do Investors Overpay for Stocks with Lottery-like Payoffs? An Examination of the Returns of OTC Stocks. **Journal of Financial Economics,** v.115, n.3, p.486-504, 2015.

FAMA, Eugene; FRENCH, Kenneth. A five-factor asset pricing model. **Journal of Financial Economics,** v.116, n.1, p.1-22, 2014.

FAMA, Eugene; FRENCH, Kenneth. The CAPM: Theory and Evidence. **Journal of Economic Perspectives**, v.18, n.3, p.25-46, 2004.

FAMA, Eugene; FRENCH, Kenneth. The Cross-Section of Expected Stock Returns. **Journal of Finance,** v.47, n.2, p.427-465, 1992.

KRAUS, Alan; LITZENBERGER, Robert. Skewness Preference and the Valuation of Risk Assets. **Journal of Finance**, v.31, n.4, p.1085-1100, 1976.

KUMAR, Alok. Who Gambles in the Stock Market? **Journal of Finance,** v.64, n.4, p.1889-1933, 2009.

SHARPE, William. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. **Journal of Finance,** v.19, n.3, p.425-442, 1964.

TVERSKY, Amos; KAHNEMAN, Daniel. Advances in Prospect Theory: Cumulative Representation of Uncertainty. **Journal of Risk and Uncertainty,** v.5, n.4, p.297-323, 1992.