Retail Derivatives Sentiment and Stock Returns[☆]

Jieyu Wang[a], Neil D. Pearson[b,c], and Qi Zhang[d]

January 31, 2026

Abstract

Issuance volumes of retail structured equity products (SEPs) predict negative returns on the SEPs' underlying stocks. A limitation of this finding is that each month the number of unique underlying stocks is typically only about 100; thus, each month SEP issuances predict the returns of only about 100 stocks. We extend the predictions to the full cross-section of stocks by training large language models to use news headlines to predict a SEP-based measure of sentiment, and then compute the extrapolated sentiment measure for all stocks. The extrapolated sentiment measure is a strong predictor of stock returns.

Keywords: Investor sentiment, structured equity products, retail derivatives, large language models. news, stock market bubbles, label noise

---

☆ This research is supported by the National Science Foundation of China [grant 72373097].

[a] Antai College of Economics and Management, Shanghai Jiaotong University, No. 1954 Huashan Road, Minhang District, Shanghai, China 200030. E-mail: jieyuwang@sjtu.edu.cn.

[b] Department of Finance, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, Illinois, USA 61820. Tel.: +1 (217) 244-0490. E-mail: pearson2@illinois.edu

[c] Canadian Derivatives Institute, 3000, Chemin de la Côte-Sainte-Catherine, Montréal, Québec H3T 2A7, Canada.

[d] Antai College of Economics and Management, Shanghai Jiaotong University, No. 1954 Huashan Road, Minhang District, Shanghai, China 200030. Tel.: +86 (21) 5230 1280. E-mail: zhang.qi@sjtu.edu.cn.

Retail Derivatives Sentiment and Stock Returns

January 31, 2026

Abstract

Issuance volumes of retail structured equity products (SEPs) predict negative returns on the SEPs' underlying stocks. A limitation of this finding is that each month the number of unique underlying stocks is typically only about 100; thus, each month SEP issuances predict the returns of only about 100 stocks. We extend the predictions to the full cross-section of stocks by training large language models to use news headlines to predict a SEP-based measure of sentiment, and then compute the extrapolated sentiment measure for all stocks. The extrapolated sentiment measure is a strong predictor of stock returns.

Keywords: Investor sentiment, structured equity products, retail derivatives, large language models. news, bubble, label noise

1. Introduction

Recent research shows that issuance volumes of U.S. retail structured equity products (SEPs) predict negative abnormal returns on the SEPs' underlying stocks (Henderson, Pearson, and Wang (2023; HPW)). HPW interpret SEP issuance volume to be a proxy for positive retail sentiment, and the negative returns to be reversals of overvaluation due to that sentiment. A limitation of their analysis is that each month the number of unique underlying stocks is typically only about 100; as a result, each month SEPs issuance volumes predict the next-month returns of only about 100 stocks. We extend the predictions to the full cross-section of U.S. stocks by training large language models (LLMs) to use news headlines to predict the HPW SEP-based sentiment measure, and then compute the extrapolated sentiment measure for all stocks. The extrapolated sentiment measure computed during a month is a strong predictor of a stock's abnormal returns during the next month.

The main contribution of this paper is to show that measures of retail investor sentiment constructed by processing news headlines using LLMs are strong predictors of monthly stock returns. Using a range of factor models that have been proposed in the literature, the estimated abnormal returns (factor model "alphas") on long-only portfolios formed using the extrapolated sentiment measure range from –0.65% to –0.45% per month, which translate to between –7.8% and –5.4% per year, and are highly significant.[1] The fact that the sentiment predictor is constructed using news headlines indicates that news is an important determinant of retail investor sentiment. While the relation between news and sentiment is complicated, it can be captured by LLMs. A second contribution of this paper is to show that news-based sentiment is correlated with of Greenwood, Shleifer and You (2019) stock-level bubbles. This relation between bubbles and news also can be captured by LLMs.

We do not train the LLMs using returns directly, but instead train them to predict the HPW SEP-based sentiment measure, which predicts returns. To understand

[1] SEPs issuances, being truncated at zero, identify only overvalued stocks. As a result, our sentiment measure based on these issuances identifies only overvalued stocks.

why we do this, suppose, counterfactually, that one knew stocks' expected returns. Then one would of course train the LLMs on expected rather than realized returns, thereby avoiding the noise in realized returns and the associated risk of overfitting. While we do not know stocks' expected returns, a good predictor, the SEP-based sentiment measure, is available for a subset of stocks. Thus, we face the common bias-efficiency tradeoff. We can either train the LLMs on realized returns, which are unbiased but noisy estimates of expected returns, or we can train the LLMs on the SEP-based sentiment measure, a presumably biased proxy for expected returns. In our case, training the LLMs on the presumably biased proxy is very effective.

We begin with two datasets, news headlines from Thomson Reuters and the data on monthly issuance volumes of retail SEPS collected by HPW. We exclude the headlines of news articles that are exclusively provided to Thomson Reuters subscribers, because these are presumably not available to most retail investors. The HPW SEPs issuance data include, for each stock and month, a normalized measure of the volume of SEPs based on that stock issued during each month. HPW show that the normalized measure of SEP issuance volume is a strong predictor of stock returns over the month following the month the SEPs were issued, and a weaker predictor of returns over the next month. As indicated above, a limitation of the HPW analysis is that the SEPs-based predictor is typically available for only about 100 stocks per month. Each month we use the set of stocks for which the predictor is available and the news headlines to train LLMs, and then use the trained models to extend the predictions to all stocks.

First, we use LLaMA2 to estimate investors' emotional responses to the news headlines. Then, each month, we take the set of stocks for which the SEP-based sentiment measure is available that month. We combine the estimated emotional responses and the SEP-based sentiment measure to train BERT under the unique methodology that ensures the BERT predictions approach the real potential SEP sentiment.[2] For each month, we then use the trained BERT model and the news

[2] The data used for BERT training extend through Oct 2018 and the training and fine tuning data of LLaMa2 extend through July 2023.

headlines for all stocks to compute a sentiment measure for each stock. The result of this is that we have a sentiment measure for each stock. For the 100 or so stocks on which SEPs were issued, we have the HPW measure constructed from the issuance volumes. For the other stocks, we have the extrapolated sentiment measure computed from the trained BERT model.

SEPs issuance volumes, which we hypothesize are associated with positive sentiment and overvaluation, cannot be negative. Thus, we expect that the sentiment measures are associated with future negative abnormal stock returns.

To test this prediction, we examine the abnormal returns on portfolios constructed using the sentiment measures. Specifically, we follow HPW and normalize the sentiment measures so that they sum to one, which allows them to be interpreted as portfolio weights. Then, for each month $t + 1$, we form a sentiment-weighted portfolio return using the extrapolated sentiment measure computed using month t data. We then use the sentiment-weighted portfolio returns to estimate calendar time factor model regressions using a range of different factor models, and test whether the abnormal returns (i.e., the factor model "alphas") are significantly different from zero.[3] The estimated abnormal returns (factor model "alphas") on long-only portfolios formed using the LLM and SEP-based sentiment measure range from –0.65% (t-statistic –4.05) to –0.45% (t-statistic –2.65) per month. We find similar results when we exclude the smallest stocks.

We also explore whether the sentiment measure is correlated with stock-level bubbles, as defined by Greenwood, Shleifer and You (2019). Results from panel regressions show that the sentiment measures are correlated with Greenwood, Shleifer, and You (2019) bubbles.

We provide additional evidence associated with our hypotheses. This includes showing that there is positive and significant relationship between the original HPW SEP sentiment measure and a stock-level options-based sentiment computed by

[3] The benchmarks are the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, the Hou, Xue and Zhang (2015) q-factor model, the Hou et al. (2021) $q5$-factor model, the Stambaugh and Yuan (2017) mispricing factor model, and the Daniel, Hirshleifer and Sun (2020) behavioral factor model.

applying the Gao and Martin (2021) approach to compute market-level sentiment to individual stocks. However, there is no relation between extrapolative SEP sentiment and the options-based sentiment measure.

As discussed above, our approach involves training LLMs to predict SEP sentiment, which predicts returns, rather than training them directly on returns. We do this because realized returns, while unbiased estimates of expected returns, are very noisy; in the context of LLMs, they suffer from label noisiness. This refers to inaccuracies, errors, or inconsistencies in the labels or target values within a dataset used to train an LLM. This phenomenon is detrimental to model training, causing poor generalization, overfitting, and inaccurate predictions

We develop a theoretical model to analyze the impact of label noisiness on LLM learning, and show that if the informativeness of the return η_Y is sufficiently low compared to the informativeness of SEP sentiment η_Z, then training an LLM on SEP sentiment yields better return predictions compared to training the LLM on returns. The exact condition is determined by the correlation between SEP sentiment and expected returns.

Lastly, we validate our interpretation of how the low label noise in SEP sentiment contributes to LLM learning and extrapolation. We conduct another classification learning task in which we compare two possible labels, (i) an indicator variable that equals one if there was a SEP issue, and (ii) an alternative indicator variable that equals one if the underlying stock return is negative in the next month. Both indicator variables predict next-month returns. They differ in that the SEP indicator is more informative, as we show. The model predicts that the more informative variable, the SEP indicator, should be a stronger predictor, and it is.

1.1 Related literature

Our paper is related to the literatures on retail investor sentiment, the use of news to predict stock returns, and especially the use of LLMs to process news data. HPW propose the SEP sentiment measure and show that the SEP sentiment predicts negative abnormal returns on the SEPs underlying stocks. This is the starting point of our paper. Henderson and Pearson (2011) and Vokata (2020) show that expected

returns on SEPs are typically less than the risk-free rate and often negative, while Célérier and Vallée (2017) argue that investors are attracted to SEPs by their salient high headline rates. These papers are consistent with the hypothesis that SEP issuance volumes reflect the demands of unsophisticated investors.[4]

Several other papers also use LLMs to process news data and either predict returns or estimate sentiment. Chen, Kelly and Xiu (2024) show that LLMs, BERT, RoBERTa, LLaMA, ChatGPT, LLaMA2, can process news to predict stock returns. They find that the Sharpe ratios of equal-weighted strategies are over 2.87, while those of portfolios formed using the traditional NLP method are only around 2.3. Chen et al. (2025) use ChatGPT and DeepSeek to extract information from *Wall Street Journal* news, and find that it can predict the performance of both the stock market and the macroeconomy. Bybee (2024) uses LLMs to analyze historical news over the past 120 years and simulate how beliefs are generated and biasness arise. Our paper differs from these in that it uses a proxy for sentiment to better estimate the relation between news sentiment and abnormal returns.

Other papers that use news or news-derived sentiment to predict stock returns or stock return volatility includes Audrino, Sigrist, and Ballinari (2020) Barbaglia, Consoli, and Manzan (2023), Bybee et al. (2024), Hirshleifer, Mai, and Pukthuanthong (2025), Jeon et al. (2022), and Telock (2007).

Gao and Martin (2021) propose a market-wide sentiment measure based on listed option price, and provide evidence that it is associated with bubbles. Their computation can be applied to individual stocks with sufficiently many options, to compute a stock-level bubble measure. We study the relationship between this sentiment and our SEP sentiment, and whether it predicts stock returns.

Greenwood, Shleifer and You (2019) propose an empirical definition of a stock-level bubble. Using a version of the Greenwood, Shleifer and You (2019) definition, Arif and Sul (2024) provide evidence that industry-level accruals can predict industry-level bubbles.

[4] Calvet et al. (2023) take a more positive view of retail structured products.

1.2 Plan of the paper

The next section of the paper describes the news and SEP sentiment data. Section 3 describes the approach used to extrapolate SEP sentiment, and Section 4 presents the results showing that the extrapolated sentiment predicts abnormal stock returns. Section 5 examines the association between SEP sentiment and an estimate of stock-level sentiment computed by applying the Gao and Martin (2021) approach for estimating market-level sentiment to individual stocks. Section 6 presents evidence that the sentiment measures predict the stock-level bubbles defined by Greenwood, Schleifer, and You (2019). Section 7 analyses the theoretical model about LLM learning in the presence of noisy labels, and shows that the model implies that our approach of training the model on a proxy for expected returns is superior to training it on realized returns, if the correlation between the proxy and expected returns is sufficiently high and realized returns are noisy. Section 8 briefly concludes.

2. Data

We use two main data, U.S. news from Thomson Reuters and the stock-level sentiment measure that HPW construct from issuances of retail SEPs. In some analyses, we also use options prices and dividend-price ratios to compute a stock-level version of the market-wide sentiment measure proposed by Gao and Martin (2021). We also use stock returns and some other data from the Center for Research in Securities Prices (CRSP).

2.1 U.S. news data and pre-processing

The news dataset consists of third-party news from Thomson Reuters covering the period running from January 2004 through December 2024. Figure 1 shows that the number of news items per year varies from slightly under 150,000 in 2004 to about 350,000 in 2024, for a total of about 5.3 million items. The mean number of news items on a stock in a month is about four. From January 2004 to December 2024, there are only 364 stock-date pairs for which the number of news items exceeds 50.

As the terminology suggests, third party news excludes those Thomson

Reuters news articles provided exclusively to Thomson Reuters subscribers. Excluding the Thomson Reuters exclusive news is desirable, because those articles are unlikely to be seen by many retail investors. In contrast, the third-party news, which includes news from widely circulated sources such as the *Wall Street Journal* and other leading newspapers, are more likely to be noticed by individual investors and more likely to affect retail investor sentiment.

The news items include a timestamp indicating when the item became available, stock identifier (ticker symbol), news title (headline), and news content. In this version of the paper, we use only the news title, not the news content. One reason to do this is that many investors are likely to suffer from limited attention, and might not take the time to read and understand the full news content.[5] We associate a CRSP PERMNO identifier with each news items by matching the CRSP data to the news items using the ticker symbols, which are available in both databases and unique for each date. The CRSP PERMNOs are used to identify the appropriate stock returns in the CRSP database.

The timestamps are based on Coordinated Universal Time (UTC+00:00), while the trading time of U.S. stock market is based on the Eastern Time (UTC+04:00 in summer and UTC+05:00 in winter). We need to convert Eastern time to UTC+0, since the raw timestamp is based on Greenwich Time (UTC+0). After that, to avoid data leakage, we calculate the corresponding trading date of each piece of news under the rule that if news happened before stock market opening time then the corresponding trading date is today else the next trading date. Similarly, if a news happened during a weekend or holiday, the corresponding trading date is also the next trading date after holiday ends. For instance, in Eastern Time, if a news happening at 8 o'clock in 8 January 2004 (Thursday), the corresponding trading date is 8 January 2004, while for news happening at 10 o'clock in 8 January 2004, the corresponding trading date is 9 January 2004 (Friday) and for news happening in 10 January 2004 (Saturday), the corresponding trading date is 12 January 2004 (Monday).

[5] Of course, it is possible that the full news content helps explain and predict sentiment. We plan to also use the full news content in future work.

This procedure results in an intraday-level news dataset that includes the stock identifier (PERMNO), trading date, timestamp, news title, and news content, from January 2004 through December 2024. For each combination of stock and trading date, we concatenate the news titles into a long text. In the rare case when the number of news items for a stock-date exceeds we take only news written by English and sourced from one of several national media.[6] The result of this procedure is a stock-date level news dataset with approximately three million observations.

2.2 SEP sentiment data

For each combination of stock and month, HPW constructed a normalized measure of issuance volumes of SEPs based on that stock, and interpret it as a measure of the sentiment of unsophisticated retail investors during the month.[7] Among other things, HPW note that most U.S. SEPs based on individual stocks have expected returns less than the risk-free rate, and many have negative expected returns (Henderson and Pearson (2011) and Vokata (2020)). Thus, it is unlikely that many sophisticated investors buy SEPS.

Letting $Q_{i,t}$ denote the total proceeds of all SEPs based on of stock i in month t and $V_{i,t-1}$ denote the capitalization of stock i at the end of the previous month $t-1$, HPW first compute the ratio of SEP issuance volume to underlying stock market capitalization, $Q_{i,t}/V_{i,t-1}$. They then compute their sentiment measure $G_{i,t}$ by normalizing the ratios $Q_{i,t}/V_{i,t-1}$ to that the normalized ratios sum to one:

$$G_{i,t} = \frac{Q_{i,t}/V_{i,t-1}}{\sum_i Q_{i,t}/V_{i,t-1}}.$$

If no SEP on based on stock *i* is issued during month *t* then $Q_{i,t} = G_{i,t} = 0$. An advantage of the normalization is that the resulting $G_{i,t}$ can be interpreted as portfolio weights. HPW use the $G_{i,t}$ to form sentiment-weighted portfolios, and find that these portfolios have large negative abnormal returns in the first month after portfolio formation, with some evidence of negative returns in the second post-

[6] From January 2004 to December 2024, there are only 364 stock-date pairs whose number of news is over 50.

[7] Among other things, HPW note that most U.S. SEPs based on individual stocks have expected returns less than the risk-free rate, and many have negative expected returns. (See Henderson and Pearson (2011) and Vokata (2020).) Thus,

formation month.

The replication package for the HPW paper provides information about the approximately 38,000 SEPs issues in their sample, which runs from January 2004 through December 2019. The HPW replication package also provides, for each stock and month, the value of the SEP sentiment measure $G_{i,t}$.

A limitation of HPW is that the sentiment measures $G_{i,t} = 0$ for all but about 100 stocks in most months. The underlying stocks are commonly volatile technology stocks. The stocks of well-known energy and financial companies also frequently appear as SEPs' underlying stocks (HPW Figure 2). We want to extend the HPW sentiment measure to the full cross-section of stocks.

Why are SEPs based on only a subset of stocks issued? The quantity of SEPs based on a stock depends upon both investor demand for exposure to the stock and the issuers' willingness to supply SEPs based on the stock. Most SEPs based on individual stocks are yield enhancement products, especially various forms of reverse convertibles, in which the issuer provides a high coupon on the SEP and retains an embedded put option (Vokata (2020), HPW). The issuer can only provide a high coupon if the retained embedded put option is valuable, that is, if the underlying stock volatility is high.[8] Thus, even if retail investor sentiment about a stock is positive, there is unlikely to be a SEP issue unless volatility is also high. Second, issuance of SEPs based on mid- or small-capitalization stocks seem unlikely even if retail sentiment is high, because the trading revenue from a small SEP issue is less likely to cover the fixed costs of issuing the SEP. Third, we conjecture that SEPs based on larger stocks with limited retail name recognition are unlikely for the same reason.

2.3 Additional data

We use the Gao and Martin (2021) stock-level sentiment measure *B* in some analyses. We construct it as they do, using options and forward prices from OptionMetrics and dividend-price ratios from the stocks characteristics database

[8] Cite the HPW probit regressions.

contributed by Jensen, Kelly, Pedersen (2023).[9] We also use the stock-level change in net operating assets (ΔNOA), computed as in Arif and Sul (2024) do using total assets, cash and short-term investments, total liabilities, short-term debt, and long-term debt. These data items are from Compustat.

We also use monthly stock returns, stock prices, and shares outstanding from CRSP. We obtain the factor portfolio returns needed to estimate the factor models from the various researchers' websites. The Fama and French (1993, 2015) factors and the Carhart (1997) momentum factor are from Kenneth French's website.[10] The Hou, Xue and Zhang (2015) and Hou et al. (2021) q and $q5$ factors are from Lu Zhang's website.[11] The Daniel, Hirshleifer and Sun (2020) behavioral factors are from Lin Sun's website,[12]and the Stambaugh and Yuan (2017) mispricing factors are from Robert Stambaugh's website.[13]The risk-free rate data used to compute excess returns are from Kenneth French's website.

3. Extrapolating SEP Sentiment to the Full Cross-Section of Stocks

For most stock months, SEP issuance volume $Q_{i,t} = 0$. As an initial step, we first consider the set of stock-months for which we need to extrapolate the sentiment measure. If the stock has a history of previous SEP issues, we interpret $Q_{i,t} = 0$ to mean that retail investor sentiment is negative, and do not compute an extrapolated sentiment measure for this stock. However, if $Q_{i,t} = 0$ and the stock has no history of SEP issues, we compute an extrapolated sentiment measure for the stock. The extrapolated sentiment measure can turn out to be zero or close to zero, indicating low sentiment for the stock.

We extrapolate the sentiment measures using two LLMs, LLaMA2 released by Meta AI in 2023, and BERT, released by Google in 2018.[14] We use LLaMA2 to estimate investors' emotional responses to the news headlines. We then use BERT to

[9] https://wrds-www.wharton.upenn.edu/pages/get-data/contributed-data-forms/global-factor-data/
[10] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[11] https://global-q.org/factors.html
[12] https://sites.google.com/view/linsunhome
[13] https://finance.wharton.upenn.edu/~stambaug/
[14] The data used for BERT training is up to Oct 2018 and the training and finetuning data of LLAMA2 is up to July 2023.

extrapolate SEP sentiment to the stock-months for which the HPW sentiment measure $G_{i,t}$ is not available. In training BERT, for each stock the monthly emotional response is the feature and the monthly SEP sentiment measure G_{it} is the label. We call the prediction of SEP sentiment derived from BERT after training extrapolated SEP sentiment.

We use BERT in two steps, conducting two types of supervised learning. First, we carry out a classification step to predict which stocks would have SEP issuance, and thus positive SEP sentiment. Second, we use regression to predict the magnitude of positive SEP sentiment would be. If prediction from the classification training is zero, i.e., no SEP issuance, then the extrapolated SEP sentiment is also zero.

3.1 Construction of the full dataset for Classification and Regression

Feature and label are important items in the construction of a dataset for supervised learning. We construct a well-designed prompt for emotional response generation by LLaMA2, and we take the LLaMA2 emotional response as the feature. The prompt is organized as: task description, company name, all news titles for that company on a trading date, and the instruction. An example follow:

> Now, your role is an individual investor holding stock of DELL.O company, how do you feel about after reading ALL these news?
>
> # News 1 #
>
> REMINDER/Dell Outlines Activities at 2004 International Consumer <DELL.O>
>
> # News 2 #
>
> Dell Lowers Networking Costs for Customers with New PPowerConnect <DELL.O>
>
> News End! Please give me a SHORT, BRIEF but PRECISE description about your emotion to ALL these NEWS within ONE SENTENCE!

Given this instruction, LLAMA2 will generate an emotional response under order. An example response is:

> After reading these news, I feel optimistic about Dell's efforts to expand its product offerings and improve customer experience, which could potentially drive growth and increase shareholder value.

This procedure transforms daily stock-level news into daily stock-level responses.

This makes the underlying optimism or pessimism explicit, and also make our later training more efficient. After that, for each stock and month we concatenate all responses together into a long text. This text is the emotional response that is used as the feature in both classification and regression training of BERT.

In the classification step, for each stock and month, the label is an indicator that equals one if there was a SEP issue based on the stock during the month. If no SEP based on a stock is issued during a month, the SEP indicator for this combination of stock and month is zero. In the regression step, the label is the (positive) SEP sentiment measure $G_{i,t}$ and the feature is also the emotional response.

3.2 Choice of training, validation, and testing datasets

How to balance fitting and over-fitting may be a permanent question in machine learning, and the design of the validation method may be a possible solution. The validation must ensure that the distribution of validation of the validation sample is close to that of testing, or out-of-sample part, so that the training would be efficient and avoid over-fitting.

In classification training, the testing, or out-of-sample part, consists of those stock-month pairs that have no SEP issuance history and the in-sample part (including training and validation) is those stock-month pairs that have SEP issuance history or newly have SEP issuance in this month. By the definition of $Q_{i,t}$, if one stock-month pair's SEP sentiment is missing, this pair should be in the out-of-sample part; if not, this pair should be in-sample part.

Obviously, we can directly find out the most significant difference of in-sample and out-of-sample data is how many times does one stock have SEP issuance record from past time to this month, we denote this as SEP count. The SEP count of in-sample data are all at least one, while are all zero in the out-of-sample data.

Based on this, we propose a novel rule to select the validation part. We choose the part of in-sample data whose SEP count is the last 25% in every month as validation dataset and the remaining part, whose SEP count is the top 75%, as the training dataset. Under this design, machine would first learn the relation between

response and SEP sentiment of most frequent SEP issuance stocks, then this relation would be tested in those less frequent SEP issuance stocks and save the best for prediction of out-of-sample, also the extrapolation.

Similarly, in regression training, the training part also consists of the most frequent SEP issuance stocks whose SEP count is the top 75% and the validation part also consists of the less frequent SEP issuance stocks whose SEP count is the last 25%. Construction of testing dataset is different from what we do in classification training. Only stock-month pairs that would have extrapolative SEP issuance are in the out-of-sample of regression training set.

3.3 BERT training design

We design a recursive rolling-by-month pipeline for BERT classification and regression training. The meaning of recursive and rolling-by-month is that every month we train new BERT models for both the classification and regression task where the start point of training in every month is the optimal model derived in the training in previous month, except for the first month (January 2004). The starting poinst of both classification and regression training in the first month (January 2004) are the uncased BERT from Hugging Face.[15] This design lets the machine adopt the time-varying relation between response and SEP issuance or SEP sentiment, and also contributes to more efficient learning.

Figure 2 exhibits the full view of our training design from classification training to regression training. The first column represent the data used for each type of training, the second column shows the starting point of each type of training, the third column represents the optimal model after each training, and the last column displays the predictions from the optimal model in third column.

Consider training in January 2004 as example. Since January 2004 is the first month in our extrapolation data, the starting point is the uncased BERT. As for the other months, like February 2004, the starting point is the BERT after learning the data in previous month, January 2004. We train this uncased BERT in several epochs

[15] https://huggingface.co/google-bert/bert-base-uncased.

by using training data in January 2004, where the training data entirely consists of frequent SEP issuance stocks whose historical SEP issue times are the top 75% in this month. We use validation data to check which model is the best model in all epochs' training and save the best model as BERT after classification training in January 2004 shown in Figure 2. Subsequently, we use this model to predict whether a stock in out-of-sample, also having no SEP issuance record, would have extrapolative SEP issuance in January 2004. If the prediction is one, this stock will be added into out-of-sample of regression training since the accurate extrapolative SEP sentiment of this stock need a step further prediction from regression training, otherwise the extrapolative SEP sentiment is zero.

The approach to regression trading resembles the approach for classification. Like the classification training, the starting point is the uncased BERT. We train this uncased BERT in several epochs by using training data from January 2004, where the training data entirely consists of frequent SEP issuance stocks whose SEP sentiment is positive and historical SEP issue frequency is in the top 75%. We use validation data to check which model is the best model in all epochs' training, and save the best model as BERT after regression training in January 2004, as shown in Figure 2. Subsequently, we use this model to predict the magnitude of the extrapolative SEP sentiment of a stock in January 2004.

3.4 Construction of extrapolative SEP Sentiment from BERT after training

Similar to the normalization used with the HPW SEP sentiment, we define the extrapolative SEP sentiment $G_{i,t}^{Ex}$ for stock i in month t to be the predicted SEP sentiment scaled by the sum of all predicted SEP sentiments in month *t*,

$$G_{i,t}^{Ex} = \frac{Pred\ SEP\ Sent_{i,t}}{\sum_i Pred\ SEP\ Sent_{i,t}}$$

This scaling results in a set of monthly portfolio weight that sum to one, and span all stocks that have no previous SEP issuance.

4. Does the Extrapolated Sentiment Measure Predict Returns?

HPW show that their SEP sentiment measure $G_{i,t}$ is a strong negative predictor of returns. Their interpretation of this predictability is that $G_{i,t}$ reflects

positive sentiment resulting in overvaluation, and the negative abnormal returns are the reversals of the overvaluation. We hypothesize that extrapolative SEP sentiment is also associated with overvaluation, and is also a negative predictor of returns.

4.1 Abnormal returns on portfolios of stocks with no SEP issuances formed using extrapolated sentiment $G_{i,t}^{Ex}$

We explore whether extrapolated SEP sentiment $G_{i,t}^{Ex}$ predicts returns using the same approach that HPW use to examine whether their measure $G_{i,t}$ predicts returns. Specifically, at the end of every month *t*, we use the $G_{i,t}^{Ex}$ to construct a portfolio, which is held during month *t* +1. The monthly return on the portfolio is

$$R_{t+1} = \sum_{i} G_{i,t}^{Ex} \times r_{i,t+1} ,$$

where $G_{i,t}^{Ex}$ is the extrapolative SEP sentiment for stock i in month t and $r_{i,t+1}$ is the monthly return on stock i in month $t + 1$. Because the SEP data are available from January 2004 through December 2019, this procedure results in a set of monthly portfolio returns extending from February 2004 through January 2020, a total of 192 months. We then calculate the portfolio excess return $R_{t+1} - r_{f,t}$, where $r_{f,t}$ is the risk-free rate quoted at the end of month t, that applies to month *t* + 1.

We test whether the extrapolated sentiment $G_{i,t}^{Ex}$ predicts returns using calendar-time regressions in which we regress the portfolio excess returns $R_{t+1} - r_{f,t}$ on the factors from a range of linear factor models, and test whether the intercepts ("alphas") are different from zero. The factor models we use are the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, the Hou, Xue and Zhang (2015) *q*-factor model, the Hou et al. (2021) *q*5-factor model, the Stambaugh and Yuan (2017) mispricing factor model, and the Daniel, Hirshleifer and Sun (2020) behavioral factor model. We expect the intercepts (alphas) to be negative, i.e., that extrapolated SEP sentiment is a negative predictor of returns. For each factor model, the null hypothesis is that the intercept of the factor model regression is zero.

We report the regression results in Table 1. The results are based on the extrapolated SEP sentiment starting from January 2004 to December 2019. The related investment period is from February 2004 through January 2020, a total of 192 months, for all models except the Stambaugh and Yuan (2017) mispricing factor model. For this model, the mispricing factor data are only available through December 2017.

The results show that intercepts or alphas, across all seven asset pricing models, are all negative and highly significant. The estimated intercepts range from –0.65% *(t*-statistic –4.05) for the Fama-French three-factor model to –0.45% (*t*-statistic –2.65) for the Hou et al. (2015) *q*5-factor model. The associated *p*-values for the tests of the null hypothesis that the intercept equals zero are all much less than 1%. The corresponding annualized abnormal returns range from –7.8% per year (= 12 × –0.65% per month) to –5.4% per year (= 12 × –0.45% per month).

The estimated alphas using extrapolated sentiment in the sample of stock-months for which the original HPW sentiment measure is not available are about half the magnitude of the alphas estimated by HPW when they formed portfolios using their original measure. In HPW, the estimated regression intercepts ranged from –1.25% (*t*-statistic –4.51) in Fama and French (1993) three-factor model to –0.93% (*t*-statistic –3.30) in the Hou et al. (2021) *q*5-factor model. This attenuation of the abnormal returns should not be surprising, because the extrapolation of sentiment to the full cross-section of stocks inevitably results in an extrapolated measure $G_{i,t}^{Ex}$ that is a noisy estimate of the unobserved latent sentiment.

4.2 Abnormal returns on portfolios of stocks with no SEP issuances excluding small stocks

The literature documents the low abnormal returns and apparent mispricing of micro-capitalization stocks, so negative alpha in these stocks may not be surprising. We check the robustness of our results to excluding micro-capitalization stocks by excluding stocks with market capitalization in the lowest 5% as of the end of the month prior to the month during which we compute the sentiment measure. For

example, let *t* be the month during which we compute the sentiment measure $G_{i,t}^{Ex}$ and let *t* + 1 be the performance evaluation period. We exclude stocks with market capitalization in the lowest 5% as of the end of month $t-1$, recompute the monthly portfolio returns excess returns, and re-estimate the factor models.

We report the results in Table 2. Similar to the results in Table 1, the regression intercepts are negative highly significant across all seven asset pricing models. The estimated intercepts range from –0.66% (*t*-statistic –4.16) Fama-French three-factor model to -0.43% (t-statistic -2.61) for the Hou et al. (2015) q5-factor model. The *p*-values for tests of the null hypotheses that the intercepts equal zero are all much less than 1%.

4.3 Abnormal returns on portfolios of stocks formed using $G_{i,t}$ and $G_{i,t}^{Ex}$

For some combinations of stock and month we have the original HPW measure $G_{i,t}$, and for others we have extrapolated sentiment $G_{i,t}^{Ex}$. We now use both to form portfolios from the full cross-section of stocks. Specifically, we construct the full market SEP sentiment as

$$G_{i,t}^{Full} = \frac{G_{i,t}1(G_{i,t})}{2} + \frac{G_{i,t}^{Ex}1(G_{i,t}^{Ex})}{2}$$

where $1(G_{i,t})$ and $1(G_{i,t}^{Ex})$ are both indicator functions that equal one if the argument is positive, and equal zero otherwise. With this definition of $G_{i,t}^{Full}$, if the original HPW sentiment measure $G_{i,t}$ is available for stock i in month t, then $1(G_{i,t}) = 1$, $1(G_{i,t}^{Ex}) = 0$, and $G_{i,t}^{Full} = G_{i,t}/2$. If $G_{i,t}$ is not available, then $G_{i,t}^{Ex}$ is available, $1(G_{i,t}^{Ex}) = 1$, and $G_{i,t}^{Full} = G_{i,t}^{Ex}/2$. With this definition, the sentiment measure $G_{i,t}^{Full}$ sums to one each month, and thus can be used as portfolio weights.

Using these portfolio returns, we estimate seven calendar-time regressions, one for each factor model, and report the estimated alphas in Table 3. Panel A displays the results when we form the portfolios using all stocks, including small stocks. Panel B displays the results when we form the portfolios after first excluding the stocks with market capitalization in the lowest 5% of market capitalizations as of the end of

month t – 1. In both panels, the estimated alphas are negative and highly statistically significant. In Panel A, they range from –0.95% (t-statistic –5.44) to –0.68% (t-statistic –3.83). In Panel B, they range from –0.96% (t-statistic –5.46) to –0.68% (t-statistic –3.81). Thus, these results provide robust evidence that SEP sentiment is associated with negative abnormal returns.

5. Exploring the SEP Sentiment Mechanism

Gao and Martin (2021) combine the options-based bound on expected returns due to Martin (2017) with the dividend to price ratio to construct a market-level sentiment measure. We adapt the Gao and Martin (2021) market-wide measure to create an alternative stock-level sentiment measure.

5.1 Construction of the stock-level options-based sentiment B

Gao and Martin (2021)'s options-based sentiment B is defined as

$$B_t = r_{f,t+1} + LVIX_t + E_t[r_{t+1} - g_{t+1}].$$

where B_t is the market-level sentiment, $r_{f,t+1}$ is the risk-free rate from month t to month $t+1$, and $LVIX_t$, the key component, is an options-based bound on expected returns computed form the prices of listed options and the stock forward price. The options prices, stock forward price, and interest rate are from Option Metrics. The last term $E_t[r_{t+1} - g_{t+1}]$ is the expectation in month t of the difference between stock return and dividend growth rate in the next month. Gao and Martin (2021) measure the dividend growth rate as $y_t = \log(1 + D_t/P_t)$, where D_t and P_t are the dividend and price of the S&P 500 in month t, and D_t/P_t is calculated using dividend and security price data from Option Metrics.

Mimicking the Gao and Martin (2021) measure, we define the options-based stock-level sentiment as

$$B_{i,t} = r_{f,t+1} + LVIX_{i,t} + \log(1 + D_{i,t}/P_{i,t}),$$

where $B_{i,t}$ is the stock-level sentiment for stock i in month t, and $LVIX_{i,t}$ is calculated using the stock forward price and the prices of options on stock using the same approach that Gao and Martin (2021) use to compute the market-wide $LVIX_t$. The dividend price ratio is taken from the stocks characteristics database contributed

by Jensen, Kelly, Pedersen (2023).

5.2 Predictability of stock-level sentiment $B_{i,t}$ on returns over different horizons

We construct a long-short strategy to test the predictability of stock-level sentiment $B_{i,t}$ on return over different horizons, including one month, two months and three months. Consider one month, for example. We buy stocks associated with positive sentiment $B_{i,t}$ by weight of sentiment B after normalization in next month and short stocks associated with negative sentiment $B_{i,t}$. The return on this long-short strategy is:

$$R_{t+j} = \frac{B_{i,t} I(B_{i,t} > 0)}{\sum_i B_{i,t} I(B_{i,t} > 0)} r_{i,t+j} - \frac{B_{i,t} I(B_{i,t} < 0)}{\sum_i B_{i,t} I(B_{i,t} < 0)} r_{i,t+j}$$

where $j \in \{1, 2, 3\}$ indicates the horizon (in months), $I(.)$ is the indicator function, $B_{i,t}$ is the sentiment for stock i in month t, and $r_{i,t+j}$ is the monthly return on stock *i* over a *j* month horizon.

We estimate abnormal returns as the intercepts (alphas) from regressing the long-short returns on the factors from the seven factor models. Table 4 displays the resulting estimates of the alphas of this long-short strategy over one (Panel A), two (Panel B), and three-month (Panel C) horizons. The coefficients and *t*-statistics of alphas in next two and three month are much stronger than that in next one month, which implies the predictability of sentiment *B* mainly focuses on the next two or three months.

5.3 Association of original SEP sentiment and sentiment B

Since sentiment *B* can predict negative return over the next two month and SEP sentiment can predict negative return in next one month, there is one-month lag with sentiment B and SEP sentiment. Thus, we hypothesis that higher sentiment B is associated to higher original SEP sentiment in next month while no significant relationship with extrapolative SEP sentiment since these extrapolative SEP issuances do not virtually exist. Thus, the null hypothesis is that there is no or negative relationship between the original SEP sentiment and the listed option-based sentiment *B*.

To test this null hypothesis, we conduct a panel regression of the original SEP sentiment on the sentiment B with lag one.

$$G_{i,t} = a + bB_{i,t-1} + \sum_{n} \gamma_n Controls_{n,i,t-1} + \varepsilon$$

where $G_{i,t}$ is the positive original SEP sentiment for stock i in month t, $B_{i,t-1}$ is the corresponding sentiment *B* for stock i in month $t-1$ and control variables including $B_{i,t-2}$, market equity, book-to-market, momentum, return on equity and asset growth rate with lag one. Furthermore, we also conduct another panel regression of the extrapolative SEP sentiment on the sentiment B with lag one.

$$G_{i,t}^{Ex} = a + bB_{i,t-1} + \sum_{n} \gamma_n Controls_{n,i,t-1} + \varepsilon$$

where $G_{i,t}^{Ex}$ is the positive extrapolative SEP sentiment for stock i in month t, the variables are the same as the former.

Table 5 reports the results of above panel regressions. Column (1)-(4) report the results of original SEP sentiment and column (5) to (8) report the results of extrapolative SEP sentiment, respectively. The coefficients of $B_{i,t-1}$ are all significantly positive when predicting original SEP sentiment, and the corresponding R^2 coefficients are over 35%. The coefficients on $B_{i,t-1}$ are all not significant when predicting extrapolative SEP sentiment.

5.4 More evidence about the association of original SEP sentiment and sentiment B

To provide more evidence about the association between HPW SEP sentiment and sentiment B, we hypothesis that original SEP sentiment associated with top quantile sentiment B in previous month imply stronger overpricing than full sample.

We also use the long-only strategy to show the impact of the selection based on sentiment B on the magnitude of alpha. For example, in this long-only strategy, we only buy stocks in month $t+1$ that not only have positive original SEP sentiment in month t but also the associated sentiment B is the top 66% among all positive original SEP sentiment stocks in month $t-1$. The return on this long-only strategy is:

$$R_{t+1} = \frac{G_{i,t}^{Ex} I(B_{i,t-1} > q66_{t-1})}{\sum_i G_{i,t}^{Ex} I(B_{i,t-1} > q66_{t-1})} r_{i,t+1}$$

Table 6 reports results of this long-only strategy conducted on different quantile thresholds of sentiment B, including 66%, 50%, 33%, 25%, which are shown by Panel A, Panel B, Panel C and Panel D, respectively. The coefficients of alphas shown by Panel A, presenting case of 66%, are range from -1.36% to -1.12%, which is more negative than that of original SEP sentiment in full sample. Furthermore, as the quantile threshold going up, from Panel A (66%) to Panel D (25%), the coefficients of alphas are more and more negative and also associated with highly significance. These results reveal that given these positive original SEP sentiments, the larger sentiment *B* is associated with the large overpricing, which is consistent with our hypothesis.

6. Full Market SEP Sentiment and Stock-level Bubble Predictability

Greenwood, Shleifer and You (2019) introduce an ex-poste empirical measure of a stock-level bubble. According to that paper, a stock is in a bubble at time *t* if the stock experience a runup of 100% or more in the two years leading up to time t and a decline of a least 40% in the two years after time t. In this section we explore whether the sentiment measures $G_{i,t}$ and $G_{i,t}^{Ex}$ predict stock-level bubbles.

6.1 Construction of the bubble indicator and ΔNOA

To do this, we construct a bubble indicator, as follows:

$$Bubble_{i,t} = I(\exists m \in [1,24], r_{t-m \to t} \geq 100\%) \times I(\exists m \in [1,24], r_{t \to t+m} \leq -40\%),$$

where m is integer $m \in \{1, 2, \ldots, 24\}$, $I(.)$ is indicator function, and $I(\exists m \in \{1,2, \ldots,24\} | \, r_{t-m \to t} \geq 100\%)$ is the runup condition and equals one if there exists a month $t - m$ such that the stock return from that time to month t is greater than 100%. Similarly, $I(\exists m \in \{1,2, \ldots,24\} | \, r_{t \to t+m} \leq -40\%)$ is the crash down condition and equals one if there exist one month $t + m$ such that the stock return from month t to that time is less than –40%. If these two indicator functions hold at the same, we can say there is a bubble for stock i in month t.

Separately, according to Arif and Sul (2024), the change in net operating assets

(ΔNOA) predicts industry-level bubble across many countries, including the United States. For this reason, we construct the U.S. stock-level ΔNOA as control variable in our paper. By definition, ΔNOA is the change in net operating assets between the past fiscal year scaled by the average of total assets in the most recent and prior fiscal year, where the net operating assets equals Total Assets – Cash and Short-term Investment – Total Liabilities + Short-term Debt + Long-term Debt, (Richardson et a. (2005)), where the data are from Compustat.

6.2 Is the HPW sentiment measure $G_{i,t}$ correlated with stock-level bubbles?

To test the predictability of original SEP sentiment, we conduct a set of panel regressions in which we regress the bubble indicator on either an indicator variable for a SEPs issue on stock *i* in month *t* (which implies $G_{i,t} > 0$) or SEP sentiment $G_{i,t}$:

$$Bubble_{i,t} = a + bI(G_{i,t} > 0) + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

$$Bubble_{i,t} = a + bG_{i,t} + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

where $Bubble_{i,t}$ is the bubble indicator defined above, $I(G_{i,t} > 0)$ is the SEP indictor that equals to 1 when SEP sentiment is positive, and $G_{i,t}$ is the SEP sentiment. The control variables include ΔNOA defined above and other stocks characteristics, specifically market equity, book-to-market, momentum, return on assets, and the asset growth rate. These explanatory variables are synchronous with the dependent variable, bubble indicator.

We report the results of the panel regressions in Table 7 reports the results of above panel regressions where column (1) to (4) report the results of the former regressions (SEP indicator as independent variable) and column (5) to (8) report the results of the latter regressions (SEP sentiment as independent variable). The coefficients on $I(G_{i,t} > 0)$ and $G_{i,t}$ are all positive and statistically significant, which is consistent with our hypothesis that original SEP sentiment can predict stock-level bubble.

6.3 Is the extrapolated sentiment measure $G_{i,t}^{Ex}$ correlated with stock-level bubbles?

We use the same approach to test the predictability of extrapolative SEP sentiment, that is, we estimate panel regressions

$$Bubble_{i,t} = a + bI\left(G_{i,t}^{Ex} > 0\right) + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

$$Bubble_{i,t} = a + bG_{i,t}^{Ex} + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

where $Bubble_{i,t}$ is the bubble indicator of stocks having no SEP issuance till month t, $I(G_{i,t}^{Ex} > 0)$ is the extrapolative SEP indicator that equals to 1 when this stock would have SEP issuance predicted by BERT after training, $G_{i,t}$ is the extrapolative SEP sentiment also given by BERT after training and control variables which is the same as former regression.

We report the results in Table 8. Columns (1)-(4) display the results of the first regressions (extrapolative SEP indicator as an explanatory variable) and columns (5)-(8) report the results of the second regression (extrapolative SEP sentiment as an explanatory variable). The coefficients on $I(G_{i,t}^{Ex} > 0)$ and $G_{i,t}^{Ex}$ are all significantly positive, consistent with the hypothesis that extrapolative SEP sentiment predicts stock-level bubbles.

6.4 Is the full-market sentiment measure $G_{i,t}^{Full}$ correlated with stock-level bubbles?

We carry out the same exercise to test the predictability of full market SEP sentiment $G_{i,t}^{Full}$, and estimate the panel regressions

$$Bubble_{i,t} = a + bI\left(G_{i,t}^{Full} > 0\right) + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

$$Bubble_{i,t} = a + bG_{i,t}^{Full} + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon,$$

where $G_{i,t}^{Full}$ is the full market SEP sentiment measure, $I(G_{i,t}^{Full} > 0)$ is the full market SEP indicator that equals to one when this stock actually have or would have SEP issuance, and the control variables are the same as in the previous regression models.

We report the results in Table 9. Columns (1)-(4) display the results of the former regressions (full market SEP indicator as independent variable), and columns (5)-(8)

show the results of the latter regressions (full market SEP sentiment as independent variable). The coefficients on $I(G_{i,t}^{Full} > 0)$ and $G_{i,t}^{Full}$ are positive and significantly different from zero, consistent with the hypothesis that full market SEP sentiment predicts stock-level bubbles.

7. Can Label Noisiness Explain Extrapolative SEP Sentiment's Ability to Predict Returns?

We hypothesize that the predictability of extrapolative SEP sentiment on return or bubble may be owing to the low noisiness attribute of original SEP sentiment, compared with other noisier label, like raw return or return indicator. Existing literatures documents hundreds of factors related to return showing the complexity of return while SEP sentiment is a quite simple label that focuses on whether stock is overpriced.

To investigate this hypothesis, we develop a theoretical model about the effect of LLM learning labels with different noisiness and train BERT to learn return indicator in the same training design and the same stocks in testing dataset.

7.1 LLM Learning Model

7.1.1 Definition of feature and labels

We set up a theoretical model of LLM learning labels with different noisiness and start with the data generating process. Let S denote a latent true signal about fundamental and $\phi(S), h(S), g(S)$ are all transformed fundamental known by insiders.

Empirically, $\phi(S), h(S), g(S)$ can be approximated by linear combination of various fundamental factors and behavioral factors. Take $h(S)$ for example, that means $h(S) \approx \sum_{i=1}^{k} F_i \beta_i$, where F_i is fundamental factor, like size, or behavioral factor, like mispricing.

Feature X and Label Y and Z are generated as

$$X = \phi(S) + \varepsilon_X, Y = h(S) + \varepsilon_Y, Z = g(S) + \varepsilon_Z.$$

where $\varepsilon_X, \varepsilon_Y, \varepsilon_Z$ are mutually independent and also independent of S, have zero mean, and variances $Var(\varepsilon_X) = \sigma_X^2, Var(\varepsilon_Y) = \sigma_Y^2,\ Var(\varepsilon_Z) = \sigma_Z^2,\ Var(Y) > 0$

and $Var(Z) > 0$. Meanwhile, we give some notations here:

$$\rho := \frac{Cov(h(S), g(S))}{\sqrt{Var(h(S))Var(g(S))}}, \eta_Y := \frac{Var(h(S))}{Var(Y)}, \eta_Z := \frac{Var(g(S))}{Var(Z)}.$$

where ρ is the correlation coefficient of $h(S)$ and $g(S)$, η_Y and η_Z are the informativeness of Y and Z, also $1 - \eta_Y$ and $1 - \eta_Z$ can be seen as the noisiness of Y and Z.

7.1.2 Definition of LLM prediction

As the latest modern machine learning model, LLM's structure usually consists of Encode and Decode, whose foundation is many Transformer layers. In addition, a trained (or fine-tuned) LLM usually includes a classifier containing sigmoid (for bio-classifier) or SoftMax (for multi-classifier) and indicator. Thus, we define a trained (or fine-tuned) bio-classifier LLM as a composite function of encoder and decoder function $DE(.)$, sigmoid function $S(.)$, indicator function $I(.)$. In our case, that is,

$$LLM(X) = I\{S[DE(X)] > 0.5\}.$$

Specifically, function $DE(.)$ is a very complicated function not having explicit closed form but the main part of LLM that can transform input X (text and something else) into tokens that can be understand and done linear algebra calculation by LLM. For reverse of input X, like $-X$, output of $DE(.)$ would be also the reverse of $DE(X)$.

For instance, output of word 'optimism' is positive and output of word 'pessimism' is approximately the corresponding opposite number. Thus,

$$DE(-X) \approx -DX(X)$$

Sigmoid function $S(x) = 1/(1 + e^{-x})$ can transform output of $DE(.)$ into probability of positive class in bio-classification and is symmetric about the point (0, 0.5). Finally, through the indicator function $I(.)$, if probability of positive class is greater than 0.5, the output of LLM would be 1, else 0. Like traditional machine learning does, we train LLM for mapping feature to label under supervised learning method. Also, we get out-of-sample prediction about label from LLM and denote predictions as $\hat{Y}$ and $\hat{Z}$.

7.1.3 Assumptions

Assumption 1 (SNR Condition): Signal-to-Noise Ratio(SNR) of Y is less than Z:

$$\frac{Var(h(S))}{\sigma_Y^2} < \frac{Var(g(S))}{\sigma_Z^2} \Leftrightarrow \frac{Var(h(S))}{Var(Y)} < \frac{Var(g(S))}{Var(Z)} \Leftrightarrow \eta_Y < \eta_Z.$$

Assumption 2 (LLM predictions approach to Bayes Optimal): We assume in ideal case predictions of LLM can reach Bayes optimal, that is $E[Y|X]$. While, in real case, there is still a step further driven by a zero-mean random variable ξ from trained or fine tuned LLM to ideal LLM whose prediction can reach Bayes optimal, that is:

$$LLM_Y^{ideal}(X) = I\{S[DE_Y^{ideal}(X)] > 0.5\} = E[Y|X]$$

and

$$LLM_Y^{real}(X) = I\{S[DE_Y^{real}(X)] > 0.5\}$$

where $DE_Z^{real}(X) = DE_Z^{ideal}(X) + \xi$ and ξ follows symmetry distribution. Assumption 2 is also supported by recent research conducted by Agarwal, Dalal, and Misra (2026), a computer science team partly from Columbia University.

Assumption 3 (Consistency of in-sample and out-of-sample): We assume our learning by LLM is efficient and our learning keep the same signal-to-noise ratio (SNR) and the same correlation coefficient compare with in-sample data, that means:

$$\begin{aligned} \rho &:= Corr[h(S), g(S)] = Corr[E[Y|X], E[Z|X]], \\ \eta_Y &:= \frac{Var(h(S))}{Var(Y)} = \eta_{\hat{Y}} := \frac{Var[E[Y|X]]}{Var(\hat{Y})}, \\ \eta_Z &:= \frac{Var(g(S))}{Var(Z)} = \eta_{\hat{Z}} := \frac{Var[E[Z|X]]}{Var(\hat{Z})}. \end{aligned}$$

7.1.4 Theorems

Theorem 1 (Unbiasedness of LLM Predictions): Under Assumption 2, the predictions of real LLM is unbiased that means $E[LLM_{label}^{real}(X)] = E[LLM_{label}^{ideal}(X)]$ and $E[residual] = 0$

Theorem 2(Signal-to-Noise Ratio Threshold Theorem): Under Assumptions 1 and 3 and Theorem 1, the condition $Corr^2(Y, \hat{Y}) < Corr^2(Y, \hat{Z})$ is equivalent to $\eta_Y/\eta_Z < \rho^2$.

where proof of Theorem 1 and Theorem 2 please refer to mathematical appendix.

Our theoretical model give a brief but powerful result that even though our target is to predict Y in out-of-sample, learning Z is better than learning Y, if the informativeness of *Y* is less than the informativeness of Z too much and the threshold of this depends on the two information parts in Y and Z, $g(S)$ and $h(S)$, are how correlated.

7.2 Can LLMs predict returns by learning the return indicator?

We construct a monthly return indictor of next month return, named ret_flag, indicator equals to one if next month return is negative otherwise equals to zero and then we train BERT to learn the relation between emotional response and this return indictor under the same training design and the same in-sample data and out-of-sample data, compare with learning SEP indicator in the former part. Since there is forward looking bias in learning return indicator[16], the predictions should have stronger predictability compared with learning SEP indicator. Thus, our null hypothesis is that predictions of LLM after learning return indicator can also predict return.

To examinate this null hypothesis, we use these predictions to construct an equal-weighted long-only strategy. We equally buy stocks that would fall in month $t+1$ predicted by BERT learning return indicator and the portfolio return of this long-only strategy is:

$$R_{t+1}^{learn\ ret} = \frac{I(pred_{i,t}^{learn\ ret} = 1)}{\sum_i I(pred_{i,t}^{learn\ ret} = 1)} r_{i,t+1}$$

Furthermore, as comparison, we also construct another equal-weighted long-only strategy in the same stock sample and the same period. We equally buy stocks that would have SEP issuance in month t predicted by BERT learning SEP indicator and the portfolio return of this long-only strategy is:

$$R_{t+1}^{learn\ sep} = \frac{I(pred_{i,t}^{learn\ sep} = 1)}{\sum_i I(pred_{i,t}^{learn\ sep} = 1)} r_{i,t+1}$$

Figure 3 exhibits the cumulative return of these two long-only strategy

[16] We already know whether stock price in month $t+1$ will raise or fall in month t for in-sample stocks.

excluding market return, using S&P 500 monthly return, where blue line is the result of learning SEP indicator and green line is the result of learning return indicator. It is straightforwardly that the result of learning SEP indicator is better than learning return indicator for predicting negative return in next month.

Additionally, we also conduct an alpha test for the examination of predictability on return by using various asset pricing models. Table 10 presents the predictability on return of learning SEP indicator and return indicator. The coefficients of alpha of learning SEP indicator vary from -0.55% (t-statistic -3.58) to -0.32% (t-statistic -2.02) while the coefficients of alpha of learning return indicator vary from -0.34% (t-statistic -3.10) to -0.04% (t-statistic -0.50), which is inconsistent with our null hypothesis that learning return indicator can predict return and implies that learning SEP indicator is better than learning return indicator for predicting return.

7.3 Can LLM predictions predict bubble by learning return indicator?

Moreover, we examinate whether predictions from BERT learning return indicator can predict bubble. We take the same way to test the predictability by conducting panel regressions.

$$Bubble_{i,t} = a + bI(pred_{i,t}^{learn\ ret} = 1) + \sum_{n} \gamma_n Controls_{n,i,t} + \varepsilon$$

where $I(pred_{i,t}^{learn\ ret} = 1)$ is the predictive return indictor for stock i in month t and equals to one if price of this stock would fall in month $t+1$ predicted by BERT learning return indicator. Control variables are the same as the former panel regressions.

Table 11 reports the results of above panel regressions and the coefficients of predictive return indictor are all insignificant which also implies that learning SEP indicator is better than learning return indicator for predicting return and bubble.

7.4 Test for signal-to-noise ratio threshold theorem empirically

We already show that the predictions of learning SEP indicator can predict return and bubble, compare with learning return indictor, which is the conclusion of our theoretical model, but for the key signal-to-noise ratio threshold condition we need more evidence to support.

To show the consistency of theoretical result and empirical result, we conduct a group of panel regressions to fit the transformed fundamental part in SEP indicator and return indicator using various fundamental factors and behavioral factors, like market equity, mispricing, sentiment proxies and so on. The R-square is just the informativeness in dependent variables.

Table 12 displays the results of R^2s of SEP indicator and return indicator. The R^2s of the SEP indicator range from 36.2% to 41.0% while the R-squares of return indicator range from 20.4% to 22.5% on different factors combinations. We take the regressions of the most factors as an example, that is, columns (4)-(8) in Table 12, the informativeness η_Y is 22.5% and η_Z is 41.0%.

We record the coefficients in columns (4)-(8) of Table 12 and use these coefficients to reproduce the transformed fundamental part $g(S)$ and $h(S)$ empirically and then calculate the correlation coefficient of these two parts. By doing so, we got the correlation coefficients is 0.9 and ρ^2 is 0.81, which exactly meet the condition that $\eta_Y/\eta_Z < \rho^2$ and show the consistency of theoretical result and empirical result.

8. Conclusion

This paper uses news data and trains BERT in classification and in regression tasks under unique methodology to extrapolate an existing SEP sentiment measure, proposed by Henderson, Pearson and Wang (2023), to the full cross-section of U.S. stocks. It examines the ability of the extrapolated SEP sentiment measure to predict returns stock returns and stock-level bubbles, as defined by Greenwood, Shleifer, and You (2019).

As original SEP sentiment does, the extrapolated SEP sentiment measure is a strong negative predictor of abnormal returns over the 2004 to 2019 sample period. We show this by using the measure to compute sentiment-weighted portfolio returns, and estimate the abnormal returns using several different linear factor models as benchmarks. These are the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, the Hou, Xue and Zhang (2015) q-factor model, the Hou et al. (2021) $q5$-factor model, the

Stambaugh and Yuan (2017) mispricing factors model, and the Daniel, Hirshleifer and Sun (2020) behavioral factor model.

The estimates of the alphas across all asset pricing models of long-only portfolio weighted by the extrapolative SEP sentiment are between –0.65% and –0.45%, and highly significant. These results reveal that trained BERT can learn the underlying relationship between news titles and SEP sentiment, and moreover can extrapolate this overpricing phenomenon to stocks with no history of SEP issues.

We also examine the relationship between SEP sentiment and a stock-level version of the options-based sentiment *B* proposed by Gao and Martin (2021). We find that higher sentiment *B* is highly significantly associated with higher SEP sentiment in next month and consistent with larger overpricing if sentiment *B* combines with SEP sentiment. However, this association with sentiment *B* and SEP sentiment is not valid in extrapolative SEP sentiment since these extrapolative SEP sentiment's pseudo issuance don't happen in fact.

We then test the predictability of SEP sentiment on stock-level bubble and we find original, extrapolative and full market SEP indicator and SEP sentiment all predict bubble significantly, which strengthens the long-term negative predictability of SEP sentiment on return.

Lastly, we develop a theoretical model to investigate the impact of label noisiness to LLM learning effect and test consistency of theoretical results and empirical results. We document that noise in label would prevent LLM from learning the truly useful information and resulting in the worse learning effect. In our paper, SEP indicator is a less noisy label, compared with return indicator, and thus even though our target is to predict return, learning the SEP indicator is still better than learning return indicator and eventually contributes to the predictability of our extrapolative SEP sentiment on return and bubble, which is also a return-based variable.

References

Arif, Salman, and Edward Sul, 2024, Does accounting information identify bubbles for Fama? Evidence from accruals, *Journal of Accounting and Economics* 78, 101711.

Audrino, F., Sigrist, F., and Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36, No. 2, 334–357

Barbaglia, L., Consoli, S., and Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics* 41, 708–719.

Bybee, J. Leland, 2025. The ghost in the machine: Generating beliefs with large language models. Working Paper, University of Chicago.

Bybee, J. Leland, Brian T. Kelly, Asaf Manela, and Dacheng Xiu, 2024. Business news and business cycles. *Journal of Finance* 79, 3105–3147.

Calvet, Laurent, Claire Célérier, Paolo Sodini, and Boris Vallée, 2023, *Journal of Finance* 78, No. 4, 1917–1966.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57–82.

Célérier, Claire and Boris Vallée, 2017, Catering to Investors Through Security Design: Headline Rate and Complexity, *Quarterly Journal of Economics* 132, No. 3, 1469–1508.

Chen, Jian, Guohao Tang, Guofu Zhou, and Wu Zhu, 2025. ChatGPT and Deepseek: Can They Predict the Stock Market and Macroeconomy?. Working Paper, Washington University.

Chen, Yifei, Bryan T. Kelly, and Dachen Xiu, 2024. Expected Returns and Large Language Models. Working Paper, University of Chicago.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short- and Long-Horizon Behavioral Factors, *Review of Financial Studies* 33, 1673–1736.

Fama, Eugene F., 2014, Two Pillars of Asset Pricing, *American Economic Review* 104, 1467–1485.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Gao, Can, and Ian W. R. Martin, 2021, Volatility, Valuation Ratios, and Bubbles: An Empirical Measure of Market Sentiment, *Journal of Finance* 76, 3211–3254.

Greenwood, Robin, Andrei Shleifer, and Yang You, 2019, Bubbles for Fama, *Journal of Financial Economics* 131, 20–43.

Henderson, Brian J., and Neil D. Pearson, 2011, The dark side of financial innovation:

A case study of the pricing of a retail financial product, *Journal of Financial Economics* 100, 227–247.

Henderson, Brian J., Neil D. Pearson, and Li Wang, 2023, Retail Derivatives and Sentiment: A Sentiment Measure Constructed from Issuances of Retail Structured Equity Products, *Journal of Finance* 78, 2365–2407.

Hirshleifer, D., Mai, D., and Pukthuanthong, K. (2025). War discourse and disaster premium:160 years of evidence from the stock market. *Review of Financial Studies* 38, 457–506.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An Augmented q -Factor Model with Expected Growth*, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting Anomalies: An Investment Approach, *Review of Financial Studies* 28, 650–705.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is There a Replication Crisis in Finance?, *Journal of Finance* 78, 2465–2518.

Jeon, Y., McCurdy, T. H., and Zhao, X. (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics* 145, 1–17.

Martin, Ian, 2017. What is the Expected Return on the Market? *Quarterly Journal of Economics* 132, No. 1, 367–433.

Richardson, Scott A., Richard G. Sloan, Mark T. Soliman, and İrem Tuna, 2005, Accrual reliability, earnings persistence and stock prices, *Journal of Accounting and Economics* 39, 437–485.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2015, Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle, *Journal of Finance* 70, 1903–1948.

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing Factors, *Review of Financial Studies* 30, 1270–1315.

Tetlock, Paul C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, No.3, 1139–1168.

Vokatá, Petra, 2021, Engineering lemons, *Journal of Financial Economics* 142, 737–755

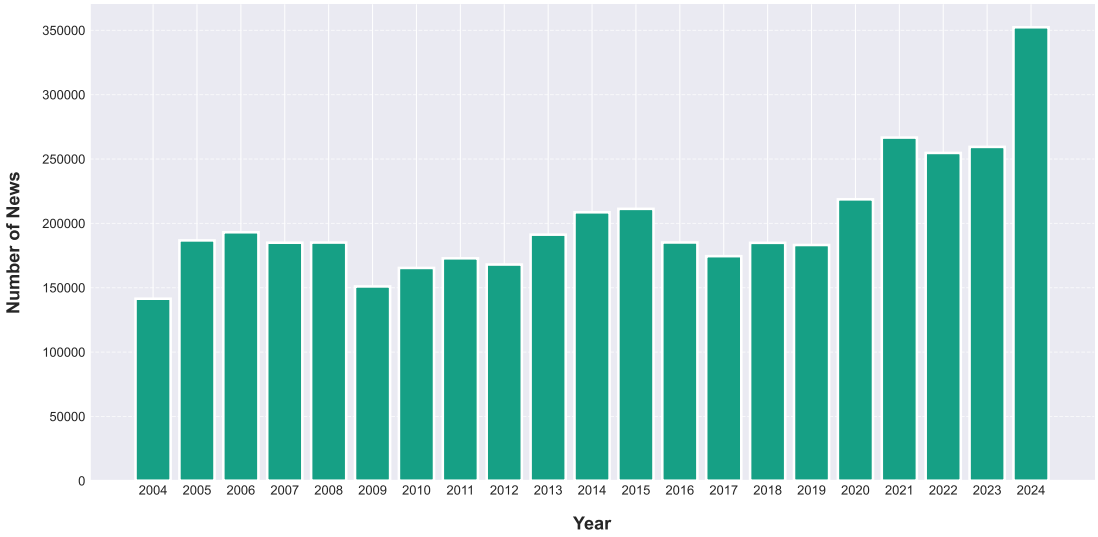


Figure 1

Yearly U.S. Stock Third Party News Amount from 2004 to 2024.

The figure displays the number of pieces of news in our news dataset from 2004 to 2024 which is counted in year level. The number of pieces of news varies from nearly 150k per year to 350k per year. The news amount fluctuates from 150k to 200k from 2004 to 2019.

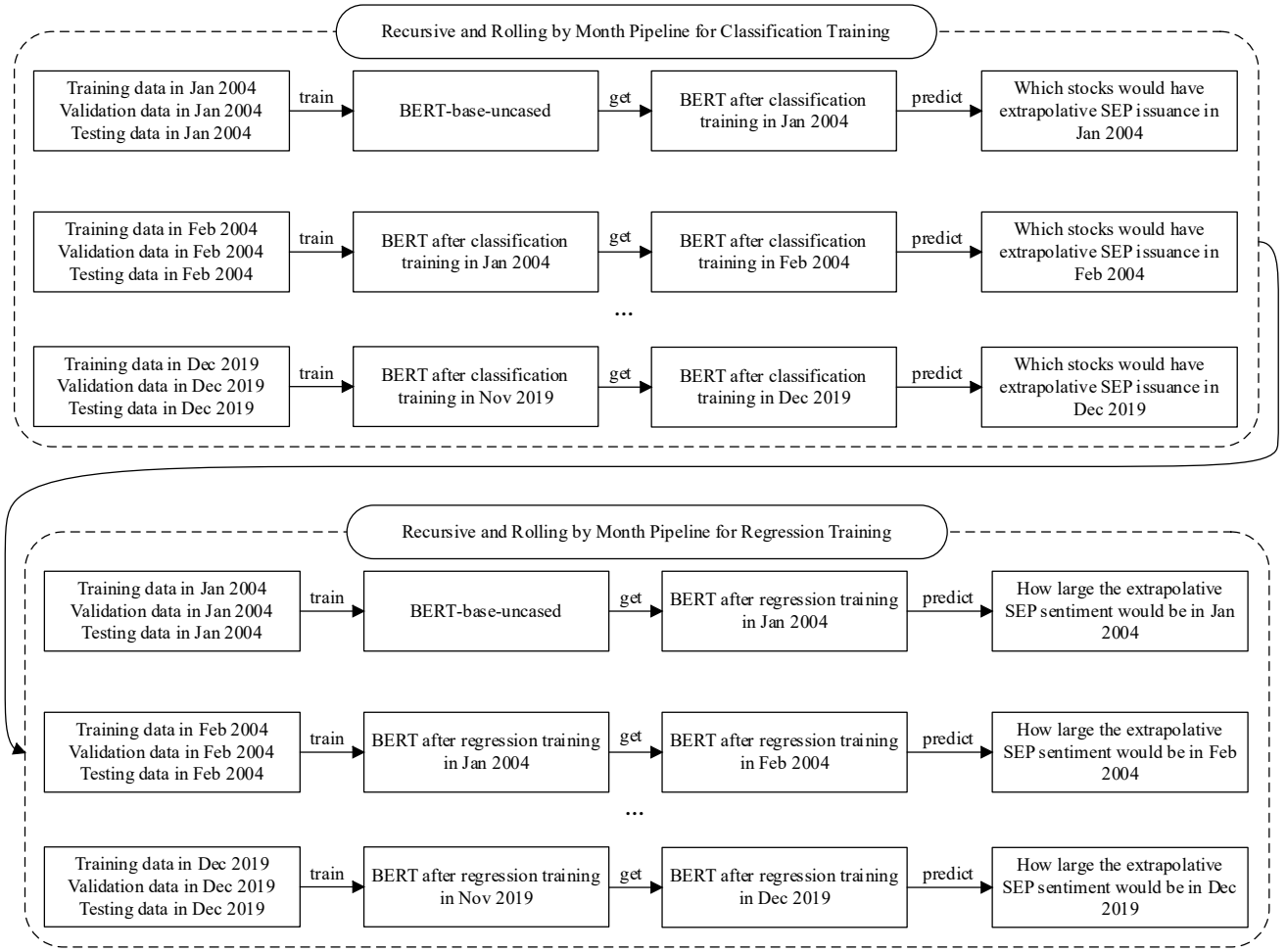


Figure 2. Recursive Training Rolling by Month

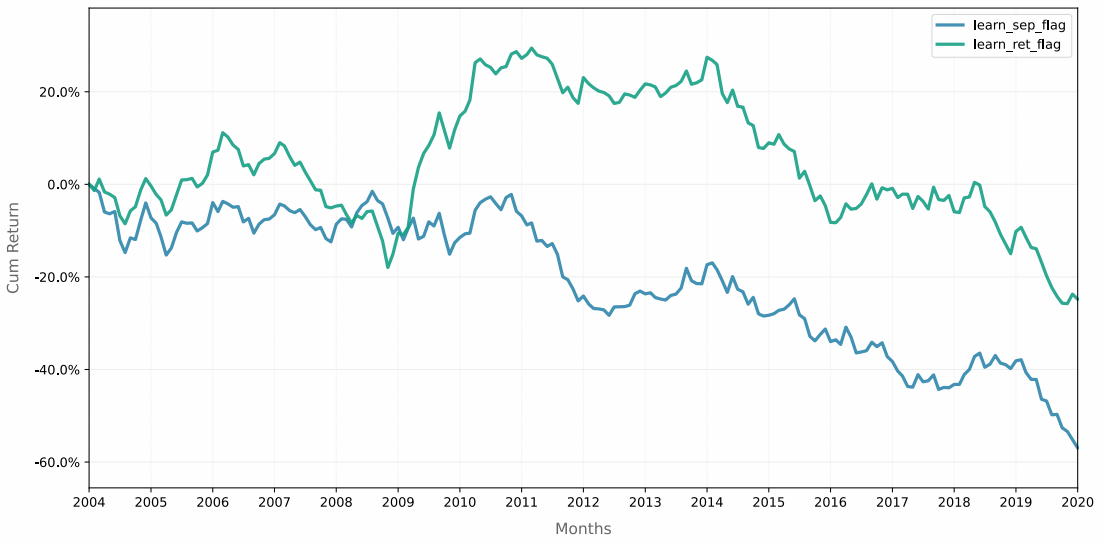


Figure 3

Cumulative excess return of two equal-weighted portfolio strategies

The figure displays the cumulated return of equal-weighted portfolio return excluding market return, using S&P 500 monthly return, from Feb 2004 to January 2020 where portfolio consists of stocks from LLM positive predictions of learning different labels. The blue line is the result of learning SEP indicator and the green line is the result of learning return indicator.

Table 1

Predictability of Extrapolative SEP Sentiment on Return

This table presents results of regressing monthly excess returns of portfolios formed using the extrapolative SEP sentiment $G_{i,t}^{Ex}$ in the full sample on the factor returns of various asset pricing models, including the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, the Hou, Xue and Zhang (2015) q-factor model, the Hou et al. (2021) $q5$-factor model, the Stambaugh and Yuan (2017) mispricing factors model and the Daniel, Hirshleifer and Sun (2020) behavioral factors model. From left to right, each columns shows the asset pricing model name, coefficient of alpha, t-statistics, corresponding p-value of t-statistics, the R^2 of this regression and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 to 2019.

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00648	-4.04519	7.63E-05	0.80072	192
Carhart 4	-0.00642	-3.99425	9.32E-05	0.801086	192
FF5	-0.00512	-3.16538	0.00181	0.811945	192
q	-0.00535	-3.28796	0.001206	0.804964	192
q5	-0.00445	-2.6546	0.008628	0.809035	192
mispricing	-0.00493	-2.83956	0.00514	0.834123	156
DHS	-0.00489	-2.74794	0.006581	0.768591	192

Table 2
Predictability of Extrapolative SEP Sentiment excluding last 5% capitalization stocks

This table presents results of regressing monthly excess returns of portfolios formed using the extrapolative SEP sentiment after excluding small size stocks whose capitalization is the last 5% in the entire market in month $t-1$ on the factor returns of various asset pricing models, including the Fama and French (1993) three factors model, the Carhart (1997) four factors model, the Fama and French (2015) five factors model, the Hou, Xue and Zhang (2015) q factors model, the Hou et al. (2021) q5-factor model, the Stambaugh and Yuan (2017) mispricing factors model and the Daniel, Hirshleifer and Sun (2020) behavioral factors model. From left to right, each columns shows the asset pricing model name, coefficient of alpha, t-statistics, corresponding p-value of t-statistics, the R-square of this regression and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 to 2019.

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00661	-4.16451	4.75E-05	0.803223	192
Carhart 4	-0.00655	-4.11071	5.9E-05	0.803649	192
FF5	-0.0052	-3.25005	0.00137	0.81525	192
q	-0.00544	-3.38155	0.000878	0.808394	192
q5	-0.00429	-2.60781	0.009852	0.81506	192
mispricing	-0.00534	-3.17309	0.001828	0.841778	156
DHS	-0.00494	-2.79947	0.005653	0.771426	192

Table 3

Predictability of the full SEP sentiment measure

This table presents results of regressing monthly excess returns of portfolios formed using the full market SEP sentiment $G_{i,t}^{Full}$ in the full sample and excluding small size stocks on the factor returns of various asset pricing models, including the Fama and French (1993) three factors model, the Carhart (1997) four factors model, the Fama and French (2015) five-factor model, the Hou, Xue and Zhang (2015) q-factor model, the Hou et al. (2021) $q5$-factor model, the Stambaugh and Yuan (2017) mispricing factor model and the Daniel, Hirshleifer and Sun (2020) behavioral factors model. From left to right, each column shows the asset pricing model name, coefficient of alpha, *t*-statistics, corresponding p-value of t-statistics, the R^2 of this regression, and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 through 2019.

Panel A: Full Sample

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.0095	-5.44071	1.64E-07	0.825351	192
Carhart 4	-0.0094	-5.37312	2.28E-07	0.826056	192
FF5	-0.00824	-4.64836	6.32E-06	0.833343	192
q	-0.0084	-4.76971	3.7E-06	0.831261	192
q5	-0.00682	-3.82683	0.000177	0.840505	192
mispricing	-0.00821	-4.32219	2.79E-05	0.859226	156
DHS	-0.00714	-4.1051	6.02E-05	0.836781	192

Panel B: Excluding last 5% capitalization stocks

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00963	-5.46133	1.48E-07	0.82278	192
Carhart 4	-0.00954	-5.39353	2.07E-07	0.823503	192
FF5	-0.00836	-4.6635	5.92E-06	0.830874	192
q	-0.00852	-4.79024	3.38E-06	0.828994	192
q5	-0.00683	-3.80794	0.00019	0.839512	192
mispricing	-0.00846	-4.39587	2.07E-05	0.855987	156
DHS	-0.00725	-4.12652	5.53E-05	0.834778	192

Table 4

Predictability of Listed Option-based Sentiment *B* on Return in Horizons

This table presents the predictability of the option-based sentiment $B_{i,t}$ on return in different horizons where since $B_{i,t}$ containing both negative and positive, we conduct the long-short strategy to construct portfolio using sentiment $B_{i,t}$. Then we also regress the monthly excess returns of portfolios formed using the option-based sentiment $B_{i,t}$ on the factor returns of various asset pricing models, including the Fama and French (1993) three factors model, the Carhart (1997) four factors model, the Fama and French (2015) five factors model, the Hou, Xue and Zhang (2015) q factors model, the Hou et al. (2021) q5-factor model, the Stambaugh and Yuan (2017) mispricing factors model and the Daniel, Hirshleifer and Sun (2020) behavioral factors model. From left to right, each columns shows the asset pricing model name, coefficient of alpha, t-statistics, corresponding p-value of t-statistics, the R-square of this regression and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 to 2019.

Panel A: Next 1 month

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00303	-0.76731	0.443862	0.328449	192
Carhart 4	-0.00244	-0.62618	0.531959	0.3489	192
FF5	0.001272	0.325042	0.745514	0.389028	192
q	0.000512	0.131937	0.895176	0.382906	192
q5	0.001407	0.348829	0.727613	0.385147	192
mispricing	5.49E-05	0.01209	0.99037	0.374763	156
DHS	0.001643	0.396122	0.692464	0.302594	192

Panel B: Next 2 month

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00619	-1.23214	0.219435	0.151581	192
Carhart 4	-0.00518	-1.05761	0.291596	0.19861	192
FF5	-0.00137	-0.27463	0.783902	0.221095	192
q	3.19E-05	0.00678	0.994597	0.290167	192
q5	0.002087	0.428138	0.669046	0.299383	192
mispricing	0.000891	0.158082	0.874603	0.271695	156
DHS	7.22E-05	0.014794	0.988212	0.246896	192

Panel C: Next 3 month

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00647	-1.43229	0.153721	0.18543	192
Carhart 4	-0.00511	-1.20418	0.230042	0.288841	192
FF5	-0.00239	-0.52479	0.600351	0.236665	192
q	4.94E-05	0.011982	0.990452	0.356753	192
q5	-0.0006	-0.13953	0.889178	0.357838	192
mispricing	0.000229	0.047239	0.962385	0.333217	156
DHS	-0.00074	-0.16689	0.867639	0.268835	192

Table 5
Association between HPW sentiment and lagged options-based sentiment *B*
This table presents results of panel regressions of positive original SEP sentiment $G_{i,t}$ and positive extrapolative SEP sentiment $G_{i,t}^{Ex}$ on listed option-based sentiment $B_{i,t}$ and various control variables where columns (1) to (4) are results based on 16,146 $G_{i,t}$ and (5) to (8) are results based on 50,808 $G_{i,t}^{Ex}$. Compared with the number of observations shown in this table, we can find that the overlapping of $G_{i,t}$ and $B_{i,t}$ over 90% while the overlapping of $G_{i,t}$ and $B_{i,t}$ is less, around 40%.

VARIABLES	(1) $G_{i,t}$	(2) $G_{i,t}$	(3) $G_{i,t}$	(4) $G_{i,t}$	(5) $G_{i,t}^{Ex}$	(6) $G_{i,t}^{Ex}$	(7) $G_{i,t}^{Ex}$	(8) $G_{i,t}^{Ex}$
b_l1	0.231***	0.166***	0.117***	0.118***	-0.001	-0.002	-0.001	-0.001
	(5.24)	(4.37)	(3.23)	(3.26)	(-0.66)	(-0.72)	(-0.59)	(-0.63)
b_l2		0.153***	0.117***	0.118***		-0.001	-0.002	-0.002
		(3.93)	(3.27)	(3.33)		(-0.26)	(-0.74)	(-0.67)
me_l1			-0.004***	-0.004***			-0.000	0.000**
			(-3.38)	(-3.35)			(-1.54)	(-2.00)
b2m_l1			-0.001	-0.001			-0.000*	-0.000*
			(-0.76)	(-0.74)			(-1.68)	(-1.95)
mom_l1			-0.001	-0.001			0.000	0.000
			(-1.58)	(-1.41)			(0.77)	(0.98)
roe_l1				0.001				0.000
				(1.14)				(1.25)
agr_l1				0.001				0.000
				(0.87)				(1.54)
Constant	0.008***	0.007***	0.045***	0.047***	0.004***	0.004***	0.006***	0.006***
	(22.03)	(14.77)	(4.04)	(3.97)	(112.61)	(91.14)	(5.09)	(5.53)
Stock and mnth fxd. effs,	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,902	14,893	14,468	14,442	21,493	21,221	20,201	20,163
R-squared	0.353	0.354	0.361	0.361	0.859	0.859	0.861	0.860

Table 6

Predictability of Part Original SEP Sentiment associated Sentiment *B* is high

This table presents the predictability of the original SEP sentiment $G_{i,t}$ after selection by the listed option-based sentiment $B_{i,t-1}$ on return where we still construct portfolio by long-only strategy using $G_{i,t}$. Then we also regress the monthly excess returns of portfolios formed using the option-based sentiment $B_{i,t}$ on the factor returns of various asset pricing models, including the Fama and French (1993) three factors model, the Carhart (1997) four factors model, the Fama and French (2015) five factors model, the Hou, Xue and Zhang (2015) q factors model, the Hou et al. (2021) q5-factor model, the Stambaugh and Yuan (2017) mispricing factors model and the Daniel, Hirshleifer and Sun (2020) behavioral factor model. From left to right, each column shows the asset pricing model name, coefficient of alpha, t-statistics, corresponding p-value of t-statistics, the R-square of this regression and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 to 2019.

Panel A: Option-based Sentiment in top 66% in 1 month before SEP issuance

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.01364	-4.27053	3.09E-05	0.666871	192
Carhart 4	-0.01351	-4.21475	3.89E-05	0.667672	192
FF5	-0.01289	-3.90848	0.00013	0.671669	192
q	-0.01274	-3.90477	0.000132	0.670131	192
q5	-0.01059	-3.16509	0.001812	0.679891	192
mispricing	-0.01334	-3.75526	0.000247	0.722004	156
DHS	-0.01122	-3.45345	0.000684	0.675943	192

Panel B: Option-based Sentiment in top 50% in 1 month before SEP issuance

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.01517	-4.19806	4.15E-05	0.620179	192
Carhart 4	-0.01494	-4.12803	5.51E-05	0.622234	192
FF5	-0.01428	-3.83679	0.000171	0.627589	192
q	-0.01392	-3.76509	0.000223	0.622465	192
q5	-0.01136	-3.00118	0.003058	0.634794	192
mispricing	-0.01495	-3.67951	0.000325	0.676317	156
DHS	-0.01246	-3.37712	0.00089	0.627442	192

Panel C: Option-based Sentiment in top 33% in 1 month before SEP issuance

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.01764	-4.0622	7.13E-05	0.550266	192
Carhart 4	-0.01742	-4.00045	9.1E-05	0.551868	192
FF5	-0.01699	-3.78553	0.000207	0.555875	192
q	-0.01632	-3.6639	0.000323	0.550455	192
q5	-0.01463	-3.16872	0.00179	0.554909	192
mispricing	-0.01768	-3.54315	0.000526	0.605183	156
DHS	-0.01463	-3.28528	0.001215	0.555521	192

Panel D: Option-based Sentiment in top 25% in 1 month before SEP issuance

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.01893	-3.98258	9.74E-05	0.514886	192
Carhart 4	-0.0186	-3.90901	0.000129	0.518055	192
FF5	-0.01833	-3.71879	0.000265	0.517625	192
q	-0.01727	-3.55767	0.000474	0.519259	192
q5	-0.01586	-3.14634	0.001926	0.522041	192
mispricing	-0.01887	-3.53144	0.000548	0.583467	156
DHS	-0.01605	-3.28596	0.001213	0.518172	192

Table 7

Predictability of Original SEP Sentiment on Stock-level Bubble

This table presents results of panel regressions of bubble indicator on original SEP sentiment *G* and original SEP indicator G_flag and various control variables where columns (1) to (4) are results of *G* and (5) to (8) are results of G_flag.

VARIABLES	(1) bubble_flag	(2) bubble_flag	(3) bubble_flag	(4) bubble_flag	(5) bubble_flag	(6) bubble_flag	(7) bubble_flag	(8) bubble_flag
g_flag	0.046***	0.032***	0.030***	0.030***				
	(9.52)	(7.51)	(7.23)	(7.23)				
size		0.035***	0.035***	0.035***		0.037***	0.037***	0.037***
		(6.05)	(5.89)	(5.89)		(6.34)	(6.16)	(6.16)
b2m		-0.005	-0.004	-0.004		-0.005	-0.004	-0.004
		(-0.77)	(-0.65)	(-0.62)		(-0.70)	(-0.58)	(-0.55)
mom		0.182***	0.182***	0.181***		0.182***	0.182***	0.182***
		(26.90)	(26.62)	(26.61)		(26.83)	(26.55)	(26.54)
roe			-0.025***	-0.025***			-0.025***	-0.025***
			(-3.32)	(-3.34)			(-3.35)	(-3.36)
agr			0.044***	0.044***			0.045***	0.045***
			(5.51)	(5.51)			(5.59)	(5.59)
d_noa				0.040				0.041
				(1.42)				(1.47)
g					1.460***	0.979***	0.888***	0.887***
					(4.72)	(3.43)	(3.13)	(3.13)
Stock & month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	112,898	112,898	112,898	112,898	112,898	112,898	112,898	112,898
R-squared	0.179	0.264	0.266	0.266	0.176	0.263	0.265	0.265

.

Table 8

Predictability of Extrapolative SEP Sentiment on Stock-level Bubble

This table presents results of panel regressions of bubble indicator on extrapolative SEP sentiment G and extrapolative SEP indicator G_flag and various control variables where columns (1) to (4) are results of G and (5) to (8) are results of G_flag.

VARIABLES	(1) bubble_flag	(2) bubble_flag	(3) bubble_flag	(4) bubble_flag	(5) bubble_flag	(6) bubble_flag	(7) bubble_flag	(8) bubble_flag
g_flag	0.008***	0.006***	0.006***	0.006***				
	(3.77)	(3.17)	(3.17)	(3.17)				
size		0.086***	0.086***	0.086***		0.070***	0.070***	0.070***
		(18.37)	(18.37)	(18.38)		(24.53)	(24.34)	(24.33)
b2m		0.004***	0.004***	0.004***		0.010***	0.013***	0.013***
		(5.99)	(5.99)	(5.99)		(3.60)	(4.48)	(4.52)
mom		0.094***	0.094***	0.094***		0.180***	0.181***	0.181***
		(5.94)	(5.94)	(5.94)		(66.91)	(67.00)	(67.00)
roe			0.000	0.000			-0.017***	-0.017***
			(1.00)	(1.00)			(-8.16)	(-8.16)
agr			-0.000***	-0.000***			0.026***	0.026***
			(-3.05)	(-3.05)			(10.22)	(10.21)
d_noa				0.000				0.031**
				(0.01)				(2.30)
g					1.969***	1.514***	1.498***	1.500***
					(2.96)	(2.59)	(2.58)	(2.58)
Stock & month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	446,812	446,812	446,812	446,812	446,812	446,812	446,812	446,812
R-squared	0.185	0.278	0.278	0.278	0.185	0.305	0.307	0.307

Table 9
Predictability of Mix SEP Sentiment on Stock-level Bubble
This table presents results of panel regressions of bubble indicator on full market SEP sentiment G and full market SEP indicator G_flag and various control variables where columns (1) to (4) are results of G and (5) to (8) are results of G_flag.

VARIABLES	(1) bubble_flag	(2) bubble_flag	(3) bubble_flag	(4) bubble_flag	(5) bubble_flag	(6) bubble_flag	(7) bubble_flag	(8) bubble_flag
g_flag	0.019***	0.012***	0.012***	0.012***				
	(9.23)	(6.46)	(6.45)	(6.44)				
size		0.062***	0.063***	0.063***		0.062***	0.063***	0.063***
		(24.49)	(24.32)	(24.32)		(24.50)	(24.34)	(24.33)
b2m		0.009***	0.011***	0.011***		0.009***	0.011***	0.011***
		(3.38)	(4.36)	(4.41)		(3.39)	(4.36)	(4.42)
mom		0.180***	0.180***	0.180***		0.180***	0.180***	0.180***
		(72.39)	(72.37)	(72.36)		(72.40)	(72.38)	(72.37)
roe			-0.019***	-0.019***			-0.019***	-0.019***
			(-9.22)	(-9.23)			(-9.21)	(-9.22)
agr			0.029***	0.029***			0.029***	0.029***
			(11.76)	(11.75)			(11.76)	(11.75)
d_noa				0.033***				0.033***
				(2.70)				(2.70)
g					5.602***	3.716***	3.601***	3.599***
					(6.84)	(4.90)	(4.78)	(4.78)
Stock & month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	570,578	570,578	570,578	570,578	570,578	570,578	570,578	570,578
R-squared	0.173	0.288	0.290	0.290	0.173	0.288	0.290	0.290

Table 10

Predictability on return of learning SEP indicator and return indicator

This table presents the predictability of the LLM positive predictions after learning SEP indicator and return indicator. Then we also regress the monthly excess returns of equal-weighted portfolios formed using the LLM positive predictions on the factor returns of various asset pricing models, including the Fama and French (1993) three factors model, the Carhart (1997) four factors model, the Fama and French (2015) five factors model, the Hou, Xue and Zhang (2015) q factors model, the Hou et al. (2021) q5-factor model, the Stambaugh and Yuan (2017) mispricing factors model and the Daniel, Hirshleifer and Sun (2020) behavioral factors model. From left to right, each columns shows the asset pricing model name, coefficient of alpha, t-statistics, corresponding p-value of t-statistics, the R^2 of this regression and lastly the number of observations, also the number of months in time-series, respectively. The sample period runs from 2004 to 2019.

Panel A: learning SEP indicator

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00551	-3.58038	0.000437	0.817729	192
Carhart 4	-0.00548	-3.54633	0.000494	0.817818	192
FF5	-0.00422	-2.7208	0.00713	0.828469	192
q	-0.00422	-2.71679	0.007211	0.823265	192
q5	-0.00323	-2.02584	0.04421	0.828158	192
mispricing	-0.00404	-2.41809	0.016793	0.843772	156
DHS	-0.00389	-2.23018	0.026918	0.779806	192

Panel B: learning return indicator

Model	**Coeff.**	***t*-statistic**	***p*-value**	R^2	**No. obs.**
FF3	-0.00343	-3.10568	0.002192	0.912282	192
Carhart 4	-0.00302	-3.03177	0.002776	0.929531	192
FF5	-0.00232	-2.10052	0.037031	0.919039	192
q	-0.00149	-1.55204	0.122344	0.937078	192
q5	-0.00048	-0.50137	0.6167	0.941814	192
mispricing	-0.00063	-0.57213	0.568086	0.937538	156
DHS	-0.00142	-1.0348	0.302091	0.873091	192

Table 11

Predictability on stock-level bubble of learning ret_flag

This table presents results of panel regressions of bubble indicator on LLM positive predictions of learning return indicator and various control variables.

VARIABLES	(1) bubble_flag	(2) bubble_flag	(3) bubble_flag	(4) bubble_flag
pred_ret_flag	0.000	0.000	0.000	0.000
	(0.06)	(0.57)	(0.57)	(0.57)
size		0.086***	0.086***	0.086***
		(18.37)	(18.37)	(18.38)
b2m		0.004***	0.004***	0.004***
		(6.00)	(5.99)	(6.00)
mom		0.094***	0.094***	0.094***
		(5.94)	(5.94)	(5.94)
roe			0.000	0.000
			(1.00)	(1.00)
agr			-0.000***	-0.000***
			(-3.17)	(-3.17)
d_noa				0.000
				(0.01)
Stock and month fixed effects	Yes	Yes	Yes	Yes
Observations	446,812	446,812	446,812	446,812
R-squared	0.185	0.278	0.278	0.278

Table 12

Noise in SEP-based and Return-based Labels

This table presents results of panel regressions of original SEP indicator and return indicator of training dataset on various fundamental and behavioral factors to fit the true signal part in SEP indicator and return indicator, respectively. Columns (1) to (4) are results of SEP indicator and (5) to (8) are results of return indicator. Theoretical model show that if $\eta_Y/\eta_Z < \rho^2$, then learn Z is better to predict Y. Specifically, in our study, take columns (4) and (8) for example, we construct linear combination of these factors under the coefficients shown in columns (4) and (8) and then calculate the correlation coefficient of these two series. By doing so, we get the correlation coefficient ρ is 0.9. Thus, the threshold condition holds, since $0.225/0.410 < 0.9^2 = 0.81$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	sep_flag	sep_flag	sep_flag	sep_flag	ret_flag	ret_flag	ret_flag	ret_flag
size	0.074***	0.073***	0.075***	0.105***	0.040***	0.040***	0.041***	0.101***
	(8.91)	(8.65)	(8.83)	(7.04)	(7.45)	(7.53)	(7.61)	(12.00)
b2m	0.003	0.003	0.002	-0.014	0.001	0.001	0.001	0.011
	(1.46)	(1.47)	(1.17)	(-1.43)	(0.43)	(0.43)	(0.33)	(1.59)
mom	-0.014**	-0.014**	-0.012**	0.010	-0.011**	-0.011**	-0.010*	-0.020**
	(-2.44)	(-2.40)	(-2.14)	(1.58)	(-2.05)	(-2.06)	(-1.94)	(-2.29)
roe		-0.000	-0.000	0.000*		0.000	0.000	-0.000
		(-1.36)	(-1.42)	(1.71)		(0.16)	(0.15)	(-0.76)
agr		0.010	0.009	0.008		-0.001	-0.001	-0.004
		(1.49)	(1.47)	(1.36)		(-0.47)	(-0.50)	(-0.89)
ret			-0.076***	-0.027			-0.018	-0.035
			(-6.91)	(-1.40)			(-1.13)	(-1.31)
ret_l1			-0.056***	-0.047***			-0.014	0.000
			(-5.66)	(-3.21)			(-0.85)	(0.00)
misp				0.146***				0.021
				(3.59)				(0.74)
ivol_ff3				-0.496*				0.233
				(-1.83)				(0.78)
hw52				-0.156***				-0.028

				(-5.14)				(-1.02)
vol_ratio				6.432***				2.547*
				(3.97)				(1.71)
si_ratio				-0.021***				0.012***
				(-5.39)				(3.85)
Stock and month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	68,321	68,239	68,239	43,670	68,314	68,232	68,232	43,668
R-squared	0.362	0.362	0.363	0.410	0.204	0.204	0.204	0.225