2020

Noise Traders in Large-cap and Small-cap Portfolios: Impact of Sentiments on the Mispricing

Eunjun Choo
Oberlin College

Noise Traders in Large-cap and Small-cap Portfolios:

Impact of Sentiments on the Mispricing*

Eunjun Choo

Oberlin College Honors Economics Thesis

May 4, 2020

Abstract

This paper analyzes the impact of "irrational" investor's sentiment on the abnormal returns of low and high cap stock portfolios. The "rational" and "irrational" sentiments are constructed using asset pricing fundamentals and the AAII sentiment survey data. The abnormal returns are calculated as the difference between the actual and Fama-French model returns. I note that due to higher limits of arbitrage for the small cap stocks, the main effect of the "rational" and "irrational" sentiments on the small-cap portfolio seems stronger than on the large-cap portfolio. Moreover, I note that the mispricing of the large-cap stocks seems to revert to its fundamental value faster than the low cap stocks, supporting the theory of greater limits of arbitrage in a small-cap market than the large-cap market.

Furthermore, I discover that the "irrational" sentiment has statistically significant interaction with the previous week's large-cap mispricing on both the current week's small and large cap mispricing. I attribute this statistical significance to the broader visibility of the large-cap stocks compared to the small-cap stocks. However, my interpretation of the mispricing interaction would need to be further tested with the market volume movements.

*I would like to thank the Oberlin Economics Department for this valuable research opportunity. Specifically, I would like to thank Professor Cotter for the model suggestions, Professor Duca for the draft feedback, and Professor Craig for not only the draft feedback but many office hours of guidance.

1 Introduction

In theoretical finance, many models, such as the Capital Asset Pricing Model, assume that investors are rational, and asset prices are efficient. However, market anomalies such as the Dot-com bubble of the 1990s and the price deviations of twin shares like Royal Dutch/Shell Group in the 1990s raise doubts about the efficient market hypothesis, because there are empirical instances of returns based on not on fundamental factors.

Therefore, instead of assuming the market efficiency, researchers such as Delong, Shleifer, Summers, and Waldmann (1990) attempted to explain such anomalies by relaxing the assumption of rational investors and instead introduce the model of both informed and uninformed investors. In their model, they illustrated how the uninformed investors with misleading or no information on the asset could bring more noise and volatility to the market. Thus, since the uninformed investor's influence on the fundamental price deviations is unpredictable, the rational investors–at least in the short run–face the "noise trader risk": the risk of price further deviating from the rational investor's calculation.

There have been arguments against the significant impact of noise trading from economists scholars such as Milton Friedman (1953). He argues that the effect of irrational agents is negligible. If the price of the risky asset were not equal to the present value that reflects all of the fundamentals, the rational investors would trade against the irrational agents and drive the price back to its fundamental value. Therefore, the investors who mistakenly judge the price of the risky asset can lose to the arbitrageurs, eventually disappearing from the market. However, De Long et. al (1990) challenges this theory by presenting a model of irrational agents–driven by its sentiment rather than the fundamental information–impacting the price and volatility of the risky asset.

This paper attempts not to limit its analysis on De Long et al. (1990) model. Instead, I will focus on the impact of the rational individual investors and the noise traders on the mispricing. The variable for the rational individual investors is the sentiment survey index explained by the asset pricing fundamentals, while the noise traders would be the unexplained residuals. Furthermore, the Fama-Macbeth rolling regression was used to predict the Fama-French model's theoretical price. Then, I construct the mispricing variable by using the difference between the actual and constructed theoretical price.

However, the Fama-French predictions and the sentiment survey constructions are one of the many ways to measure the deviations from the fundamental price and noise trading. The regressions on the noise trader's impact on the mispricing may be difficult to interpret on its own. Therefore, this paper focuses on comparing the noise trader's effect on the mispricing of the low and high market cap stock portfolios instead of a single market portfolio. This focus allows the interpretation of the noise trader to not solely focus on the coefficients of the sentiment on the mispriced returns but instead on the varying effects of the noise trader on different market cap portfolio.

Podolski et al. (2009) have concluded that small-cap stocks were prone to noise trader risk compared to the large-cap stocks. Thus, not only do the two portfolio mispricing allow noise trader's relative impact on different market cap stocks, but also having two mispricings allow another examination to the conclusion from Podolski et al. (2009).

Section (2) introduces a brief survey of the theory and the empirical studies of the noise traders and mispricing to motivate the research on noise trader's impact on mispricing. Section (3) introduces the methodology and the construction of the noise trader and mispricing variables. Section (4) describes the data of the constructed variables and the data corrections. Section (5) describes the result of the noise trading on different market cap portfolios and the result's implications. Finally, section (6) concludes with comments for further research.

2 Literature Review

2.1 Noise trader

Ramiah et al. (2015) defines a noise trader as "a market participant who makes investment decisions without the use of finance fundamentals, exhibits poor market timing, follows trends and tends to overreact or underreact to good and bad news." Black (1986) describes the noise trader as traders who "trade on noise as if it were information." These definitions of noise trader imply that the noise trader acts on noisy information, which is not based on fundamental information.

While Black (1986) described the nature of noise traders, DeLong et al. (1990) present a model that formalized the noise trader behavior and its impact on returns & volatility. In their model, the noise traders as a group can influence the equilibrium of the stock price. The changes of noise traders' sentiment influence the price of the asset to deviate from its fundamental value, creating a systematic risk and higher price volatility in the market. This noise trader risk may reduce the attractiveness of the informed investors to carry out the arbitrage. Therefore, DeLong et al. (1990) argue that because of the additional risk that the uninformed (noise) traders expose to the market, the risk-averse arbitrageurs will hold less risky portfolios than they would have without the noise traders. Thus, if there are a significant group of noise traders dominating the market, the additional risk–and the returns–that the noise traders create in the market would benefit the noise traders if they are on average bullish and suffer if on average bearish. In other words, by driving out the arbitrageurs, the noise traders can benefit from their "bullish" space and their overpriced assets. Therefore, the noise traders may have positive relationship on the returns.

2.2 Firm Size difference

Because the rational investor needs access to information and stocks to carry out the arbitrage, the limited information and the market equity can impact the price of the asset. For example, case studies of anomalies like Palm/ 3Com share in the early 2000s illustrate that short sale restrictions can limit the arbitrage, resulting in further mispricing of assets. Since the short lending fees are dependent on the supply and demand, and small-cap stocks tend to be in a small market, the short sale market for the small-cap stocks tends to be less liquid than the large-cap stocks. Therefore, when there is a sudden shift of strong demand for short lending small-cap stocks, the costlier its lending fee, limiting the short-cap stocks to be priced according to its fundamentals (Thaler and Lamont., 2002).

This observation is supported in the empirical study of Jones and Lamont (2002). They found that small-cap stocks tend to have more short-sale constraints and, thus, more expensive to short. Furthermore, small-sized firms may be more costly to gather their fundamental information. Mitchell et al. (2002) argue that one of the most important factors that limit arbitrage is the cost of collecting the appropriate information. Since small firms tend to have less news coverage compared to large firms, small-cap stocks seem to have a higher cost of information. Therefore, the asymmetry of the information cost may limit the arbitrage opportunity in small market sized firm stocks.

Thus, I hypothesize that small-cap stocks are more vulnerable to noise trader risk. I predict that the correlation between the noise trader risk and the weekly returns is stronger for the small-cap stocks compared to the returns of the large-cap stock.

2.3 Measuring Investor Sentiment

Empirically, measuring the noise trader's behavior is ambiguous and difficult. Nevertheless, some studies like Lee et al. (1991) and Sias et al. (1995) indirectly used closed-end funds to study the noise trader's risk. Due to the nature of the closed-end funds, the intrinsic value of the closed-end fund share can be calculated and compared with the market price of the share. Thus, its difference can illustrate the closed-end fund's behavior of investors–including the noise traders–in the market. In contrast, other studies use the market sentiment as a proxy to noise trader's behavior. While there is still an on-going debate as to whether or not the market sentiment reflects the noise trader's behavior (Ramiah et al., 2015), number of researches used the market sentiment as a proxy to noise trading (Lee et al., 2002; Verma and Verma, 2005; Podolski et al., 2009; Qiang and Shu-e, 2009; Corredor et al., 2015).

Brown (1999) attempted to test the investor sentiment as a proxy to the noise trader's behavior. Brown argues that following the noise traders theory from DeLong et al. (1990), if the noise traders affect prices, the noisy signal[1] of the market is a sentiment, and the noise traders cause a systematic risk (volatility), then the sentiment should correlate with the asset's volatility. By using the volatility of the closed-end funds as a dependent variable and the market sentiment as one of the independent variables, Brown concluded that the deviation of the average investor sentiment is associated with the increases in the volatility of the closed-end funds. Thus, Brown (1990) implies that when the investors' sentiments are volatile, so does the number of trades in the closed ends funds, increasing the closed-end funds' volatility. Therefore, the investor sentiment is a valid measure of noise trader behavior.

Following Brown (1990)'s conclusion, I assume that the investor's sentiment is an appropriate characterization of the noise traders' behavior. There are indirect and direct methods to measure an investor's sentiment. The indirect method used by Baker and Wurgler (2006 and 2007), consisted of using macroeconomic factors. For the direct measure of investor sentiment, a handful of studies uses investor's intelligence data or the American Association of Individual Investors (AAII) surveys.

The investor's intelligence data is from the editors of Investors' Intelligence, which reads and rates over 100 advisory services on the market movements (Lee et al., 2002; Corredor et al., 2015). On the other hand, AAII data is the survey data from the American Association of individual investors. Before 2000, every week, AAII conducts the survey via mail by randomly selects from approximately 100,000 AAII members and asks the participants to predict the likely direction of the stock market during the six months. AAII

[1]From Fisher Black (1986), the noisy signal refers to the opposite of the information, false information, or hype of the assets instead of information of the fundamentals.

then measures the percentage of participants responding “up” as bullish, “down” as bearish, and “the same” as neutral (Corredor et al., 2015). However, the beginning of 2000, AAII conducted its survey on-line.[2]

The average AAII member is a male in his late-50s with a graduate degree. Moreover, over half of the AAII members have an investment with over $500,000. Therefore, the AAII sentiment survey represents the market sentiment of “hands-on” active individual investors.

Thus, investor’s intelligence is a measure of the market’s sentiment from analyzing a handful of analysts, while the AAII survey measures the sentiment directly from the investors. While Brown and Cliff (2005) and W. Y. Lee et al. (2002) used investor’s intelligence data, Verma and Verma (2006) and P. Corredor et al. (2015) used AAII as an investor’s sentiment and a proxy for the noise trader’s behavior. An indirect measure of sentiment was constructed by Baker and Wurgler (2006 and 2007). They constructed a sentiment variable by combining trading volume, dividend premium, closed-end fund discount, number and the first-day returns on IPOs, and the equity share in new issues.

However, there may be problems of using indirect sentiment measures along with other macroeconomic variables in a model. Specifically, in P. Corredor et al. (2015), they argued that when attempting to examine the role of investor sentiment on spot and futures markets, using macroeconomic data to indirectly measure the sentiment may bias the result since the macroeconomic variables would be used for independent (sentiment construction) and dependent variables (spots and futures returns) instead of actual behavior of the investors. Moreover, since P. Corredor et al. (2015) was interested in a variable that proxies the observation of noise trader behavior, the AAII’s nature of the target respondent and the survey construction makes AAII survey the preferred measure of the noise trader behavior. Furthermore, the previously mentioned study of investor sentiment as a proxy of noise trader risk (Brown, 1999) used AAII as investor sentiment.[3] Thus, the AAII survey seems to be an

[2]https://www.aaii.com/journal/sentimentsurveyarticle

[3]The assumption being that the volatility of the closed-end fund represents the systematic risk.

appropriate variable for investor sentiment and noise trader behavior.

However, AAII survey results contain individual investors, not necessarily the uninformed (noise trader) individual investor. Thus, studies like Brown and Cliff (2005), Verma and Verma (2005), Verma et al. (2008), and Sayim et al. (2013) controlled for the macroeconomic variables that may have influenced the investor sentiment. Since the noise trader is an investor without the fundamental knowledge of the market, only part of the overall investor sentiment may reflect the behavior of the noise trader. Therefore, Brown and Cliff (2005) regressed the investor sentiment variable along with popular variables in asset pricing literature that influence the market valuation, such as treasury bill returns and Fama french factors. Using the monthly intervals data, Verma and Verma (2005), Verma et al. (2008), and Sayim et al. (2013) were able to examine the impact of both the rational (fundamental and information-based) investor and irrational investor sentiment on asset returns by first regressing the sentiment variable on fundamentals as shown below:

$$Sent_t = \gamma_0 + \sum_{i=1}^{I} \gamma_i \ Fund_{i,t} + \xi_t \text{ , where} \tag{1}$$

$Sent_t$ is the investor sentiment data like AAII survey, γ_0 is the constant, γ_i is the parameter estimate of the impact on the fundamental variable to $Sent_t$, and ξ_t is the error term. After the regression, Verma and Verma (2005), Verma et al. (2008), and Sayim et al. (2013) used the sentiment estimate $\hat{Sent}_t$ as the rational investor and ξ_t as the irrational investor.

2.4 Measuring Mispricing

To calculate the mispricing of an asset, studies have tried to calculate a theoretical value based on stock's fundamentals or factor models. Studies like Doukas et al. (2010) and Aabo et al. (2017) measured the fundamental return of each stock rather than a portfolio by using the firm's fundamental information such as residual income value of the firm or book and net income value of the firm. On the other hand, studies such as Guidolin and

Ricci (2010) and Cao and Han (2016) utilized the Fama-French 3-factor model [4] to calculate the fundamental return of stocks. Afterward, they took the difference between the observed return of the stock and the Fama-French prediction to estimate the mispricing measure. This mispricing measure was used as the error of the factor pricing model in Joon and Yang (2016).

2.5 Contribution

This paper contributes to the growing literature of noise trader and market anomalies by examining the noise trader and the rational investor's impact on the mispricing of small-cap and large-cap portfolios. This approach differs from Podolski et al. (2009) since I am interested in examining the noise traders in United States with AAII survey data rather than Australia data with the broker's net initiated order flows as an investor's proxy.

Furthermore, while Podolski et al. (2009) and Qadan and Aharon (2019) studied the impact of sentiment on firm sizes and stock-cap sizes, respectively, the two studies did not separate the rational and irrational sentiments. However, taking into consideration that the individual investors may have rational and non-informed judgments, I use methods similar to Verma et al. (2008), Verma and Verma (2005), and Sayim et al. (2013) to separate the rational and irrational individual investors.

Studies have found that arbitrageurs care about their short-term investment performance (Shleifer and Vishny, 1995) and that the arbitrage is limited when arbitrageurs have shorter investor horizons than noise traders (DeLong et al., 1990). Thus, the noise trader's impact may be more evident during the frequent period (Cao and Han, 2016). Therefore, unlike Verma et al. (2008), Verma and Verma (2005), and Sayim et al. (2013), I examine the impact of noise traders on a weekly basis rather than monthly intervals.

Taking into consideration the noise trader theory, noise trader's impact on firm size difference, and the construction of our mispricing model, I hypothesize the following:

[4]For more information on Fama–French three-factor model, see appendix (9.1)

H1: The small-cap stock portfolio will have a higher deviation from its fundamental returns compared to the high-cap stock portfolio.

H2: The irrational investor sentiment will have positive association to the weekly mispriced returns

H3: The impact of irrational individual investor (noise trader) on mispricing will be more significant for the low-cap stock portfolio compared to the high-cap stock portfolio

Note that the *H2* does not mean the noise traders will earn higher or lower returns. Instead, the focus of this paper is restricted to the impact of portfolio's mispriced returns, not the expected value of noise traders or rational investors.

3 Model and Methodology

There have been several measures of mispricing in the literature. In T. Aabo et al. (2017), they have used several mispricing models on groups of stocks, such as utilizing residual income value of the firm or using the book and net income value of the firm. However, these mispricing measures are for the individual firm stocks rather than a portfolio. Since factor models can be used to valuate a stock or a portfolio, the Fama-French 3-factor model was chosen to model the fundamental price of the small and large-cap stock portfolios.

As described in Guidolin and Ricci (2010), the price predicted by the Fama-French model will be the price in the absence of arbitrage risk and other limits to arbitrage. While several studies have noted that each factor in the Fama-French model may constitute a market anomaly factor, the combination of the three factors to produce a single measure should reduce the residual noise and allow a more accurate measure of the theoretical price. Therefore, the Fama-French 3-factor model was used as an intrinsic rate of return for the small and large-cap stocks portfolios. The difference between the Fama-French prediction and the actual return is the mispricing—the return not explained by the market risk premium and the size and value premium.

Similar to Joon and Yang (2016), which tested the Fama French model's predictability strength, I use the Fama-French three-factor model to regress on previous years and use that regression coefficient to predict the current return. This way, the Fama-French β_j coefficients are adjusted depending on the time. Furthermore, this paper assumed that the mispriced returns is the difference between the observed returns and the returns that would prevail in the absence of noise traders.

3.1 Mispricing

The mispricing is defined as follows: for each year i, take the weekly portfolio returns R_i, risk-free rate R_f, and the Fama-French factors, which are R_{Mkt}, SMB, and HML from years $(i-6)$ to $(i-1)$. Then the theoretical Fama-French 3 factor model is regressed using OLS as

$$R_{i,t} - R_{f,t} = \beta_{1,t}\ (R_{Mkt,t}, R_{f,t}) + \beta_{2,t}\ SMB_t + \beta_{3,t}\ HML_t + \epsilon_t, \tag{2}$$

which would result in the coefficient estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ from the years $(i-6)$ to $(i-1)$ sample. For the year i, the estimated Fama-French return is

$$\overline{R_{i,t} - R_{f,t}} = \hat{\beta}_1\ (R_{Mkt,t} - R_{f,t}) + \hat{\beta}_2\ SMB_t + \hat{\beta}_3\ HML_t, \text{ where} \tag{3}$$

$\overline{R_{i,t} - R_{f,t}}$ is the market excess expected return.

The mispricing, or the difference between the actual and the predicted, is

$$Mispriced = (R_{i,t} - R_{f,t}) - \overline{R_{i,t} - R_{f,t}}\ . \tag{4}$$

Note that the intercept is not included in the Fama French regression (2) since we are interested in the abnormal return of the portfolio. Thus, α value would be included in the $Mispriced$ variable. This is a process also used by researchers such as Chae and Yang (2016).

The weekly returns instead of monthly or daily returns were used as our metric since according to Cao and Han (2016), Koller et al. (2010), and Koller et al. (2010), daily returns may lead to an extreme returns in certain days influencing the returns while the monthly returns may not represent the frequent investor's sentiment affecting the future returns.

Moreover, five years of weekly returns were used for the Fama-French B_i estimation, since according to Cao and Han (2016) and Koller et al. (2010), five years of data was chosen as a compromise between precise estimates and to avoid too recent systematic biases.

This paper will have two different mispriced variables: $MispricedL20_t$ for small-cap stocks portfolio and $MispricedH20_t$, for a large-cap stocks portfolio.

3.2 Rational and Irrational sentiments

A similar process from Verma and Verma (2005), Verma et al. (2008), and Sayim et al. (2013) was used to construct the sentiment variables. Using an investor survey variable $Sent$ and I number of fundamental information, $Fund_i$ for $i = 1, ..., I$, the rational and irrational sentiments were modeled using the OLS regression as the equation (1):

$$Sent_t = \gamma_0 + \sum_{i=1}^{I} \gamma_i \; Fund_{i,t} + \xi_t \; . \tag{5}$$

After estimating γ_0 and all of the γ_i's, we estimate the rational and irrational sentiment as follows:

$$Rational_t = \hat{\gamma}_0 + \sum_{i=1}^{I} \hat{\gamma}_i \; Fund_{i,t} \; , \tag{6}$$

and $Irrational_t = \xi_t$. Intuitively, the $Rational_t$ is the investor's sentiment explained by the fundamental variables at time t while the $Irrational_t$ is the part of the sentiment survey that is not explained by the fundamental variables. This paper assumes that the $Irrational_t$ is the proxy for the noise trader's behavior, and the set of $Fund_i$ variables contain justifiable information (e.g. macroeconomic state) for an appropriate outlook of the future market.

4 Data

4.1 Returns and Mispricing

The daily returns on a portfolio of the bottom 20% and a portfolio of top 20% market cap stocks were from the Kenneth R. French's website. The daily returns on the two portfolios were accumulated to weekly returns to match the weekly Fama French 3 factors. For more information on the weekly returns calculations, see appendix. Furthermore, the risk-free weekly rate, the excess return on the market, SMB, and HML are from French's website. Due to AAII conducting the survey on-line in 2000, only the sentiment survey after 2000 was used in our analysis. Thus, only the Fama-French returns from 2000 to Oct-2019 were constructed.

For brevity, throughout this paper, the "low cap" portfolio refers to the portfolio with stocks in the bottom 20% of the market cap, while the "high cap" portfolio refers to the portfolio with shares in the top 20% of the market cap.

Table 1: Returns and Fama-French Statistics

Data from Jan-07-2000 to Oct-25-2019. Note that $MispricedL20_t$ and $MispricedH20_t$ variables in year 2000 to 2005 were constructed from the Fama-French factors and returns data in years 1995 to 2000. $RL20_t$ and $RH20_t$ are excess returns for low 20 percentile market cap portfolio and high 20 percentile market cap portfolio.

Statistic	N	Mean	St. Dev.	Min	Median	Max
$\mathrm{Mkt_t}$	1,034	0.109	2.472	−18.000	0.235	12.610
$\mathrm{SMB_t}$	1,034	0.045	1.367	−10.710	0.070	6.990
$\mathrm{HML_t}$	1,034	0.057	1.490	−8.710	−0.005	9.940
$\mathrm{RL20_t}$	1,034	0.161	3.026	−21.538	0.324	15.078
$\mathrm{RH20_t}$	1,034	0.097	2.376	−18.266	0.217	11.502
$\mathrm{FFL20_t}$	1,034	0.168	2.920	−19.305	0.338	12.870
$\mathrm{FFH20_t}$	1,034	0.095	2.378	−18.316	0.192	12.178
$\mathrm{MispricedL20_t}$	1,034	−0.007	0.660	−2.829	−0.006	4.484
$\mathrm{MispricedH20_t}$	1,034	0.002	0.137	−0.910	−0.001	0.905

Table 1 summarizes the descriptive statistics of the Fama French factors, actual weekly returns, predicted returns, and the mispriced returns. The three factors (Mkt, SMB,

HML) and the weekly returns are from the Fama-French website. The estimated returns were constructed every year from the July of year t to July of year t + 1 by using the actual returns and Fama French factors from year $(t-6)$ to $(t-1)$ to estimate the coefficients and using those coefficients to estimated the returns from the July of year t to July of year $t+1$.[5] Since the Fama-French coefficients were estimated every year t from the end of June to the end of June at year $t+1$, the Fama-French beta coefficients were estimated 21 times per each low and high cap portfolios to construct the $Mispriced_t$ variables. The $Mispriced_t$ variables are the difference between the actual and Fama-French predicted returns.

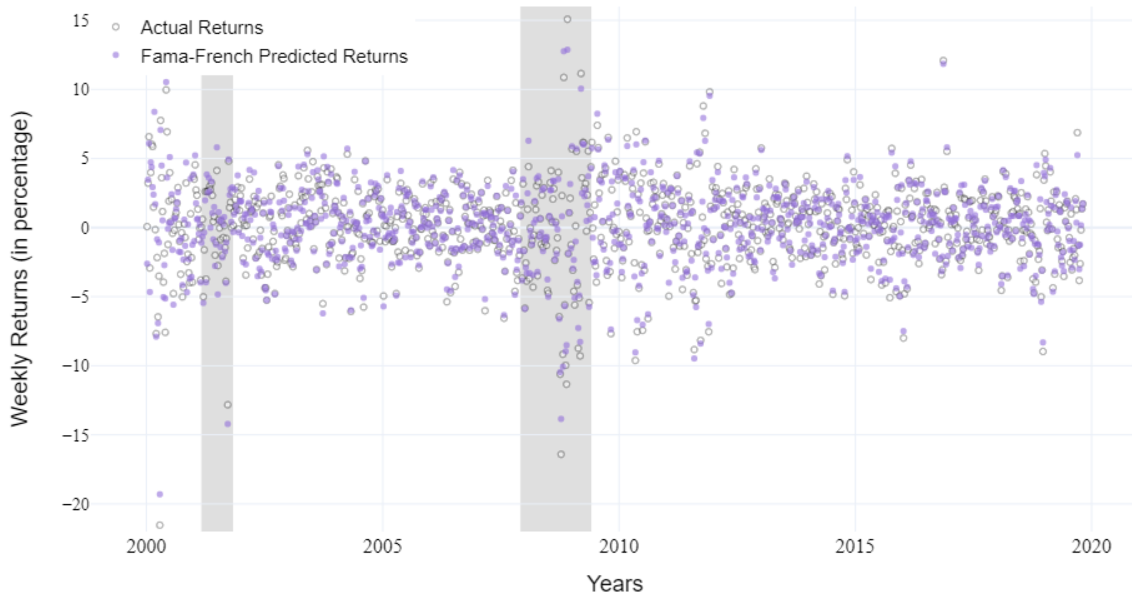
The mean of $RL20_t$ and $RH20_t$ are weekly returns of 0.161 percent and 0.097 percent, respectively. In contrast, the mean of predicted returns ($FFL20_t$ and $FFH20_t$) are 0.168 percent and 0.095 percent, respectively. The Fama-French returns for low cap seem to have slightly more bullish estimation than the high cap. Nevertheless, the standard deviation for both the estimated and the actual returns are around 2.3 to 3.1, indicating that the difference of the mean between low and high cap stock portfolios are not statistically different. One interesting note is the $Mispriced_t$ variables. The mean of $MispricedL20_t$ is around -0.007 percent, while the mean of $MispricedH20_t$ is 0.002 percent. Furthermore, the standard deviation is 0.660 for $MispricedL20_t$ while 0.137 for the $MispricedH20_t$. While the mispricing for both portfolios is essentially around zero, the volatility of $MispricedL20_t$ is almost six times higher than $MispricedH20_t$, indicating that the low cap stocks more often deviates from the Fama-French predicted model compared to the high cap stocks.

Additional illustrations of the small and large-cap returns and the fundamental returns are in figures 1 through 4. Note that for the figure 1, one can see that the colored and the outlined circles tend to mismatch each other often, illustrating that the low cap stocks tend to deviate from the theoretical price more often than the high-cap stocks in figure 2. This illustration is more evident in figure 4, in which after taking the difference between the actual and the Fama-French predicted returns, one can see that the low cap

[5] For more information, visit French's website at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`.

portfolio exhibits much higher deviation from the Fama-French predicted returns.

Figure 1: Small-Cap stocks Returns: Actual vs. Predicted



Shaded regions indicate recession, according to FRED on US Business Cycle. Plotly library from python was used to construct the graphs.

Figure 2: Large-Cap stocks Returns: Actual vs. Predicted

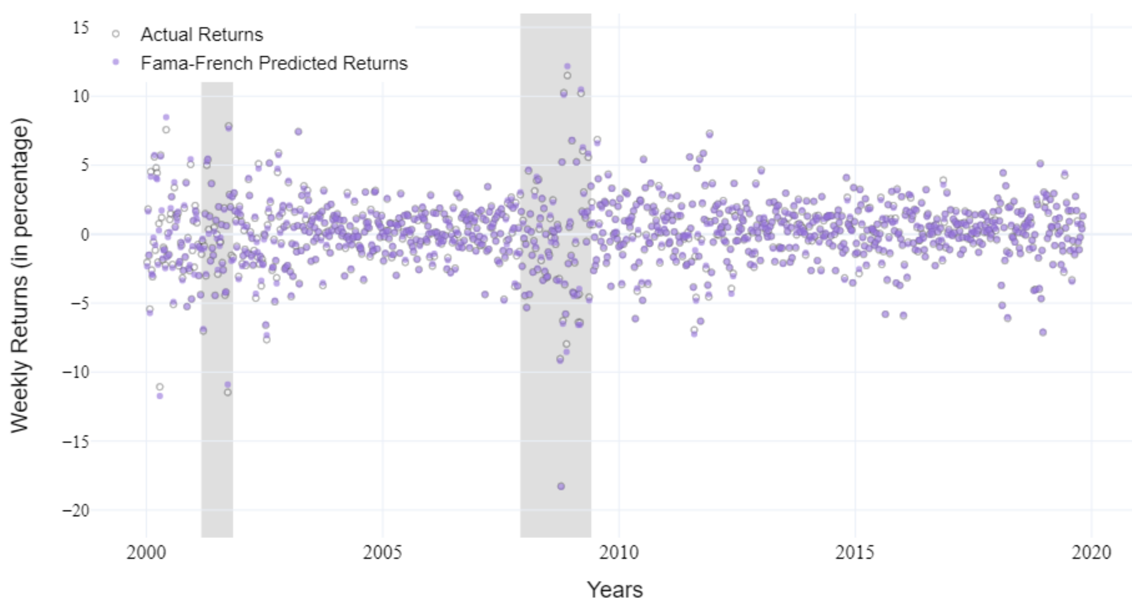
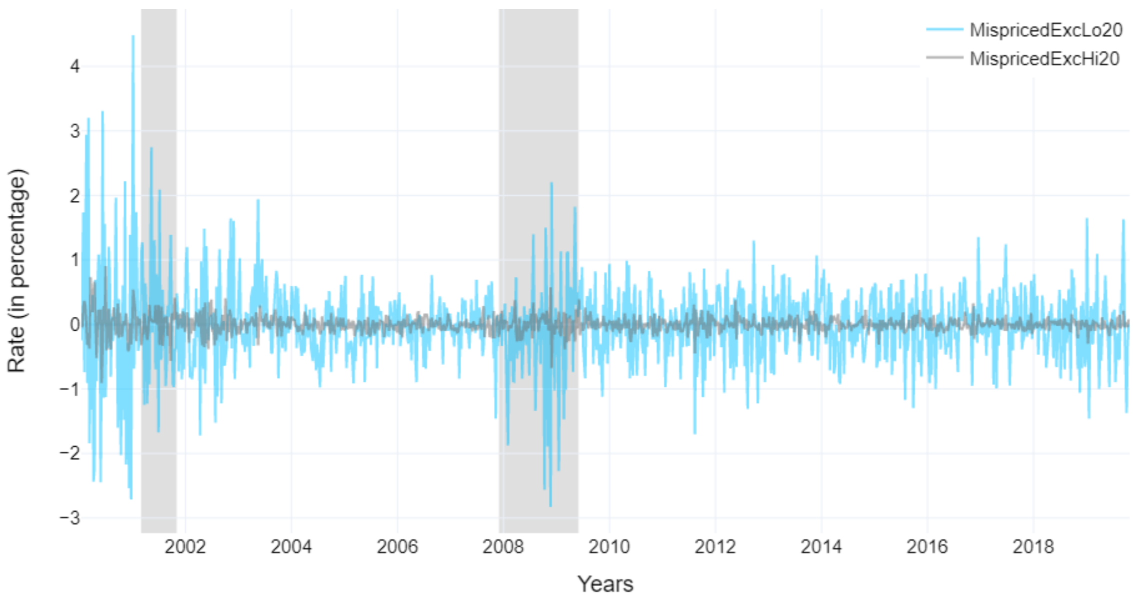


Figure 3: Mispricing Returns: from Jan 07, 2000, to October 25, 2019



Shaded regions indicate recession, according to FRED on US Business Cycle. The Mispricing of low and high cap indicates the difference between the actual weekly returns and Fama-French 3-factor model predicted returns. The rate is in weekly returns.

4.2 Rational and Irrational Sentiments

The AAII survey data is directly from American Association of Individual Investors website. The "Bull-Bear" spread in the data was used as the individual investor sentiment variable. For the macroeconomic factors, a list of commonly used and accepted variables in asset pricing literature from Verma and Verma (2005), Verma et al. (2008), and Sayim et al. (2013) were used as the fundamental factors. Specifically, they are (i) SMB, HML, Mkt—which are the Fama-French factors, (ii) $BAaa$ as the business condition, calculated by subtracting the difference in yields on Baa and Aaa corporate bonds, (iii) $T10Y3M$ as the future economic expectations variable, calculated by the difference between 10-year US treasury bond and 3-month US treasury bond, (iv) *Dividend* as the S & P 500 dividend yield, (v) MoM as the momentum factor, measured by the difference between the average return on two high prior and the average return on the low prior return portfolios, (vi) $Inflation$ as the monthly change in the consumer price index, (vii) IPI as the economic growth, calculated by the monthly change of the industrial production index, (viii) $T30$ as the short term interest rate, measured as the 1 month Treasury yield, (ix) $T90\_T30$ as economic risk premium, measured as the difference in yields on 3-month and 1-month Treasury bills.

The data on SMB, HML, Mkt, MoM, $T30$ are from the Fama-French website. The monthly dividend yields is from Robert Shiller's webpage.The $BAaa$, $T10Y3M$, $Inflation$, IPI, and $T90$ are from the FRED website. Note that since $T30$ is from the Fama-French website, $T90\_T30$ variable was constructed by taking the difference of the $T90$ and $T30$ values, after formatting for $T90$ to have consistent compounding interest yields as $T30$. See appendix for detailed construction of $T90\_T30$.[6]

Dividend is not specifically the dividend yield of the low-cap and high-cap stock portfolios. Instead, *Dividend* will be used as a market wide effect. This assumption also has been used by Sayim et al. (2013), Verma et al. (2008), Qadan and Aharon (2019), in which their dependent variable is the weighted average of different market indices but still utilized

[6] For more information about constructing $T30$ and $T90\_T30$, see appendix (9.3).

the Dividend of S &P 500 as their dividend effect.

The descriptive statistics for SMB, HML, and Mkt are illustrated in table 1 while the rest of the fundamental factors and the sentiment data is illustrated in table 2. For more information on variables, see appendix 9.4, table 14.

Table 2: Descriptive Statistics of Sentiment and Macroeconomic Factors

Statistic	N	Mean	St. Dev.	Min	Median	Max
AAII	1,034	0.073	0.180	−0.514	0.069	0.629
MoM	1,034	0.224	5.268	−34.390	0.345	18.360
Inflation	1,034	0.157	0.258	−1.621	0.170	1.139
IPI	1,034	0.067	0.643	−4.369	0.119	1.578
T90-T30	1,034	0.002	0.021	−0.110	0.000	0.080
BAaa	1,034	1.046	0.430	0.520	0.930	3.470
cBAaa	1,034	0.000	0.045	−0.230	0.000	0.490
T10Y3M	1,034	1.776	1.167	−0.766	1.923	3.782
cT10Y3M	1,034	−0.001	0.119	−0.870	−0.008	0.876
Dividend	1,034	1.887	0.374	1.110	1.900	3.600
cDividend	1,034	0.001	0.040	−0.450	0.000	0.590
chDividend	1,034	0.002	0.083	−0.450	−0.010	0.590
T30	1,034	0.134	0.152	0.000	0.090	0.560
cT30	1,034	−0.000	0.012	−0.160	0.000	0.080
chT30	1,034	−0.001	0.025	−0.160	0.000	0.080

Unfortunately, unlike Verma et al. (2008), $BAaa$, $T10Y3M$, $Dividend$, and $T30$ variables exhibited non-stationary characteristics. When the Augmented Dickey-Fuller Test was used to test for stationary of the fundamental variables, the four variables did not pass the Augmented Dickey-Fuller Tests for stationarity.[7] Therefore, the change of the variable was used instead to make the variables stationary. From examining the four non-stationary variables, it seemed like during the year 2008, $BAaa$, $T10Y3M$, $Dividend$, and $T30$ exhibited a sharp trend (either increasing–$Dividend$, $BAaa$, $T10Y3M$, or decreasing–$T30$). Since Verma et al. (2008) used data from 1988 to 2004, they may have avoided the issue of non-stationary by having the a longer sample horizon and the data before 2008.

[7]R package tseries and the function adf.test was used to test for the stationary. The lag order of 10 was used

However, Sayim et al. (2013) used the data from 1999 to 2010. With such a short sample horizon and the recession during 2008 and 2009, I expected Sayim et al. (2013) to have experienced non-stationary variables. However, their paper indicates that their variables all pass the Augmented Dickey-Fuller (ADF) Test but did not include the results of the unit root tests.

The table 3 includes both the original and the change of variables for $BAaa$, $T10Y3M$, $Dividend$, and $T30$. Note that the new variables have "Δ" prefix attached. Since $Dividend$ and $T30$ are monthly data, $\Delta_w Dividend$ and $\Delta_w T30$ represents the weekly changes of the monthly data propagated to weekly. As a result, only the first week of the month in $\Delta_w Dividend$ and $\Delta_w T30$ may have value other than zero. Thus, the monthly change of dividends and the short term interest rate was also calculated before propagating to weekly. $\Delta_M Dividend$ and $\Delta_M T30$ represents the monthly change of dividend and the 1-month Treasury yield propagated to weekly.

Table 3: Cross-correlations of US market fundamentals and Sentiment

	AAII	SMB	HML	Mkt	MoM	Inflation	IPI	T90_T30	BAaa	Δ_wBAaa	T10Y3M	Δ_wT10Y3M	Dividend	Δ_wDividend	Δ_MDividend	T30	Δ_wT30	Δ_MT30
AAII	1																	
SMB	0.11	1																
HML	0.09	-0.22	1															
Mkt	0.11	0.23	0.07	1														
MoM	-0.06	0.03	-0.08	-0.14	1													
Inflation	0.01	0.02	0.01	0.03	0.05	1												
IPI	0.11	-0.03	-0.01	0.01	0.09	0.05	1											
T90_T30	0.1	0.08	-0.02	0.04	0.19	0.01	0.23	1										
BAaa	-0.29	0.01	-0.04	-0.01	-0.19	-0.28	-0.4	-0.1	1									
Δ_wBAaa	-0.13	-0.06	0.03	-0.09	0.1	-0.24	0.03	-0.09	0.05	1								
T10Y3M	0.06	0.06	-0.04	0.03	-0.04	-0.12	0.03	0.06	0.28	-0.03	1							
Δ_wT10Y3M	0.02	0.07	0.11	0.21	-0.12	0.1	-0.08	-0.2	0.03	-0.07	0.06	1						
Dividend	-0.36	-0.01	-0.08	0.07	-0.27	-0.18	-0.25	0.04	0.7	-0.11	0.28	0.01	1					
Δ_wDividend	-0.1	-0.07	-0.01	-0.06	-0.02	-0.18	-0.09	-0.04	0.1	0.11	-0.02	-0.02	0.04	1				
Δ_MDividend	-0.15	-0.05	-0.03	-0.01	-0.03	-0.36	-0.18	-0.08	0.22	0.26	-0.05	-0.05	0.05	0.48	1			
T30	0.12	-0.02	0.08	-0.06	0.06	0.15	-0.03	-0.21	-0.26	0.06	-0.74	0.02	-0.59	0.03	0.05	1		
Δ_wT30	0.03	-0.04	0.05	-0.01	-0.06	0.06	0.02	-0.12	-0.06	0.01	-0.02	-0.01	0	-0.15	-0.07	0.02	1	
Δ_MT30	0.05	-0.1	0.05	0.01	-0.14	0.12	0.03	-0.25	-0.14	-0.07	-0.06	-0.08	-0.01	-0.07	-0.15	0.03	0.48	1

Table 4 reports the regression using the fundamental variables with or without adjusting for stationary characteristics. Since the AAII survey data after 2000 was conducted through online, I've assumed no or insignificant lag between the survey response and its publication. Therefore, the macroeconomic variables and the AAII data were regressed without any lags.[8] The column (1), "Original", represents the asset pricing variables without station-

[8]Regressing the asset pricing variables without any lags on the AAII survey variable is consistent with

arity check. The column (2), "All", represents using stationary check for the monthly variables *Dividend* and $T30$ by taking the weekly difference. The column (3), "Weeks.Months", represents the regression with monthly changes of *Dividend* and $T30$ propagated to weekly intervals. As expected, the "Original" regression provided higher R^2. Verma et al. (2008) had R^2 of 0.3 while Sayim et al. (2013) had R^2 of 0.25. Therefore, my R^2 of 0.24 seems within a credible range. However, adjusting for the stationary, we see that the R^2 drops to 0.07 and 0.08 when taken into account both the weekly and monthly change as illustrated in columns (2) and (3). As expected, using the difference of the monthly change instead of weekly change on monthly variables *Dividend* and $T30$ seem to explain more variations but by only 0.01 R^2 increase. While the drop is slightly disheartening, we can interpret that the lower R^2 may mean that after taking care of the unit root of our variables, there are wider ranges of residual (*Irrational*) values unexplained by fundamental factors. Therefore, lower R^2 does not mean that our *Rational* variable is erroneous. After taking into account the stationarity, the "Weeks.Months" regression in column (3) will be used to construct our *Rational* and *Irrational* variables.

The table 5 to 7 are the summary statistics of both *Rational* and *Irrational* variables. From the summary statistics, the mean, median, and the volatility of the *Rational* variable seems to be almost the same throughout the period when *Irrational* is bullish or bearish. The figure 4 is a time series graph illustrating the relationship between *Rational* and *Irrational* variables. Note that during the recession period, the *Rational* Bull-Bear spreads is around or under 0, illustrating that the *Rational* sentiment seems to react to the market factors appropriately. On the other hand, we see much volatile movement in our *Irrational* variable through out our sample size time interval.

Verma et al. (2008), Verma and Verma (2005), and Sayim et al. (2013).

Table 4: Effects of US market fundamentals on the individual investor's sentiment

All three regressions have $AAII$—the weekly individual investor sentiment survey variable—as the dependend variable. The column (1) is the regression of the asset pricing variables without the stationarity correction. The column (2) is the regression with the stationarity correction and the weekly change of the monthly variables $Dividend$ and $T30$, and the weekly change of the weekly variables $BAaa$ and $T10Y3M$. The column (3) is the regression with the stationarity correction and the monthly change of $Dividend$ and $T30$.

	AAII		
	(1)	(2)	(3)
	Original	All	Weeks.Months
SMB	0.010***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)
HML	0.007*	0.013***	0.013***
	(0.003)	(0.004)	(0.004)
Mkt	0.006***	0.004*	0.005*
	(0.002)	(0.002)	(0.002)
MoM	−0.006***	−0.002	−0.002
	(0.001)	(0.001)	(0.001)
Inflation	−0.055***	−0.030	−0.051**
	(0.020)	(0.022)	(0.023)
IPI	−0.014	0.030***	0.025***
	(0.009)	(0.009)	(0.009)
T90-T30	1.248***	0.529*	0.599**
	(0.262)	(0.287)	(0.296)
BAaa	−0.066***		
	(0.021)		
Δ_wBAaa		−0.474***	−0.390***
		(0.127)	(0.128)
T10Y3M	0.041***		
	(0.008)		
Δ_wT10Y3M		0.001	0.005
		(0.049)	(0.049)
Dividend	−0.159***		
	(0.029)		
Δ_wDividend		−0.301**	
		(0.142)	
Δ_MDividend			−0.257***
			(0.073)
T30	0.154**		
	(0.074)		
Δ_wT30		0.326	
		(0.455)	
Δ_MT30			0.331
			(0.232)
Constant	0.354***	0.073***	0.077***
	(0.057)	(0.006)	(0.007)
Observations	1,034	1,034	1,034
R^2	0.236	0.071	0.080
Adjusted R^2	0.228	0.061	0.070

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 5: Full Sample

Statistic	N	Mean	St. Dev.	Min	Median	Max
Rational	1,033	0.073	0.051	−0.295	0.075	0.330
Irrational	1,033	−0.000	0.172	−0.455	−0.010	0.544

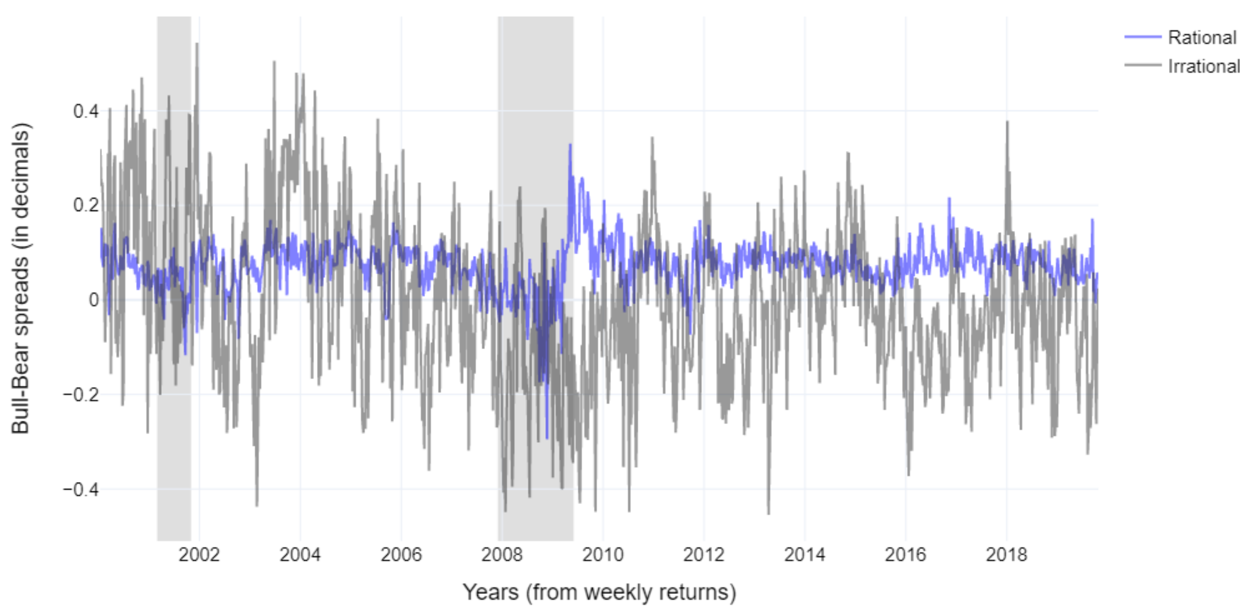
Table 6: Bullish Sample: Irrational > 0

Statistic	N	Mean	St. Dev.	Min	Median	Max
Rational	495	0.072	0.045	−0.174	0.077	0.170
Irrational	495	0.142	0.111	0.000	0.115	0.544

Table 7: Bearish Sample: Irrational < 0

Statistic	N	Mean	St. Dev.	Min	Median	Max
Rational	538	0.073	0.056	−0.295	0.073	0.330
Irrational	538	−0.132	0.098	−0.455	−0.114	−0.000

Figure 4: Sentiment: from Jan 07, 2000, to October 25, 2019



5 Results

Note that throughout this section, there will be only 1033 observations per regression. This is due to the lag of *Rational* and *Irrational* variables.

The bivariate correlations of our interested variables are illustrated in table 8. Note that there seems to be a slightly negative correlation between the current and the previous week's mispricing as indicated by the correlations of $MisrpciedL20_t$ & $MispricedL20_{t-1}$ and $MispricedH20_t$ & $MispricedH20_{t-1}$. This negative correlation between the past and current week's mispricing seem to suggest a price correction from the market for both portfolios. Other high correlations are $RL20_{t-1}$ & $RH20_{t-1}$ and $Rational_t$ & $Rational_{t-1}$, which are 0.809 and 0.889 respectively. The correlation of $RL20_{t-1}$ & $RH20_{t-1}$ is realistic since as a overall market, if the economy is doing well (poorly), then the returns of both the high and low market may go up (down). The correlation $Rational_t$ & $Rational_{t-1}$ illustrates that the positive or negative sentiments explained by the fundamentals in one week may be related to the following week.

Table 8: Cross-correlations of Fama-French excess returns, Market excess returns, and the Sentiments

	$MispricedL20_t$	$MispricedH20_t$	$MispricedL20_{t-1}$	$MispricedH20_{t-1}$	$RL20_{t-1}$	$RH20_{t-1}$	$Rational_t$	$Irrational_t$	$Rational_{t-1}$	$Irrational_{t-1}$
$MispricedL20_t$	1									
$MispricedH20_t$	-0.045	1								
$MispricedL20_{t-1}$	-0.054	0.041	1							
$MispricedH20_{t-1}$	0.045	-0.11	-0.036	1						
$RL20_{t-1}$	0.259	-0.024	0.266	0.02	1					
$RH20_{t-1}$	0.218	-0.009	0.019	0.017	0.809	1				
$Rational_t$	0.075	0.081	0.054	0.068	0.07	0.021	1			
$Irrational_t$	0.166	-0.024	0.059	0.031	0.386	0.373	-0.005	1		
$Rational_{t-1}$	0.096	0.061	0.083	0.085	0.318	0.214	0.889	0.116	1	
$Irrational_{t-1}$	0.025	0.015	0.173	-0.02	0.04	-0.001	0.01	0.522	0	1

5.1 Regressing without Interactions

Before examining how the noise traders may react to the past information and the mispricing, this section covers how the rational and irrational (noise traders) sentiment individually impact the mispricing of low and high cap portfolios. The main equation in this section is as follows:

$$\begin{aligned} Mispriced\_20_t = & \lambda_1 + \sum_{i=0}^{1} \lambda_{2,i}\ Rational_{t-i} \\ & + \sum_{i=0}^{1} \lambda_{3,i} Irrational_{t-i} + \lambda_4 Mispriced\_20_{t-1} + \lambda_5 X, \end{aligned} \tag{7}$$

where $Mispriced\_20_{t-1}$ is either $MispricedL20_t$ or $MispricedH20_t$ and X is either $RL20_{t-1}$, $FFL20_{t_1}$, $MispricedH20_{t-1}$ or $RH20_{t-1}$, $FFH20_{t_1}$,$MispricedL20_{t-1}$.

The OLS regression is on table 9. The column (1) and (5) are the impact of $Rational_t$ and $Irrational_t$ variables on low and high cap portfolio mispricing. The column (2) and (6) test the impact of low (high) cap portfolio's previous market excess returns on the current mispricing of high (low) cap portfolio. Thus, $RH20_{t-1}$ and $RL20_{t-1}$ were added to exmine the impact of previous week's returns from the opposite market-cap side of the market to the current mispricing. The column (3) and (7) test the impact of last week's Fama-French predicted low (high) cap portfolio returns on the current high (low) cap mispricing. The column (4) and (8) test the impact of other portfolio's mispricing on the current respective portfolios (e.g. $MispricedH20_{t-1}$ on $MispricedL20_t$).

The significant coefficients of $Rational_t$ and $Irrational_t$ on $MispricedL20_t$ (columns (1) through (4)) illustrate that the optimistic and pessimistic outlook for both $Rational_t$ and $Irrational_t$ uniquely influence the mispricing of low cap stocks while do not seem to have an influence on high cap stocks. So far, this is consistent with Mitchell et al. (2002) and Podolski et al. (2009). Investor's sentiment—whether based on fundamental factors or a noise—seem to have a positive and statistically significant influence on the mispricing of the low cap portfolio while not on the mispricing of the high cap portfolio.

Another interesting coefficients are $MispricedL20_{t-1}$ in columns (1) through (4) and

$MispricedH20_{t-1}$ in columns (5) through (8). While this paper focuses on the market anomalies, the negative and significant coefficients suggest the mean-reversal; the positive (negative) mispricing of time t negatively (positively) influence the next week's mispricing at time $t+1$. Nevertheless, comparing the coefficients of previous mispricing effect, even though the descriptive statistics in table 1 indicate that the value range and volatility is higher for $MispricedL20_t$ compared to $MispricedH20_t$, the coefficient of $MispricedH20_{t-1}$ is around -0.11, almost twice more negative than the coefficient of $MispricedL20_{t-1}$. Although further tests and scrutiny are needed, the coefficients may imply that the mispricing of high cap stocks tend to revert to their average more rapidly than the mispricing of small cap stocks.

In columns (2), (3), (6), and (7), the coefficients of $RH20_{t-1}$ and FFL_{t-1} compared to $RL20_{t-1}$ and FFH_{t-1} may either indicate that the past actual returns and Fama French predicted returns may have information that impacts the mispricing of the low cap market. However, the insignificant $MispricedH20_{t-1}$ in column (4) may indicate that the small cap market does not take into consideration how much the high cap market over or under performed compared to the Fama-French model. However, as indicated by the significant coefficients from $RH20_{t-1}$ and $FFH20_{t-1}$ in columns (5) and (8), the high cap portfolio information influences the mispricing of the low cap portfolio. One explanation is that the high cap market returns may include a systematic market wide risk or the additional "hype" that impact the mispricing of the low cap market. Since small firms are generally less covered by analysts compared to the large cap stocks, the large cap stocks coverage—and their returns—may influence the perspective of the market as whole, including the mispricing of the small cap stocks.

Table 9: OLS: Mispriced without interactions

Columns (1) to (4) are regressions of sentiment and lagged variables on $MispriciedL20_t$. Columns (5) to (8) are regressions on $MispricedH20_t$. All of the regressions have high F statistics to be signficicant at the 1 percent level.

	$\text{MispricedL20}_\text{t}$				$\text{MispricedH20}_\text{t}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rational_t	1.403***	1.519***	1.519***	1.394***	−0.020	−0.041	−0.048	−0.020
	(0.486)	(0.481)	(0.482)	(0.486)	(0.104)	(0.106)	(0.106)	(0.104)
$\text{Irrational}_\text{t}$	0.750***	0.493***	0.501***	0.741***	−0.004	0.006	0.011	−0.002
	(0.153)	(0.160)	(0.160)	(0.153)	(0.033)	(0.034)	(0.035)	(0.033)
$\text{Rational}_\text{t-1}$	0.642	−0.037	−0.026	0.651	0.014	0.077	0.096	0.002
	(0.501)	(0.515)	(0.516)	(0.501)	(0.106)	(0.124)	(0.124)	(0.107)
$\text{Irrational}_\text{t-1}$	−0.314**	−0.161	−0.165	−0.312**	0.040	0.036	0.032	0.035
	(0.151)	(0.153)	(0.153)	(0.151)	(0.032)	(0.032)	(0.032)	(0.032)
$\text{MispricedL20}_\text{t-1}$	−0.067**	−0.067**	−0.068**	−0.066**				0.006
	(0.031)	(0.031)	(0.031)	(0.031)				(0.007)
$\text{MispricedH20}_\text{t-1}$				0.153	−0.110***	−0.110***	−0.110***	−0.109***
				(0.145)	(0.031)	(0.031)	(0.031)	(0.031)
$\text{RH20}_\text{t-1}$		0.046***						
		(0.009)						
$\text{RL20}_\text{t-1}$						−0.002		
						(0.002)		
$\text{FFH20}_\text{t-1}$			0.045***					
			(0.009)					
$\text{FFL20}_\text{t-1}$							−0.002	
							(0.002)	
Constant	−0.159***	−0.122***	−0.123***	−0.159***	0.002	−0.000	−0.001	0.003
	(0.038)	(0.039)	(0.039)	(0.038)	(0.008)	(0.009)	(0.009)	(0.008)
Observations	1,033	1,033	1,033	1,033	1,033	1,033	1,033	1,033
R^2	0.052	0.073	0.072	0.053	0.014	0.015	0.016	0.015
Adjusted R^2	0.047	0.067	0.067	0.047	0.009	0.009	0.010	0.009

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

5.2 Regressing with Interactions

Since the impact of $Rational_t$ and $Irrational_t$ sentiments may be influenced by the previous week's mispricing, the sentiment and the previous mispricing interaction variables were added. The regression results with interaction variables are illustrated in table 10.[9]

Similar to the result from table 9, in columns (1) to (4), we see that $Rational_t$ and $Irrational_t$ are positive and significant. However, the inclusion of the interaction terms have slightly lowered the coefficients for $Rational_t$.

However, there are several differences. The coefficients of $MispricedL20_{t-1}$ in columns (1) to (4) do not have statistically significant association to $MispricedL20_t$ while the coefficients of $MispricedH20_{t-1}$ in columns (5) to (8) are still significant. Another interesting coefficients are $Irrational_t : MispricedL20_{t-1}$ and $Irrational_t : MispricedH20_{t-1}$.

In column (4), $Irrational_t : MispricedH20_{t-1}$ is statistically significant while $Irrational_t : MispricedL20_{t-1}$ is not. This implies that the information of high cap portfolio's mispricing—not the low cap portfolio's mispricing—has a significant interaction effect with $Irrational_t$ on $MispricedL20_t$. This seems conflicting, since the information of the low-cap stocks does not seem to have significant interaction with any of the sentiments on the next week's mispricing of the low-cap portfolio.[10] Furthermore, $Irrational_t : MispricedH20_{t-1}$ is -3.241, one of the most significant and negative coefficients in table 10.

On the other hand, for columns (5) to (8), $Irrational_t : MispricedH20_{t-1}$ is around 0.45. This positive interaction may indicate that when $MispricedH20_{t-1}$ is positive (negative), the optimistic (pessimistic) noise trader may buy (sell) more of the high cap stocks, influencing the mispricing of the high cap market depending on irrational's mood. The $Irrational_t$'s negative significant interaction with $MispricedH20_{t-1}$ instead of $MispricedL20_{t-1}$ in column (4) may illustrate that since large-cap stocks are more visible to the noise traders, those without enough knowledge of financial markets may take a glance of positive (negative) performance of large-cap firms and invest (sell) small-cap stocks. Furthermore, the noise traders using large-cap stock's information

[9] The interaction variable $Rational_t : Irrational_t$ was not added because its interaction was not significant in any of the regressions in table 10

[10] To elaborate, the $MispricedL20_{t-1}$ variable and its interactions like $Irrational_t : MispricedL20_{t-1}$are not significant on the dependent variable $MispricedL20_t$.

on the investment decision of small-cap stocks may lead to erroneous and negative returns on the small-cap stocks, on average. The insignificant interaction of $Irrational_t$ and $MispricedL20_{t-1}$ may mean the noise traders may not consider the past performance of low cap stocks as much as the high cap stocks for both $MispricedL20_t$ and $MispricedH20_t$.

Table 10: OLS: Mispriced with all variables interactions

Columns (1) to (4) are regressions of sentiment, lagged, and interaction variables on $MispriciedL20_t$. Columns (5) to (8) are regressions on $MispricedH20_t$. All of the regressions have high F statistics to be signficicant at the 1 percent level.

	MispricedL20_t				MispricedH20_t			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rational_t	1.345***	1.493***	1.493***	1.225**	−0.008	−0.022	−0.029	−0.001
	(0.490)	(0.486)	(0.486)	(0.486)	(0.104)	(0.106)	(0.106)	(0.104)
Irrational_t	0.737***	0.489***	0.496***	0.750***	−0.009	−0.001	0.003	−0.009
	(0.153)	(0.161)	(0.161)	(0.152)	(0.033)	(0.034)	(0.035)	(0.033)
Rational_{t-1}	0.650	−0.030	−0.019	0.783	−0.005	0.037	0.057	−0.028
	(0.503)	(0.519)	(0.519)	(0.500)	(0.106)	(0.124)	(0.124)	(0.107)
Irrational_{t-1}	−0.302**	−0.157	−0.161	−0.271*	0.038	0.035	0.031	0.033
	(0.151)	(0.153)	(0.153)	(0.150)	(0.032)	(0.032)	(0.032)	(0.032)
$\text{MispricedL20}_{t-1}$	−0.030	−0.049	−0.049	−0.016				0.006
	(0.044)	(0.044)	(0.044)	(0.044)				(0.009)
$\text{MispricedH20}_{t-1}$				0.233	−0.147***	−0.148***	−0.147***	−0.142***
				(0.224)	(0.048)	(0.048)	(0.048)	(0.048)
$\text{Rational}_t\text{:MispricedL20}_{t-1}$	−0.563	−0.260	−0.261	−0.766				0.031
	(0.514)	(0.513)	(0.513)	(0.514)				(0.110)
$\text{Irrational}_t\text{:MispricedL20}_{t-1}$	−0.144	−0.113	−0.113	−0.205				−0.046
	(0.170)	(0.169)	(0.169)	(0.170)				(0.036)
$\text{Rational}_t\text{:MispricedH20}_{t-1}$				0.948	0.273	0.289	0.291	0.215
				(2.665)	(0.567)	(0.568)	(0.568)	(0.572)
$\text{Irrational}_t\text{:MispricedH20}_{t-1}$				−3.241***	0.465***	0.453***	0.447***	0.457***
				(0.745)	(0.159)	(0.160)	(0.160)	(0.160)
RH20_{t-1}		0.045***						
		(0.010)						
RL20_{t-1}						−0.001		
						(0.002)		
FFH20_{t-1}			0.044***					
			(0.010)					
FFL20_{t-1}							−0.002	
							(0.002)	
Constant	−0.152***	−0.119***	−0.120***	−0.148***	0.002	0.000	−0.000	0.004
	(0.038)	(0.039)	(0.039)	(0.038)	(0.008)	(0.009)	(0.009)	(0.008)
Observations	1,033	1,033	1,033	1,033	1,033	1,033	1,033	1,033
R^2	0.053	0.073	0.073	0.072	0.024	0.024	0.024	0.026
Adjusted R^2	0.047	0.066	0.065	0.063	0.017	0.016	0.017	0.016

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

5.3 Serial Correlation Test

Due to the nature of our time-series data, I've tested the above regressions for serial correlation by regressing the regression's residual with 11 lags—around two and a half months of lag duration. Therefore, for each regression in the table 10, its residual e_t is regressed as follows:

$$e_t = \sum_{i=1}^{11} x_i \ e_{t-i} \text{ , where} \tag{8}$$

each x_i coefficient was checked for the serial correlation.

In table 11, we see that for low cap portfolio, its residuals at lag 3 and 4 seems to have statistically significant relationship to the current week t. For the high cap portfolio in columns (5) to (8), while the lag 7 in serial correlation seem significant, due to the low F statistics, only the lag 3 and 4 in the $MispriecedL20$ were considered. Thus, the four-weeks lagged dependent variable was added to each of the regression in section 5.2 and again, checked for serial correlation. An additional serial correlation test afterwards indicate that the regressions in table 12 does not seem to exhibit a serial correlation, up to lag 11.

The table 12 illustrates the regressions after the lags. Similar to section 5.2, many of the coefficients are still signficant. The new variables in table12 are the four weeks lagged mispricing variables: $MispricedH20_{t-4}$ and $MispricedL20_{t-4}$. For the high cap portfolio, its mispricing does not seem to be statistically impacted by its mispricing four weeks ago. However, for the low cap portfolio, its mispricing a week before does not seem to have a unique effect on the current mispricing while its mispricing from 4 weeks do. Furthermore, from the taking into account that the the regressions in table 12 does not seem to exhibit a serial correlation up to lag 11, the significance of lag 4 and lag 1 in the mispricing of the low cap portfolio and high cap portfolio, respectively, may mean the mispricing of the low cap stocks may exhibit the price correction later than the high cap mispricing. This analysis is consistent with the theory that from greater limits of arbitrage in small cap stocks, the price correction may take longer than the high cap stocks (Podolski et al., 2009).

Table 11: Serial Lags Result

	MispricedL20				MispricedH20			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x_1	0.038	0.035	0.035	0.049	−0.011	−0.011	−0.010	−0.011
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
x_2	−0.017	−0.017	−0.016	−0.031	−0.003	−0.002	−0.002	−0.007
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
x_3	0.082***	0.081***	0.080***	0.087***	0.028	0.026	0.025	0.028
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
x_4	−0.130***	−0.128***	−0.128***	−0.125***	−0.024	−0.024	−0.025	−0.026
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
x_5	0.024	0.019	0.018	0.029	−0.003	−0.005	−0.006	−0.003
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_6	−0.005	−0.010	−0.010	−0.010	−0.023	−0.021	−0.021	−0.022
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_7	−0.008	−0.015	−0.016	−0.003	−0.094***	−0.094***	−0.093***	−0.092***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_8	−0.035	−0.034	−0.033	−0.020	0.034	0.033	0.032	0.031
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_9	0.010	0.025	0.025	0.018	0.004	0.005	0.005	0.001
	(0.031)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_{10}	0.030	0.011	0.013	0.035	0.041	0.040	0.039	0.043
	(0.031)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
x_{11}	−0.019	−0.023	−0.023	−0.013	−0.042	−0.041	−0.041	−0.045
	(0.031)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
Observations	1,022	1,022	1,022	1,022	1,022	1,022	1,022	1,022
F Statistic	2.537***	2.489***	2.473***	2.506***	1.497	1.448	1.427	1.486

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 12: OLS: lag 4 added.

Columns (1) to (4) are regressions of sentiment, lagged, and interaction variables on $MispriciedL20_t$. Columns (5) to (8) are regressions on $MispricedH20_t$. All of the regressions have high F statistics to be signficicant at the 1 percent level.

	MispricedL20$_t$				MispricedH20$_t$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rational$_t$	1.368***	1.507***	1.507***	1.247***	−0.006	−0.020	−0.028	0.002
	(0.486)	(0.483)	(0.483)	(0.483)	(0.104)	(0.106)	(0.106)	(0.104)
Irrational$_t$	0.738***	0.503***	0.510***	0.752***	−0.008	−0.000	0.004	−0.008
	(0.152)	(0.160)	(0.160)	(0.151)	(0.033)	(0.034)	(0.035)	(0.033)
Rational$_{t-1}$	0.739	0.088	0.099	0.873*	−0.005	0.039	0.059	−0.027
	(0.500)	(0.516)	(0.516)	(0.496)	(0.106)	(0.124)	(0.124)	(0.107)
Irrational$_{t-1}$	−0.278*	−0.142	−0.146	−0.245	0.039	0.035	0.032	0.034
	(0.150)	(0.152)	(0.152)	(0.149)	(0.032)	(0.032)	(0.032)	(0.032)
MispricedL20$_{t-1}$	−0.024	−0.042	−0.042	−0.011				0.006
	(0.044)	(0.044)	(0.044)	(0.044)				(0.009)
MispricedH20$_{t-1}$				0.188	−0.148***	−0.148***	−0.147***	−0.142***
				(0.223)	(0.048)	(0.048)	(0.048)	(0.048)
Rational$_t$:MispricedL20$_{t-1}$	−0.485	−0.204	−0.204	−0.676				0.033
	(0.510)	(0.510)	(0.510)	(0.511)				(0.110)
Irrational$_t$:MispricedL20$_{t-1}$	−0.074	−0.049	−0.049	−0.131				−0.045
	(0.170)	(0.168)	(0.168)	(0.170)				(0.036)
Rational$_t$:MispricedH20$_{t-1}$				1.624	0.280	0.296	0.298	0.224
				(2.649)	(0.567)	(0.568)	(0.568)	(0.572)
Irrational$_t$:MispricedH20$_{t-1}$				−3.314***	0.463***	0.450***	0.444***	0.455***
				(0.740)	(0.159)	(0.160)	(0.160)	(0.160)
MispricedL20$_{t-4}$	−0.123***	−0.114***	−0.115***	−0.126***				
	(0.030)	(0.030)	(0.030)	(0.030)				
MispricedH20$_{t-4}$					−0.034	−0.034	−0.034	−0.034
					(0.031)	(0.031)	(0.031)	(0.031)
RH20$_{t-1}$		0.043***						
		(0.010)						
RL20$_{t-1}$						−0.001		
						(0.002)		
FFH20$_{t-1}$			0.042***					
			(0.010)					
FFL20$_{t-1}$							−0.002	
							(0.002)	
Constant	−0.162***	−0.130***	−0.130***	−0.157***	0.002	0.000	−0.001	0.003
	(0.038)	(0.039)	(0.039)	(0.038)	(0.008)	(0.009)	(0.009)	(0.008)
Observations	1,033	1,033	1,033	1,033	1,033	1,033	1,033	1,033
R^2	0.068	0.086	0.085	0.087	0.025	0.025	0.026	0.027
Adjusted R^2	0.061	0.078	0.077	0.078	0.017	0.017	0.017	0.017

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

6 Discussions

Without the economic analysis, our results from the above regressions allow us to conclude the following:

1. Effect of the Rational Sentiment: $Rational_t$ is statistically significant for all of our regressions on response variable $MispricedL20_t$ while not for the regressions on $MispricedH20_t$.
2. Effect of the Noise Trader:$Irrational_t$ is statistically significant for all of our regressions on $MispricedL20_t$ on $MispricedH20_t$.
3. Effect of the Noise Trader and Large-cap Mispricing Interaction: the past mispricing of high cap portfolio $MispricedH20_{t-1}$ has negative significant interaction with $Irrational_t$ on $MispricedL20_t$ while positive interaction on $MispricedH20_{t-1}$.[11]
4. Effect of Outside Market Information: The past excess returns and Fama-French predicted returns from the *high cap* portfolio ($RH20_{t-1}$ and $FFH20_{t-1}$) are significant on $MispricedL20_t$ while the past excess returns and Fama-French predicted returns from the *low cap* portfolio ($RL20_{t-1}$ and $FFL20_{t-1}$) are not significant on $MispricedH20_t$.[12]

From tables 10 and 12, while both sentiments–$Irrational_t$ and $Rational_t$–are significant on $MispricedL20_t$, the sentiments does not seem to have significance on $MispricedH20_t$. Thus, the noise traders seem to respond differently depending on the market. Specifically, for the regressions on the mispricing of the $MispricedL20_t$–from columns (1) to (4)–while $Irrational_t$ seems to have the main effect[13], the significant and large interaction value of $Irrational_t : MispricedH20_{t-1}$ suggest that the noise traders react to the information of the large-cap market and may significantly (and erroneously) impact the mispricing of the small-cap portfolio.

One explanation for the insignificance of $Rational_t$ and $Irrational_t$ on $MispricedH20_t$ may be that the large-cap stocks are more liquid and that there are less limitation of arbitrage (Hong et al., 2000; Mitchell et al., 2002). While more arbitrageurs can offset the effects of noise traders—as illustrated by the insignificant $Rational_t$ and $Irrational_t$ coefficients for $MispricingH20_t$, the higher visibility of the high cap market[14] means more noise traders may

[11] See table 10 and 12, columns (4) through (8).
[12] See table 10 and 12.
[13] The main effect here means that $Irrational_t$ variable is statistically significant.
[14] Large-cap stocks may have higher concentration of noise traders according to Podolski et al. (2009).

concentrate in high cap stock market compared to the low cap stock market. When there are cases of large $MispricingH20_{t-1}$ value in the previous week, the noise traders may have displayed significant presence in the high cap market to override the arbitrageurs. Therefore, the following (next) week, the $Irrational_t$ may have bigger influence on high cap market since they have already concentrated in the high cap market and have a big presence in the market. Thus, the positive $Irrational_t\colon MispricedH20_{t-1}$ coefficient on $MispricedH20_t$ may imply that when the $MispricedH20_{t-1}$ is positive, the high cap stocks were on average overvalued by the significant presence of the noise traders the week before. Thus, the bullish push at time t by the $Irrational_t$ may further push the $MispricedH20_t$ to further overvaluation. On the other hand, if $Irrational_t$ is bearish, the irrational noise traders concentrating the $MispricedH20_{t-1}$ pushes the price lower and selling at one of the worst time.

When the same logic is applied to the negative $MispricedH20_{t-1}$ from columns (5) to (8), it's puzzling that $MispricedH20_{t-1}$ and $Irrational_t$ being negative results in a positive interaction. The table 13 illustrates the regressions after splitting the effect of positive and negative $MispricedH20_{t-1}$. Now, we see that in table 13, columns (5) through (8), the interaction between the $Irrational_t$ and the negative $MispricedH20_{t-1}$[15] is not significant. This implies that the bullish noise trader presence and the resulting positive $MispricedH20_{t-1}$ the week before translates to the significant returns of the next week's mispricing only when there was an overvaluation in the high cap portfolio.

The $Irrational_t\colon MispricedH20_t$ on the $MispricedL20_t$ regression is more puzzling. However, if we assume that the noise traders in the market who have concentrated themselves move the large-cap stocks price from their fundamental value, then either the noise traders have immigrated their positions in the small-cap market to the large-cap market or may have utilized the information from large-cap market to erroneously judge and impact the market.

Furthermore, looking at the coefficients $Irrational_t\colon MispricedH20_{t-1|\geqslant 0}$ and $Irrational_t\colon MispricedH20_{t-1|<0}$, one interpretation of statistical significance in the $MispricedL20_t$ may be from the volume and concentration of noise trader's impact when $MispricedH20_t$ is negative. One of the Friedman's description of the noise trader is that they

[15]The variable $Irrational_t\colon MispricedH20_{t-1}$

buy and sell at the worst time (Lee et al., 2002). Thus, if the $MispricedH20_t$ is positive from last week's noise traders, more noise traders may join in and also the noise traders from the low cap market. Thus, less concentration of the noise traders in the low cap market from the noise traders going to the high cap market may lead the current risk premium/discount set by the noise traders in the low cap market to be diminished.

We can see this by an example as follows: the bullish $Irrational_t$ term would push the $MispricedL20_t$ value higher. However, when $Irrational_t$ interact with the positive $MishpeicedH20_{t-1}$, this noise trader premium that would have been present in the low cap market is now moved to the high cap market, leading to negative impact on $MispricedL20_t$. On the other hand, if the noise trader is bearish, the additional discounting from $Irratioanl_t$ is transferred to the high cap stocks, leading to positive value of $Irrational_t\colon MispricedH20_{t-1}$ on $MispricedL20_t$.[16]

However, when the previous week's mispricing of the high cap market is negative but the irrational traders are bullish, the negative $Irrational_t\colon MispricedH20_{t-1|<0}$ implies that the bullish $Irrational_t$ will have an overall positive interaction effect to the $MispricedL20_t$. One of the explanation is that the negative $MispricedH20_{t-1}$ may imply smaller concentration of the noise traders in the high cap market and larger concentrate in the low cap market. Thus, high concentration of noise traders in low cap stocks results in the $Irrational_t$ value to have an amplified effect—when the noise traders are optimistic or pessimistic, the mispricing of the low cap portfolio shifts further to the irrational sentiment's (noise trader) direction.

Ultimately, this explanation needs to be further tested with low and high cap market volume trades to examine if the $Irrational_t$ and $Rational_t$ sentiments and the reactions from the $MispricedH20_{t-1}$ have significant and main effect to the market volume. Nevertheless, the paradox of negative $Irrational_t\colon MispricedH20_t$ in $MispricedL20_t$ may illustrate the impact of noise trader's shifting volumes and concentrations in the market. We do see that $Irrational_t$ have main effect on $MispricedL20_t$. However, $Irrational_t\colon MispricedH20_t$ shades further information on how noise trader's reaction and the concentration of visible (large-cap sized) market may further influence the smaller market—the one with greater limits of arbitrage—to deviate further from the

[16]The positive value of $Irrational_t\colon MispricedH20_{t-1}$ is from negative value $Irrational_t$ multiplied by the negative value $MispricedH20_{t-1}$.

Fama-French's fundamental price.

Table 13: OLS: separating out the positive and negative $\text{MispricedH20}_{\text{t-1}}$

Columns (1) to (4) are regressions of sentiment, lagged, and interaction variables on $MispriciedL20_t$. Columns (5) to (8) are regressions on $MispricedH20_t$. All of the regressions have high F statistics to be signficant at the 1 percent level.

	$\text{MispricedL20}_{\text{t}}$				$\text{MispricedH20}_{\text{t}}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Rational}_{\text{t}}$	1.368***	1.507***	1.507***	1.277***	−0.044	−0.051	−0.060	−0.037
	(0.486)	(0.483)	(0.483)	(0.486)	(0.104)	(0.106)	(0.106)	(0.105)
$\text{Irrational}_{\text{t}}$	0.738***	0.503***	0.510***	0.812***	−0.087**	−0.082*	−0.076*	−0.085**
	(0.152)	(0.160)	(0.160)	(0.187)	(0.040)	(0.042)	(0.043)	(0.040)
$\text{Rational}_{\text{t-1}}$	0.739	0.088	0.099	0.844*	0.033	0.056	0.079	0.009
	(0.500)	(0.516)	(0.516)	(0.499)	(0.106)	(0.124)	(0.124)	(0.107)
$\text{Irrational}_{\text{t-1}}$	−0.278*	−0.142	−0.146	−0.251*	0.048	0.046	0.043	0.042
	(0.150)	(0.152)	(0.152)	(0.149)	(0.032)	(0.032)	(0.032)	(0.032)
$\text{MispricedL20}_{\text{t-1}}$	−0.024	−0.042	−0.042	−0.012				0.007
	(0.044)	(0.044)	(0.044)	(0.044)				(0.009)
$\text{MispricedH20}_{\text{t-1}}$				0.194	−0.155***	−0.155***	−0.155***	−0.151***
				(0.223)	(0.048)	(0.048)	(0.048)	(0.048)
$\text{Rational}_{\text{t}}\text{:MispricedL20}_{\text{t-1}}$	−0.485	−0.204	−0.204	−0.677				0.034
	(0.510)	(0.510)	(0.510)	(0.511)				(0.110)
$\text{Irrational}_{\text{t}}\text{:MispricedL20}_{\text{t-1}}$	−0.074	−0.049	−0.049	−0.140				−0.035
	(0.170)	(0.168)	(0.168)	(0.170)				(0.036)
$\text{Rational}_{\text{t}}\text{:MispricedH20}_{\text{t-1}}$				1.570	0.336	0.344	0.348	0.286
				(2.652)	(0.565)	(0.565)	(0.565)	(0.570)
$\text{Irrational}_{\text{t}}\text{:MispricedH20}_{\text{t-1} \mid \geqslant 0}$				−3.770***	1.053***	1.040***	1.028***	1.048***
				(1.118)	(0.238)	(0.241)	(0.241)	(0.240)
$\text{Irrational}_{\text{t}}\text{:MispricedH20}_{\text{t-1} \mid < 0}$				−2.718**	−0.315	−0.314	−0.316	−0.319
				(1.324)	(0.283)	(0.284)	(0.284)	(0.284)
$\text{MispricedL20}_{\text{t-4}}$	−0.123***	−0.114***	−0.115***	−0.124***				
	(0.030)	(0.030)	(0.030)	(0.030)				
$\text{MispricedH20}_{\text{t-4}}$					−0.037	−0.037	−0.037	−0.037
					(0.031)	(0.031)	(0.031)	(0.031)
$\text{RH20}_{\text{t-1}}$		0.043***						
		(0.010)						
$\text{RL20}_{\text{t-1}}$						−0.001		
						(0.002)		
$\text{FFH20}_{\text{t-1}}$			0.042***					
			(0.010)					
$\text{FFL20}_{\text{t-1}}$							−0.001	
							(0.002)	
Constant	−0.162***	−0.130***	−0.130***	−0.157***	0.002	0.001	−0.000	0.003
	(0.038)	(0.039)	(0.039)	(0.038)	(0.008)	(0.009)	(0.008)	(0.008)
Observations	1,033	1,033	1,033	1,033	1,033	1,033	1,033	1,033
R^2	0.068	0.086	0.085	0.088	0.035	0.035	0.035	0.037
Adjusted R^2	0.061	0.078	0.077	0.077	0.027	0.026	0.026	0.026

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

7 Conclusion

This paper examined the impact of noise trader in the financial market by first constructing the noise trader variable as the residuals of the sentiment regression not explained by the asset pricing fundamentals. The mispricing was created using the Fama-French three-factor model. Furthermore, the mispricing of the small and large-cap portfolios was used to examine the noise trader behavior, but also the extent of the small cap's limits of arbitrage on the mispricing. Consistent with Podolski et al. (2009), I conclude that the noise trader and sentiment's impact is more significant in small-cap portfolio compared to the large-cap portfolio. However, I note that the substantial interaction in the irrational sentiment and the past large-cap mispricing may indicate that the impact of noise trader—when concentrated together—has a significantly negative relationship with the small-cap portfolio while having a slightly positive significant relationship with the large-cap portfolio. Furthermore, after regressing four weeks past mispricing, I also note that compared to the large-cap portfolio, the mispricing of the small-cap portfolio reverts back to its fundamental price later than the large-cap. Although further research using market volumes are needed to test the interpretation on the impact of noise trader's concentration on the certain markets, my analysis presents not only the noise trader's main impact to the small-cap stocks but also the interaction effect with the previous week's large-cap stocks mispricing to the current week's mispricing of both the small and large cap portfolios.

8 References

[1] Aabo, Tom, Christos Pantzalis, and Jung Chul Park. "Idiosyncratic Volatility: An Indicator of Noise Trading?" Journal of Banking Finance 75 (2017): 136–51. https://doi.org/10.1016/j.jbankfin.2016.11.003.

[2] Baker, Malcolm, and Jeffrey Wurgler. "Investor Sentiment in the Stock Market." Journal of Economic Perspectives 21, no. 2 (2007): 129–51. https://doi.org/10.3386/w13189.

[3] Booth, James R., and Lena Chua Booth. "Economic Factors, Monetary Policy, and Expected Returns on Stocks and Bonds." FRBSF ECONOMIC REVIEW 2 (1997): 32–42.

[4] Brown, Gregory W. "Volatility, Sentiment, and Noise Traders." Financial Analysts Journal 55, no. 2 (1999): 82–90. https://doi.org/10.2469/faj.v55.n2.2263.

[5] Brown, Gregory W., and Michael T. Cliff. "Investor Sentiment and Asset Valuation." The Journal of Business 78, no. 2 (2005): 405–40. https://doi.org/10.1086/427633.

[6] Cao, Jie, and Bing Han. "Idiosyncratic Risk, Costly Arbitrage, and the Cross-Section of Stock Returns." Journal of Banking Finance 73 (2016): 1–15. https://doi.org/10.1016/j.jbankfin.2016.08.004.

[7] Chae, Joon, and Cheol-Won Yang. "Why Do Some Asset Pricing Models Perform Poorly? Evidence from Irrationality, Transaction Costs, and Missing Factors." Seoul Journal of Business 22, no. 1 (2016): 1–64. https://doi.org/10.35152/snusjb.2016.22.1.001.

[8] Corredor, Pilar, Elena Ferrer, and Rafael Santamaria. "Sentiment-Prone Investors and Volatility Dynamics between Spot and Futures Markets." International Review of Economics Finance 35 (2015): 180–96. https://doi.org/10.1016/j.iref.2014.09.013.

[9] Doukas, John A., Chansog (Francis) Kim, and Christos Pantzalis. "Arbitrage Risk and Stock Mispricing." Journal of Financial and Quantitative Analysis 45, no. 4 (2010): 907–34. https://doi.org/10.1017/s0022109010000293.

[10] Fama, Eugene F., and Kenneth R. French. "Common Risk Factors in the Returns on Stocks and Bonds." Journal of Financial Economics 33, no. 1 (1993): 3–56. https://doi.org/10.1016/0304-405x(93)90023-5.

[11] Friedman, Milton. "The Case For Flexible Exchange Rate." In Essays in Positive Economics, 157–203. Chicago, IL: The University of Chicago Press, 1953.

[12] Guidolin, Massimo, and Andrea Ricci. "Arbitrage Risk and a Sentiment as Causes of Persistent Mispricing: The European Evidence." The Quarterly Review of Economics and Finance, 2019. https://doi.org/10.1016/j.qref.2019.05.006.

[13] Hong, Harrison, Terence Lim, and Jeremy C. Stein. "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies." The Journal of Finance 55, no. 1 (2000): 265–95. https://doi.org/10.1111/0022-1082.00206.

[14] Jeffrey Wurgler Website. Accessed November 2019. http://people.stern.nyu.edu/jwurgler/.

[15] Jones, Charles M, and Owen A Lamont. "Short-Sale Constraints and Stock Returns." Journal

of Financial Economics 66, no. 2-3 (2002): 207–39. https://doi.org/10.1016/s0304-405x(02)00224-6.

[16] Kenneth R. French Data Library. Accessed November 2019. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[17] Tim Koller, Marc Goedhart, and David Wessels, VALUATION: Measuring and Managing the Value of Companies, 5th ed. (Wiley, 2010), p.250-251.

[18] Lamont, Owen A, and Richard H Thaler. "Anomalies: The Law of One Price in Financial Markets." Journal of Economic Perspectives 17, no. 4 (2003): 191–202. https://doi.org/10.1257/089533003772034952.

[19] Lee, Wayne Y, Christine X Jiang, and Daniel C Indro. "Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment." Journal of Banking Finance 26, no. 12 (2002): 2277–99. https://doi.org/10.1016/s0378-4266(01)00202-3.

[20] Long, J. Bradford De, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. "Noise Trader Risk in Financial Markets." Journal of Political Economy 98, no. 4 (1990): 703–38. https://doi.org/10.1086/261703.

[21] Mitchell, Mark, Todd Pulvino, and Erik Stafford. "Limited Arbitrage in Equity Markets." The Journal of Finance 57, no. 2 (2002): 551–84. https://doi.org/10.1111/1540-6261.00434.

[22] Podolski, Edward J., Petko S. Kalev, and Huu Nhan Duong. "Deafened by Noise: Do Noise Traders Affect Volatility and Returns?," 2009, 1–35. https://doi.org/10.2139/ssrn.1253042.

[23] Poterba, James, and Lawrence Summers. "Mean Reversion in Stock Prices: Evidence and Implications." Journal of Financial Economics 22, no. 1 (October 1988): 27–59. https://doi.org/10.1016/0304-405X(88)90021-9.

[24] Qadan, Mahmoud, and David Y. Aharon. "Can Investor Sentiment Predict the Size Premium?" International Review of Financial Analysis 63 (2019): 10–26. https://doi.org/10.1016/j.irfa.2019.02.005.

[25] Qiang, Zhang, and Yang Shu-E. "Noise Trading, Investor Sentiment Volatility, and Stock Returns." Systems Engineering - Theory Practice 29, no. 3 (2009): 40–47. https://doi.org/10.1016/s1874-8651(10)60010-5.

[26] Ramiah, Vikash, Xiaoming Xu, and Imad A. Moosa. "Neoclassical Finance, Behavioral Finance and Noise Traders: A Review and Assessment of the Literature." International Review of Financial Analysis 41 (2015): 89–100. https://doi.org/10.1016/j.irfa.2015.05.021.c

[27] Sayim, Mustafa, Pamela D. Morris, and Hamid Rahman. "The Effect of US Individual Investor Sentiment on Industry-Specific Stock Returns and Volatility." Review of Behavioural Finance 5, no. 1 (June 2013): 58–76. https://doi.org/10.1108/rbf-01-2013-0006.

[28] Shleifer, Andrei, and Robert Vishny. "The Limits of Arbitrage." Journal of Finance 52 (1995): 35–55. https://doi.org/10.3386/w5167.

[29] Verma, Rahul, and Priti Verma. "Noise Trading and Stock Market Volatility." Journal of Multinational Financial Management 17 (2005): 231–43. https://doi.org/10.2139/ssrn.686915.

[30] Verma, Rahul, Hasan Baklaci, and Gökçe Soydemir. "The Impact of Rational and Irrational Sentiments of Individual and Institutional Investors on DJIA and S & P500 Index Returns." Applied Financial Economics 18, no. 16 (2008): 1303–17. https://doi.org/10.1080/09603100701704272.

[31] Yan, Yuxing. "On the Consistency between the Fama-French Daily and Monthly Factors." SSRN Electronic Journal, 2008. https://doi.org/10.2139/ssrn.1108636.

9 Appendix

9.1 Fama-French Three-Factor

Unlike the traditional Capital Asset Pricing Model (CAPM), which describes the relationship between systematic risk and expected return for assets, Fama and French (1992) 3-Factor Model includes additional systematic factors, which are firm size–through Market equity breakpoints–and book-to-market ratio.[17] Since the historical-average returns on stocks of small firms and stocks with high book-to-market ratio seems to perform better than the CAPM predictions, Fama and French accounted for this size premium in their model. By constructing new factors called SMB (small minus big) to account for the small firm size risk premium and HML (high minus low) to account for the high book-to-market premium, Fama-French's 3 Factor model quantifies the size risk premium.

The SMB is constructed as follows. After double sorting the stocks by both market equity size and book-to-market ratio, the SMB factor is calculated as

$$SMB = \frac{R_{S/L} + R_{S/M} + R_{S/H}}{3} - \frac{R_{B/L} + R_{B/M} + R_{B/H}}{3}. \tag{9}$$

where $R_{S/L}, R_{S/M}$, and $R_{S/H}$ are three portfolio returns based on small/low, small/medium, small/high portfolios. Similarly, $R_{B/L}, R + B/M$, and $R_{B/H}$ are three portfolio returns based on big/low, big/medium, big/high portfolios.[18]

The HML factor is calculated as

$$HML = \frac{R_{S/H} + R_{B/H}}{2} + \frac{R_{S/L} + R_{B/L}}{2}. \tag{10}$$

where $R_{S/H}$ and $R_{B/H}$ are returns from small & high and big & high portfolios. Similarly, $R_{S/L}$and $R_{B/L}$ are returns from small & low and big & low portfolios.

Putting together, in Fama-French 3 model, the return of a stock or portfolio is expressed as

$$r_{i,t} - r_f = \alpha_t + \beta_1 \ (r_{m,t} - r_f) + \beta_2 \ SMB + \beta_3 \ HML + \epsilon_{i,t} \ , \tag{11}$$

where

α = its abnormal return at time t.

β_i = respective sensitivity to its factor.

$\epsilon_{i,t}$ = unsystematic risk of security i in period t.

[17]Book-to-Market ratio is the accounting value of the company (company's historical cost) divided by the market value of the company.

[18]The "small/low" portfolio is formed by stock of portfolios that are small quantile of market equity (small) and bottom 30% of B/M ratio (low). Similarly, the "big/high" portfolio is formed by stocks of portfolio that are high quantile of market equity (big) and top 30% of B/M ratio (high).

In comparison, the CAPM expresses the return of a stock or portfolio as

$$r_{i,t} - r_f = \alpha_i + \beta(r_{m,t} - r_f) + \epsilon_{i,t}, \tag{12}$$

where

$r_{i,t}$ =return of stock or portfolio i in period t

r_f = risk free rate

α_i = asset's alpha or its abnormal return

β = the relative asset's returns to the market excess return.

9.2 Fama-French Weekly Returns

Since the Fama-French returns were only in daily, monthly, or yearly intervals, the weekly returns were computed as follows. For a week i, it's weekly return $r_{w,i}$ is computed by compounding the daily return from week i's Monday return $r_{d,1}$ to the Friday's return $r_{d,5}$. The formula is

$$\left(1 + \frac{r_{w,i}}{100}\right) = \left[\prod_{j=1}^{5}\left(1 + \frac{r_{d,j}}{100}\right)\right].$$

The compounding method is from Yan (2008), and python was used to convert the daily returns to the weekly returns.

9.3 Fama-French Risk Free rate

The variable $T30$, the risk free rate from the 1-month Treasury bill, is from Fama-French's dataset. Since the Fama-French data i is in monthly yield but the FRED data like 3-month Treasury bill (Series id: DGS1M) is in annual yield, for the $T90\_T30$ variable, the 3-month Treasury bill was converted to the monthly yield to match the 1-month Treasury bill. For example, with j 3-month Treasury bill in annual yield, the yield j was converted to monthly yield by $\left(1 + \frac{j}{100}\right)^{1/12} * 100$.

Furthermore, we use monthly instead of weekly intervals since while FRED have weekly averaged yields of 1-month Treasury bill, the data is limited to year 2002 instead of 2000. Therefore, Fama-French's 1-month Treasury yield data was used to to keep more data observations.

9.4 Variable Reference

Table 14: Sentiment Variables Descriptions

Variables	Description	Data Value	Frequency	Source
AAII	Bull-Bear Sentiment survey data	Percentage in decimal (e.g. 50% = 0.5)	Weekly	https://www.aaii.com/sentimentsurvey
SMB	Fama-French's Small minus Big factor	Percentage in decimal * 100 (e.g. 50% = 50)	Weekly	Kenneth R. French's website
HML	Fama-French's High minus Low factor	Percentage in decimal * 100 (e.g. 50% = 50)	Weekly	Kenneth R. French's website
Mkt	Market Excess Rate (Market rate minus risk free rate)	Percentage in decimal * 100 (e.g. 50% = 50)	Weekly	Kenneth R. French's website
BAaa	Business condition from difference in Baa and Aaa corporate bonds	Percentage in decimal * 100 (e.g. 50% = 50)	Weekly	FRED series ID: WBAA_WAAA
T10Y3M	Future economic expectations. Yield spread between 10 year US treasury bond and 3 month US treasury bill	Percentage in decimal * 100 (e.g. 50% = 50)	Weekly	FRED series ID: T10Y3M
T90T30	Economic risk premium. Measured as difference in yields on 3-month and 1-month Treasury bills	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	FRED series ID: DGS3MO and French's website
T30	Short-term interest rates measured as yield on 1-month US Treasury bill	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	Kenneth R. French's website
IPI	Economic growth, measured as monthly change in industrial production index	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	FRED series ID: INDPRO_CHG
Inflation	Monthly change in CPI	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	FRED series ID: CPALTT01USM661S_CHG
Dividend	Dividend of S&P 500	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	http://www.econ.yale.edu/~shiller/data.htm
MoM	Momentum variable	Percentage in decimal * 100 (e.g. 50% = 50)	Monthly	Kenneth R. French's website

Note that the variables like $\Delta_w Dividend$ and $\Delta_M Dividend$ are constructed by either taking the weekly change (e.g. $\Delta_w Dividend$) or the monthly change (e.g. $\Delta_M Dividend$)

Table 15: Returns Variables Descriptions

Variables	Description	Data Value
Rational	Sentiment predicted from the asset pricing fundamental factors	Percentage in decimal (e.g. 50% = 0.5)
Irrational	Sentiment unpredicted from the fundamentals	Percentage in decimal (e.g. 50% = 0.5)
RH20	Excess market returns of high cap portfolio	Percentage in decimal * 100 (e.g. 50% = 50)
RL20	Excess market returns of low cap portfolio	Percentage in decimal * 100 (e.g. 50% = 50)
FFH20	Fama-French predicted excess returns of high cap portfolio	Percentage in decimal * 100 (e.g. 50% = 50)
FFL20	Fama-French predicted excess returns of low cap portfolio	Percentage in decimal * 100 (e.g. 50% = 50)
MispricedL20	Difference of RH20 and FFH20	Percentage in decimal * 100 (e.g. 50% = 50)
MispricedH20	Difference of RL20 and FFL20	Percentage in decimal * 100 (e.g. 50% = 50)