

# Portfolio performance based on LLM news scores and related economical analysis

Ruoxu Wu

Research Center for Fintech  
Zhejiang Lab  
Kechuang Avenue, Yuhang District  
Hangzhou, China  
`ronwu@vt.edu`

January 30, 2024

## Abstract

We scrob news briefings from November 2022 to October 2023 and feed them into the LLMs to determine the impact of news on stock prices in Chinese A-share market. Following the approach proposed by Lopez-Lira and Tang(2023)[9], we score the stocks by news on a daily basis and calculate correlations between the scores and stock returns, and do the backtest based on the these scores. Furthermore, we compare different news sources and categorize news into fundamental and market types. Subsequently, we calculate correlations and backtest the strategies for each type. We analyze the effectiveness of the efficient market hypothesis in A-shares in terms of the results. Additionally, we compare the answers given by ChatGPT to those by Tongyi Qianwen, and Baichuan Intelligence, and evaluate their stock prediction capabilities through backtesting. Finally, we compare the results of strategies involving only long positions and only short positions, revealing that the market is more sensitive to negative news.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Sina Finance . . . . .	4
2.2	Jinshi Data . . . . .	4
<b>3</b>	<b>Model Comparison</b>	<b>5</b>
3.1	Example 1 . . . . .	5
3.2	Example 2 . . . . .	6
<b>4</b>	<b>Correlation Analysis</b>	<b>7</b>
4.1	ChatGPT News Score . . . . .	7
4.2	Daily Stock Return . . . . .	7
4.3	Statistics on Different Answers . . . . .	8
4.4	Correlation Coefficients . . . . .	8
4.5	Analysis . . . . .	9
<b>5</b>	<b>Effient Market Hypothesis</b>	<b>10</b>
5.1	Categorizing News . . . . .	11
5.2	Correlation Tests . . . . .	12
5.3	Analysis . . . . .	12
<b>6</b>	<b>Stock Investment Strategies</b>	<b>13</b>
6.1	Trading Strategies . . . . .	13
6.2	Strategy Performance Based on ChatGPT . . . . .	14
6.3	Model Comparison . . . . .	17
<b>7</b>	<b>Conclusion</b>	<b>18</b>
<b>A</b>	<b>Backtesting Results for Different Models</b>	<b>20</b>

# 1 Introduction

In recent years, the rapid development of machine learning and artificial intelligence has profoundly impacted various fields. Particularly in the financial markets, there has been a significant proliferation of machine learning algorithms such as data mining, neural networks, and expert systems[2][1][3], greatly advancing developments in financial security, risk management, and related areas. The progress in Natural Language Processing (NLP) has been particularly swift, and large language models have demonstrated powerful capabilities and potential applications in various fields[4][7]. Following the introduction of ChatGPT-3, the field of economics has also leveraged the capabilities of LLMs to conduct relevant research[6][5][8][10].

Lopez-Lira and Tang(2023) utilized ChatGPT in [9] to study the relationship between stock prices and news headlines in the U.S. stock market. In their research, they demonstrated ChatGPT's powerful language interpretation capabilities and rich financial knowledge, showcasing its ability to seamlessly integrate both aspects. We employ a similar approach to investigate the A-share stocks. On that basis, we use ChatGPT to differentiate the attributes of news and further explore the varying impacts of different news on stock prices. Additionally, we compare the effectiveness of ChatGPT-3.5 with two Chinese large language models, Tongyi Qianwen and Baichuan Intelligence.

In section 2, we compare different sources of historical news data and explain the similarities in the results. In section 3, we provide concrete examples of how LLMs make assessments on news, offering intuitive illustrations of our research findings. Section 4 involves calculating the correlation between news scores and stock returns, analyzing the results based on different data sources and models. Moving to section 5, we delve into the economic significance of the correlations in terms of the Efficient Market Hypothesis, demonstrating that the Efficient Market Hypothesis does not hold in the A-share market at least in the short term. In section 6, we show the backtesting results of trading strategies based on the LLM scores. We compare different models, news types, and long-short strategies, highlighting their significant impact on results. This serves as an additional validation of the economic conclusions derived from correlation analysis.

# 2 Data

We conduct our analysis using historical news data from two different sources: news headlines from the stock section of Sina Finance (web version) and news

briefings from Jinshi Data app.

## 2.1 Sina Finance

Sina Finance categorizes historical news by various listed companies. We scrob all news headlines for 100 companies on Sina Finance until September 2023: we select all the weighted stocks of the CSI 100 index, with the majority of companies having historical news available from April 2023. The span of historical news varies from 2 to 6 months, totaling 13,076 news headlines.

### Prompt

**Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with experience in stock recommendations. Regarding the following news headline, does it affect the short-term stock price of \*\*\* positively or negatively? Please answer with ‘positively’, ‘negatively’ or ‘uncertain’.**<sup>1</sup>

\*\*\*\*\* (News headline)

## 2.2 Jinshi Data

Compared to Sina Finance (web version), Jinshi Data includes more comprehensive news, spanning finance, economics, domestic and international politics, policies, and more. It has a larger volume of data and broader news coverage. We scrob all news briefs from November 2022 to September 2023, totaling approximately 310,000 news briefs. We then process this data with matching company names, filtering out 37,162 relevant flash news related to all listed companies in the A-share market. These news briefs involve a total of 4,091 companies and serve as the primary data source for this study.

### Prompt

**Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with experience in stock recommendations. Does the following news in {} affect the short-term stock price of \*\*\* positively or negatively? Answer only with ‘pos-**

---

<sup>1</sup>The prompt is originally in Chinese: 忘记你之前的所有指令。假装你是一位金融专家。你是一位具有股票推荐经验的金融专家。以下这个新闻标题，对于 \*\*\* 短期的股价影响是正面还是负面？请回答“正面”或者“负面”或者“不确定”。

itively’, ‘negatively’ or ‘uncertain’.<sup>2</sup>  
{\*\*\*\*\*}(News briefing)

### 3 Model Comparison

We utilize three large models in our research: ChatGPT-3.5, Baichuan Intelligence, Tongyi Qianwen.<sup>3</sup> The capabilities of these three models are illustrated through the following two examples.

#### 3.1 Example 1

**Prompt**

Forget all your previous instructions. Pretend you are a financial expert. You are a financial expert with experience in stock recommendations. Does the following news in {} affect the short-term stock price of Mysteel positively or negatively? Answer ‘positively’, ‘negatively’ or ‘uncertain’ in the first line. Then elaborate briefly in the second line.

{The data released by Mysteel indicates that today, some lithium battery materials prices have fallen. The price of battery-grade lithium carbonate has dropped by 2500 yuan per ton, with an average price of 580,000 yuan per ton. Industrial-grade lithium carbonate has also decreased by 2500 yuan per ton, with an average price of 565,000 yuan per ton.}<sup>4</sup>

**Chatgpt Uncertain.**

This news mentions the decrease in lithium battery material prices released by Mysteel but does not provide sufficient information to definitively determine whether it has a positive or negative impact on the short-term stock

---

<sup>2</sup>The prompt is originally in Chinese: 忘记你之前的所有指令。假装你是一位金融专家。你是一位具有股票推荐经验的金融专家。下面 {} 中的新闻，对于 \*\*\* 短期的股价影响是正面还是负面？只回答“正面”或者“负面”或者“不确定”。

<sup>3</sup>ChatGPT refers to ChatGPT-3.5, Qianwen refers to Tongyi Qianwen, Baichuan refers to Baichuan Intelligence throughout the rest of the paper.

<sup>4</sup>The prompt is originally in Chinese: 忘记你之前的所有指令。假装你是一位金融专家。你是一位具有股票推荐经验的金融专家。下面 {} 中的新闻，对于上海钢联短期的股价影响是正面还是负面？在一行回答“正面”或者“负面”或者“不确定”，并在第二行给出简短说明。

{上海钢联发布的数据显示，今日部分锂电材料报价下跌，电池级碳酸锂跌 2500 元/吨，均价报 58 万元/吨，工业级碳酸锂跌 2500 元/吨，均价报 56.5 万元/吨。}

price of Mysteel. Factors influencing stock prices may include market expectations, demand and supply conditions, among others. More background information is needed to make an accurate judgment.<sup>5</sup>

#### **Qianwen Negatively.**

The data released by Mysteel indicates a decline in the prices of certain lithium battery materials today. This could potentially have a negative impact on the stock price of Mysteel, as one of its main businesses is the production and sale of lithium battery materials. Furthermore, the decrease in lithium carbonate prices may result in a decline in Mysteel's profitability, thereby affecting its stock price.<sup>6</sup>

#### **Baichuan Negatively.**

This news has a negative short-term impact on Mysteel's stock price, as the decline in lithium battery materials may affect the company's performance and stock price performance.<sup>7</sup>

Mysteel began as a provider of steel information services and gradually expanded into a data service provider for other bulk commodities and related industries. However, the news in the above example is actually a piece of information issued by Mysteel regarding the prices of lithium battery materials, with no specific information about Mysteel itself. We can see that ChatGPT provides the 'correct' answer, while Qianwen and Baichuan incorrectly link lithium battery production with Mysteel's performance.

## **3.2 Example 2**

### **Prompt**

**Forget all your previous instructions. Pretend you are a financial expert. Is the following news about the stock price?**

---

<sup>5</sup>The answer is originally in Chinese: 不确定。  
这则新闻提到了上海钢联发布的锂电材料报价下跌，但没有提供足够的信息来明确判断对上海钢联短期股价的正面或负面影响。影响股价的因素可能包括市场预期、需求和供应情况等，需要更多背景信息来做出准确的判断。

<sup>6</sup>The answer is originally in Chinese: 负面。  
上海钢联发布的数据显示，今日部分锂电材料报价下跌，这可能会对上海钢联的股价产生负面影响，因为上海钢联的主要业务之一是锂电材料的生产和销售。此外，碳酸锂价格的下跌可能会导致上海钢联的盈利能力下降，从而影响其股价。

<sup>7</sup>The answer is originally in Chinese: 负面。  
这个新闻对上海钢联的短期股价是负面的，因为锂电材料的下跌可能会影响到公司的业绩和股价表现。

**Huadong Medicine: Subsidiary HDM1002 Tablet Clinical Trial Application Approved.** <sup>8</sup>

**Chatgpt** No.

**Qianwen** Yes.

**Baichuan** Yes.

We can see that ChatGPT once again provides the ‘correct’ answer, while Baichuan and Qianwen do not give the expected answer. In subsequent research, we use ChatGPT to differentiate between fundamental news and market (stock-price-related) news.

## 4 Correlation Analysis

### 4.1 ChatGPT News Score

We utilize correlation analysis to validate whether the assessments of LLMs on news align with the actual trends in corresponding stocks. Specifically, we define a news score  $\epsilon_{i,t}$  for the  $i$ th stock before the trading day  $t$ : We map ChatGPT’s responses to three numerical values: positive are labeled as 1, uncertain as 0, and negative as -1. If a company has multiple news on the same day, the average is taken. For non-consecutive trading days, the news score is calculated as the multi-day average. Then we have

$$-1 < \epsilon_{i,t} < 1$$

### 4.2 Daily Stock Return

The stock returns are calculated in two different cases:

1. If the news is reported during the opening hours (9:30am–3:00pm) on a trading day, then the corresponding stock return is

$$r_{i,t} = \frac{C_{i,t+1} - C_{i,t}}{C_{i,t}}$$

where  $C_{i,t}$  is the close price that day, and  $C_{i,t+1}$  is the close price on the next trading day.

---

<sup>8</sup>The original prompt is in Chinese: 忘记你之前的所有指令。假装你是一位金融专家。判断以下新闻是关于股票行情的吗？  
华东医药：子公司 HDM1002 片临床试验申请获批。

2. If the news is reported during the non-open hour, then the corresponding stock return is

$$r_{i,t} = \frac{C_{i,t+1} - O_{i,t+1}}{O_{i,t+1}}$$

where  $O_{i,t+1}$  is the open price on the upcoming trading day.

### 4.3 Statistics on Different Answers

Among ChatGPT's 73,726 responses, there are a total of 34,284 positive responses (46.5%) and 11,177 negative responses (15.2%). We calculate the average returns for the stocks that have good news and bad news respectively:

$$\begin{aligned}\mathbb{E}[r_{i,t} | GPT \text{ score} = 1] &= 0.00032 \\ \mathbb{E}[r_{i,t} | GPT \text{ score} = -1] &= -0.0021\end{aligned}$$

One can see that the average decline of stocks that have bad news is significantly greater than the average increase of stocks that have good news. One might wonder if this gap stems from the difference between average decline and average increase of *all* stocks, namely, stock price might fall faster than they rise. The answer is no:

$$\begin{aligned}\mathbb{E}[r_{i,t} > 0] &= 0.027 \\ \mathbb{E}[r_{i,t} < 0] &= -0.023\end{aligned}$$

We see that the average increase in stock prices is even greater than the average decrease. So we come the conclusion that to some extent the market is more sensitive to negative news than to positive news.

### 4.4 Correlation Coefficients

We employ the Pearson correlation coefficient

$$\rho = \frac{\text{cov}(\epsilon, r)}{\sigma_{\epsilon} \sigma_r}$$

to calculate the correlation between the ChatGPT news score sequence  $\epsilon^{gpt} = (\epsilon_{i,t})$  and daily stock returns  $r = (r_{i,t})$ .

#### Sina Finance

Using the news headlines from Sina Finance to perform the above correlation calculation on all weighted stocks of the CSI 100 (totaling 13,076 news items), we obtain

$$\rho_s = 0.0226$$



## Jinshi Data

Using the news briefings from Jinshi Data to perform the correlation calculation on stocks of all listed companies in the A-share market, we have

$$\rho_j = 0.0197$$

## 4.5 Analysis

The news in the stock section of Sina Finance (web version) is concentrated under the sub-sections of various listed companies. The target audience is more inclined to be viewers interested in specific listed companies (otherwise they wouldn't access the page through layers of links), resulting in relatively more coverage of less important reports and news about that particular listed company. This may include some non-financial promotional content.

For example, in the news section for China Railway (SH601390) on September 11, 2023, there is a news headline: 'Main Force of Traditional Infrastructure Writes a New Chapter, Central Enterprise Holding Listed Company Strengthens New Infrastructure.'<sup>9</sup> Although ChatGPT's response to this news headline on stock prices is uncertain, due to a considerable number of similar news, it may still influence ChatGPT's scoring and correlation to some extent.

On the other hand, news briefings from Jinshi Data are more geared towards general subscribers interested in the financial market. Consequently, they provide relatively more important information about listed companies, especially stock-sensitive news or policies. We can infer this from the volume of news data: historical news data for various companies on Sina Finance has different time spans, with longer ones from April 2023 to September 2023, covering 6 months; shorter ones from July 2023 to September 2023, covering only 3 months. There are a total of 13,076 news headlines for 100 companies, averaging 130 news headlines per company. In contrast, the time span for news data in Jinshi Data is from November 2022 to September 2023, lasting 11 months; a total of 4,091 companies, 37,162 news briefings, with an average of fewer than 10 news briefings per company.

Surprisingly, from the above correlation calculations, we observe that despite significant differences between the two news data sources in terms of time span, number of companies, types and quantity of news, the statistically derived correlations are very close ( $\rho_s = 0.0226$ ,  $\rho_j = 0.0197$ ), within the margin of statistical error.

---

<sup>9</sup>The original headline is in Chinese: 传统基建主力军谱写新华章 央企控股上市公司发力新基建

Next, we apply the same process to the data using Baichuan and Qianwen. The resulting correlations are  $\rho^b = -0.00159$  for Baichuan and  $\rho^q = 0.0236$  for Qianwen. Firstly, we note that the correlation obtained by Baichuan is close to 0. Further analysis reveals that Baichuan assigns an ‘uncertain’ rating to 96.9% of the responses. Consequently, 96.9% of the values in the corresponding news score sequence are 0, resulting in the low correlation  $\rho^b = -0.00159$  for Baichuan. This indicates that Baichuan’s ability to assess the impact of news on stock prices is not as strong as Qianwen and ChatGPT.

Moreover, Qianwen’s correlation  $\rho^q = 0.0236$  is quite consistent with ChatGPT, demonstrating Qianwen’s comparable ability to assess the impact of news on stock prices as ChatGPT-3.5, and further confirming that the correlation between news and the A-share stock prices is around 0.02.

Lopez-Lira and Tang(2023)[9] conducted a similar correlation analysis on the U.S. stock market. They selected news headlines related to U.S. listed companies from October 2021 to December 2022 and obtained a correlation between daily stock returns and news headlines (totaling 67,586):

$$\rho_0 = 0.018$$

Despite significant differences in languages, assets, time frames, news sources, and data processing methods, by using the same statistical approach, the results we obtain on the A-share market is similar to those in the U.S. stock market. This indicates a general correlation between news and stock prices, which is independent of different financial systems, media environments, and even different time periods.<sup>10</sup>

## 5 Efficient Market Hypothesis

In 1970, Fama proposed the Efficient Markets Hypothesis (EMH), which defines an efficient market as one where prices fully reflect all available information in a securities market. EMH comes in three forms:

**The Weak Form** In a weak form of efficiency, market prices already fully reflect all historical information about securities, including stock transaction prices, trading volume, short selling amounts, financing information, and so on. If this hypothesis holds true, it is not possible to use technical analysis to predict stock price returns.

---

<sup>10</sup>Given the limited time span of these two studies, this conclusion requires further research for confirmation.

**The Semi-Strong Form** This form asserts that prices fully reflect all publicly available information about the company's operational prospects. This information includes transaction prices, trading volume, profit data, earnings forecasts, the management status of the company, and other publicly disclosed financial information. If this form of the hypothesis is valid, it would be ineffective to make investment decisions based on acquiring news or information about unexpected events.

**The Strong Form** This form posits that prices fully reflect all information about a company's operations, including both publicly available and undisclosed internal information. If this hypothesis holds true, obtaining insider information in advance would not help on investments.

## 5.1 Categorizing News

We can test the EMH by categorizing historical news from Jinshi Data. Specifically, we classify historical news into fundamental news and market news using ChatGPT.

**Prompt** Forget all your previous instructions. Pretend you are a financial expert. Regarding the news in the following {}, is the part about \*\*\* stock-price related news or fundamental news? Answer only with 'stock-price' or 'fundamental'.<sup>11</sup>  
{\*\*\*\*\*}(News briefing)

**Example 1** 【Market News】 In the afternoon, A-share stocks with tickers starting with Chinese characters continue to rise sharply. China Railway Construction (20CM) has hit the daily limit, and several other stocks such as China State Construction Engineering, China Metallurgical Group, China Unicom, and China Steel International have also hit the daily limit.<sup>12</sup>  
Chatgpt: stock-price

**Example 2** 【BYD's all-electric bus secures a large order in Norway】 Recently, BYD signed an order for 64 large buses with Nobina, the largest public transport operator in the Nordic region. These vehicles are scheduled

---

<sup>11</sup>The original prompt is in Chinese: 忘记你之前的所有指令。假装你是一位金融专家。下面 {} 中的新闻中, 关于 \*\*\* 的部分是股价新闻还是基本面新闻? 只回答“股价”或者“基本面”。

<sup>12</sup>The original news is in Chinese: 【行情】午后中字头个股继续大幅走强, 中铁装配 20CM 涨停, 中国交建、中国中冶、中国联通、中钢国际等多只个股涨停。

for delivery in the second quarter of 2023 and will be put into operation in the Nedre Glomma region of Norway. Since establishing a partnership with BYD in 2015, Nobina has operated over 300 BYD electric buses in Sweden, Norway, and Finland.<sup>13</sup>

Chatgpt: fundamental

We can see that ChatGPT’s differentiation between market news and fundamental news aligns with our general definitions of fundamental and market news. We analyze 500 news entries, and among them, 471 are consistent with our general differentiation between fundamental and market news. ChatGPT achieve an accuracy rate of 94.2%.

## 5.2 Correlation Tests

We can further conduct correlation analysis on the categorized news after filtering out fundamental news and market news with ChatGPT. We use ChatGPT, Baichuan, and Qianwen respectively to calculate the correlation for both fundamental news and market news in the approach discussed in 4.4.

	ChatGPT	Qianwen	Baichuan
Fundamental News	-0.0116	-0.0068	-0.0145
Market News	0.0350	0.0378	0.0124

Table 1: Correlations on Categorized News

## 5.3 Analysis

After calculating the correlation for categorized news, we find that the correlation between market news and stock prices is significantly higher than the correlation between fundamental news and stock prices. We believe this situation indicates the following issues:

1. The news briefings we select contain the names of listed companies, but more general fundamental news may not necessarily include company

<sup>13</sup>The original news is in Chinese: 【比亚迪纯电动巴士获挪威大单】近日，比亚迪与北欧最大公交运营商 Nobina 签订 64 台大巴订单，这批车辆计划于 2023 年第二季度交付，在挪威 Nedre Glomma 地区投入运营。据了解，自 2015 年与比亚迪建立合作关系以来，Nobina 已有超 300 辆比亚迪电动巴士在瑞典、挪威和芬兰运营。

names. For example, industry news, the release of national policies and economic indicators, changes of the exchange rates, and even international geopolitics could affect the stock price of a particular company. However, most of these types of news are not included in our filtered news list.

2. We can conclude that the impact of company-specific news on the stock prices of A-share is minimal, and it is more likely influenced by macroeconomic and policy factors.
3. On the other hand, A-shares are to a considerable extent influenced by market news. Because ChatGPT's assessment on news related to price surges is positive, the positive correlation indicates that a trend tends to continue in the A-share market. This implies that participants in the A-share market are like to buy more if the prices go up; and like to sell more if the prices go down.
4. Finally, we analyze the correlation results in the context of the Efficient Market Hypothesis:
  - (a) Firstly, at least in the short term, the weak form of the Efficient Market Hypothesis is ruled out because we observe a correlation between historical stock prices and future stock prices. Therefore, we can, to some extent, use technical analysis to predict stock price returns.
  - (b) Because strong efficiency implies semi-strong efficiency, and semi-strong efficiency implies weak efficiency, logically, the Efficient Market Hypothesis does not hold in the short term in the A-share market. Regarding whether the Efficient Market Hypothesis holds in the long term, our data cannot provide support.

## 6 Stock Investment Strategies

We can design trading strategies based on the LLMs' scores for news. The goal is to intuitively understand the relationship between different news categories and asset prices through the comparison of equity curves and the SSE Composite Index, as well as to compare the abilities of different models.

### 6.1 Trading Strategies

We adopt a simple trading strategy as to reflect the capabilities of the LLMs.

1. Before the close of each trading day, we aggregate all news that occurred after the opening of the day. For stocks with news, we calculate the average score given by the LLM. Based on whether the average score is greater than 0, we decide the investment direction: go long at the close if the average score is greater than 0, go short if it is less than 0, and do not trade if it equals 0. We close positions at the close of the next trading day.
2. Before the open of each trading day, we aggregate all news that occurred after the close of the previous trading day. For stocks with news and no existing position, we calculate the average score given by the LLM. Based on whether the average score is greater than 0, we decide the investment direction: go long at the open if the average score is greater than 0, go short if it is less than 0, and do not trade if it equals 0. We close positions at the close of the same day.

**Positions** We adopt an equal-weighted strategy, where the funds (1 yuan) are evenly distributed among all stocks with news on the given day.<sup>14</sup>

## 6.2 Strategy Performance Based on ChatGPT

First, let us examine the strategy performance based on ChatGPT's scores for news. We provide statistics for the scenarios of only going long, only going short, going both long and short, and going long for all stocks with news. We also compare these results with the SSE Index (SHA:000001).

---

<sup>14</sup>Here, we ignore restrictions in A-shares regarding minimum position sizes and integer multiples of positions. Ultimately, these requirements can be met by increasing the total funds.

## Uncategorized News

Figure 1 shows the backtesting results based on uncategorize news:

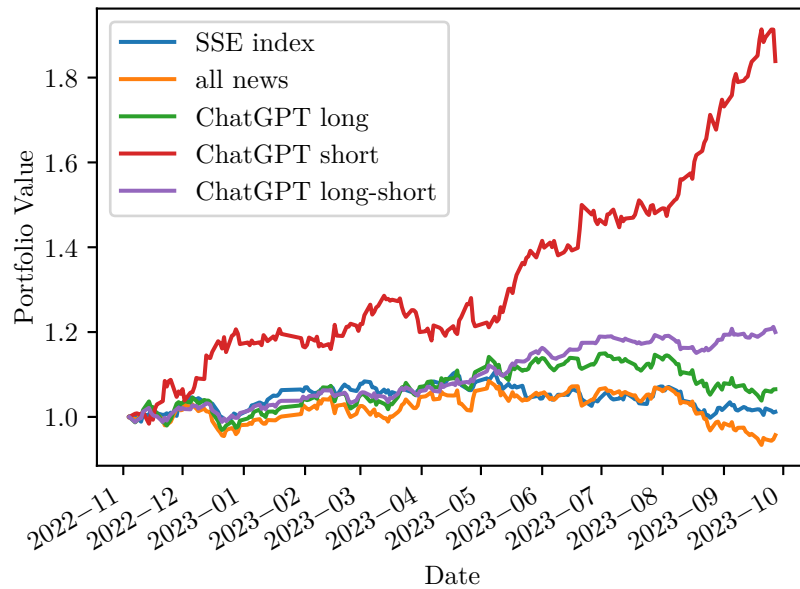


Figure 1: Equity Curve for Uncategorized News

We observe that the long strategy based on ChatGPT's scores can achieve a 6.5% return over an 11-month period, while the short strategy can yield an 83.9% return. Compared to the 1.2% growth in the SSE Index, ChatGPT's news-assessment-based strategy gains a significant outperformance.

## Fundamental News

Figure 2 shows the backtesting results based on fundamental news:

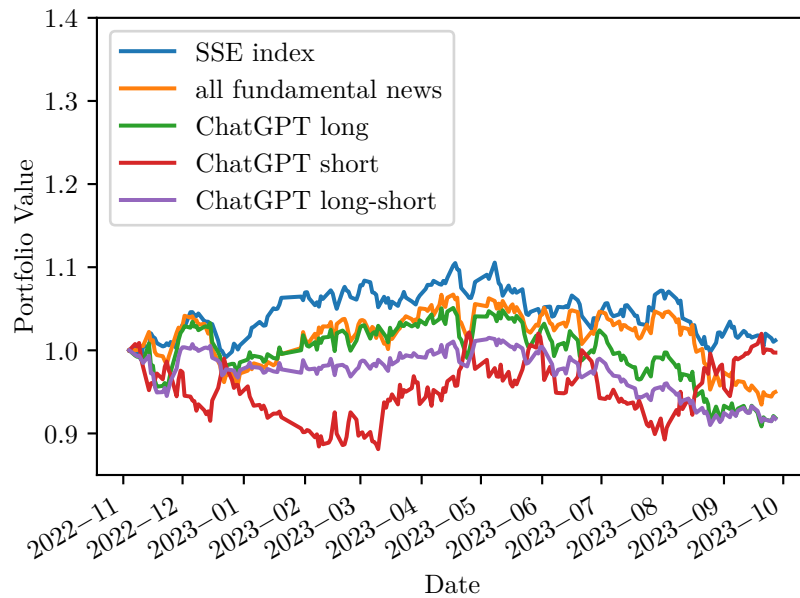


Figure 2: Equity Curve for Fundamental News

It can be seen that the strategy based on fundamental news does not outperform the index in terms of returns: the long strategy ultimately incurs a loss of -8.1%, while the short strategy results in a loss of -0.27%.



## Market News

Figure 3 shows the backtesting results based on market news:

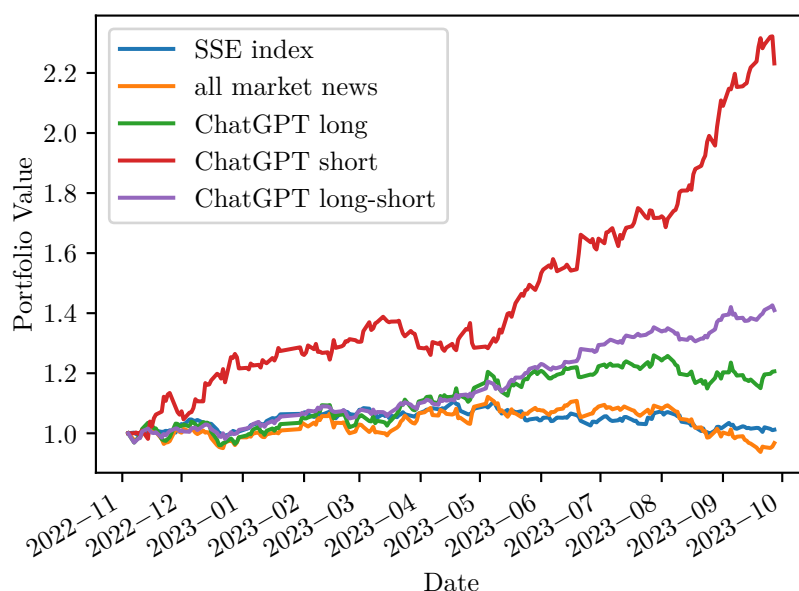


Figure 3: Equity Curve for Market News

On the other hand, the strategy based on market news exhibits the most outstanding performance: the long strategy achieves a final return of 20.6%, while the short strategy yields a final return of 123.2%.

## 6.3 Model Comparison

Next, we compare the results of ChatGPT with those based on Qianwen and Baichuan. We sequentially present the backtesting results for the three strategies of only long, only short, and both long and short, based on the scores from the three different models. Additionally, we compare the three scenarios of all news, fundamental news, and market news. See appendix A for the equity curves.

We summarize the total returns of the three models under different scenarios in the table below:

	ChatGPT	Qianwen	Baichuan
All News(long)	6.5%	5.9%	5.6%
All News(short)	83.9%	43.6%	36.4%
All News(long-short)	20.0%	15.9%	11.1%
Fundamental News(long)	-8.1%	-11.8%	-30.9%
Fundamental News(short)	-0.3%	3.5%	27.9%
Fundamental News(long-short)	-8.2%	-9.1%	-22.5%
Market News(long)	20.6%	16.6%	46.1%
Market News(short)	123.2%	93.2%	11.6%
Market News(long-short)	40.1%	31.5%	57.1%

Table 2: Returns Statistics

From the above backtesting curves and statistical results, we observe:

1. The overall performance of the three models shows that ChatGPT > Qianwen > Baichuan, which is consistent with our analysis of correlation.<sup>15</sup>
2. The overall performance based on the news types shows that market news > all news > fundamental news.
3. The overall performance of restricting positions shows that only short positon > long and short position > only long position.

## 7 Conclusion

Based on the correlations and backtestings, we analyze and summarize the results from the following aspects.

### Data Sources

First, we separately calculate the correlations between financial news from Sina Finance and Jinshi Data and stock returns. We then compare these results with the correlations calculated in [9]. We find a remarkably consistent in correlations in spite of different data sources, time periods, and financial markets:

$$\rho \approx 0.02$$

---

<sup>15</sup>See table 1.

We believe that there may be a more general correlation between stock price and news, surpassing different countries, public opinion environments, and financial markets. This correlation needs further research for confirmation.

## News Types

We use ChatGPT to classify the news briefings from Jinshi Data into fundamental and market news. The correlation between market news and stock prices is much higher than that of fundamental news. Additionally, as seen in the backtesting results 6.2, the performance of the strategy based on market news significantly outperforms that based on fundamental news. These two findings collectively suggest that, in the A-share market from November 2022 to October 2023, EMH does not hold at least in the short term. Moreover, fundamental news affecting stock prices is more likely to come from macro-level news, such as industry updates, policy changes, or the release of financial and economic indicators, rather than company-specific news.

## Model Performances

We conduct a comparison between ChatGPT, Baichuan, and Qianwen in both correlation calculations and backtesting. The results show that Qianwen are very close to ChatGPT in both correlation results and backtesting, but it performs weaker than ChatGPT in various aspects. On the other hand, Baichuan exhibits a high level of uncertainty in news scoring (96.9%), resulting in low correlation between its scores and stock prices. Additionally, backtesting based on Baichuan's scores shows significant volatility. Therefore, we observe that Baichuan performs weaker than both ChatGPT and Qianwen in news assessment.<sup>16</sup>

## Long vs. Short

During the backtesting period, the SSE Index rise by 1.2%. However, we observe that the strategy of only going short outperforms the strategy of only going long. This suggests that even in an overall uptrend, going short still generates excess returns. One possible reason is that negative news occurs less frequently than positive news in China, making the market more sensitive to negative information. This observation is consistent with the mean return calculations for positive and negative news classified by ChatGPT in Section 4.3.

---

<sup>16</sup>This is limited to the scope of the prompts used in this study.

# Appendices

## A Backtesting Results for Different Models

**Uncategorized News** The following figure displays the backtesting results of the strategy based on the LLM scores for all uncategorized news:

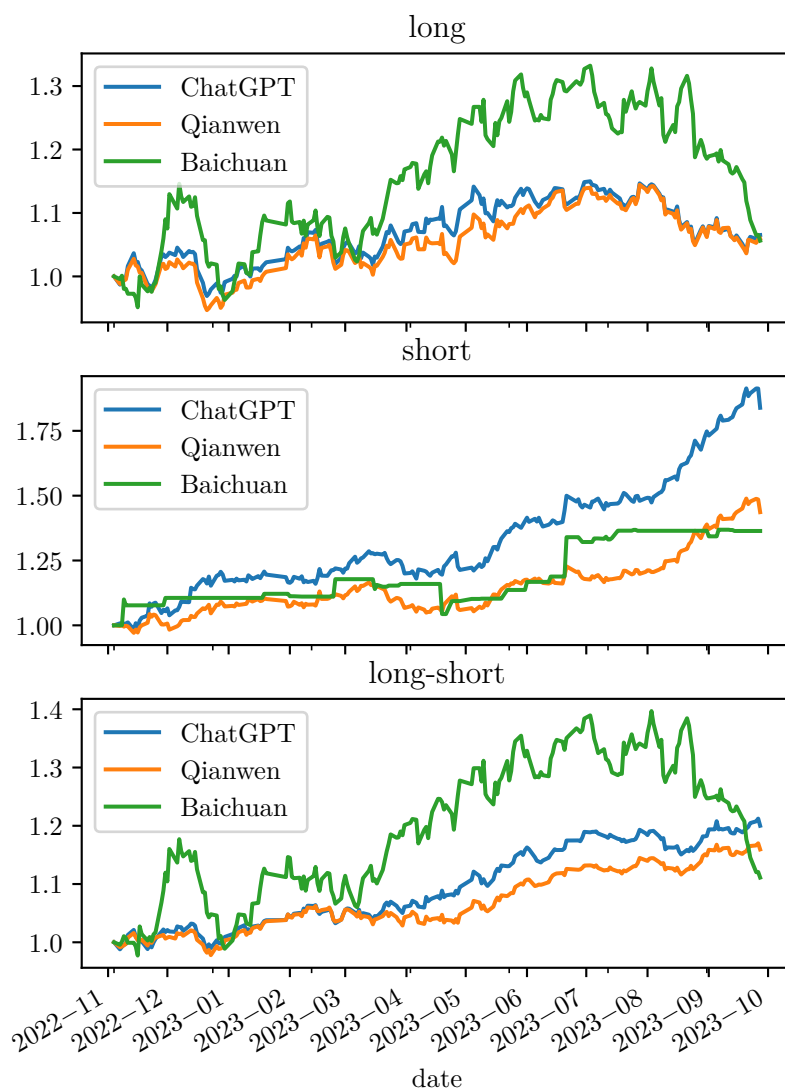


Figure 4: All News(Uncategorized)

**Fundamental News** The following figure displays the backtesting results of the strategy for the fundamental news:

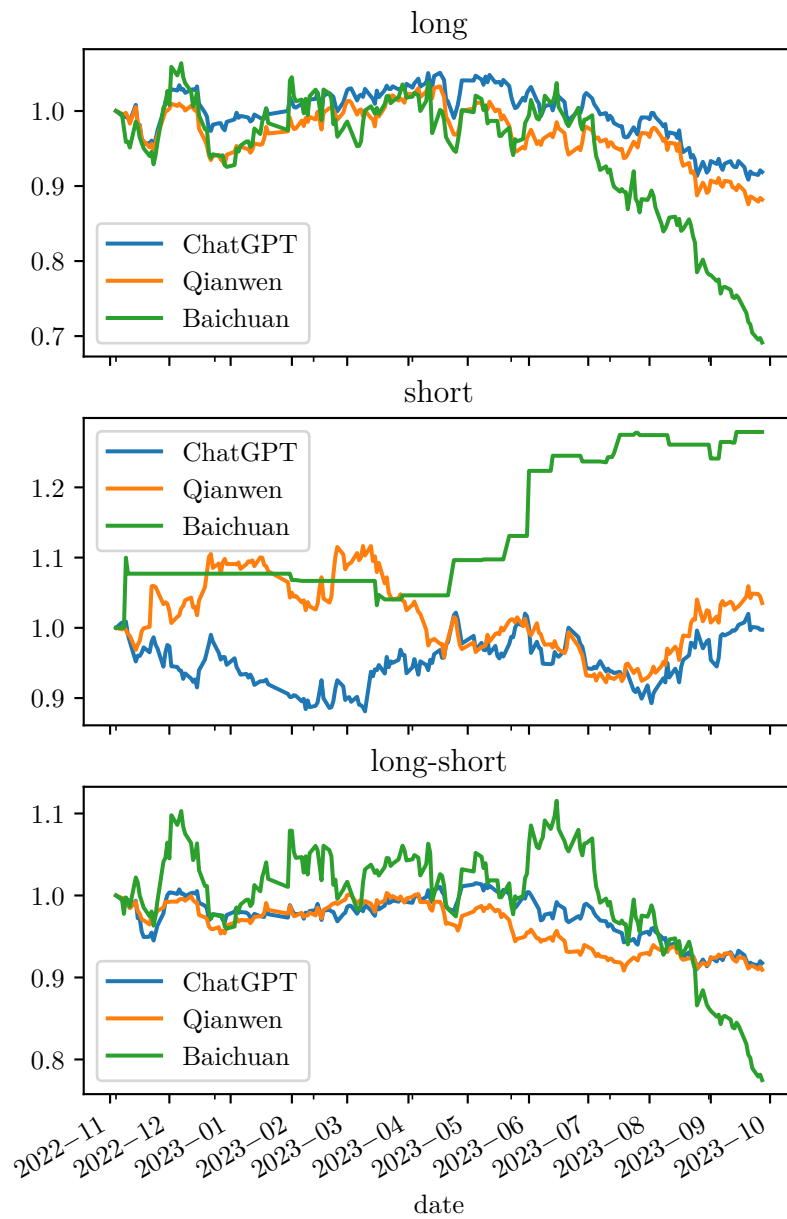


Figure 5: Fundamental News

**Market News** The following figure displays the backtesting results of the strategy for the market news:

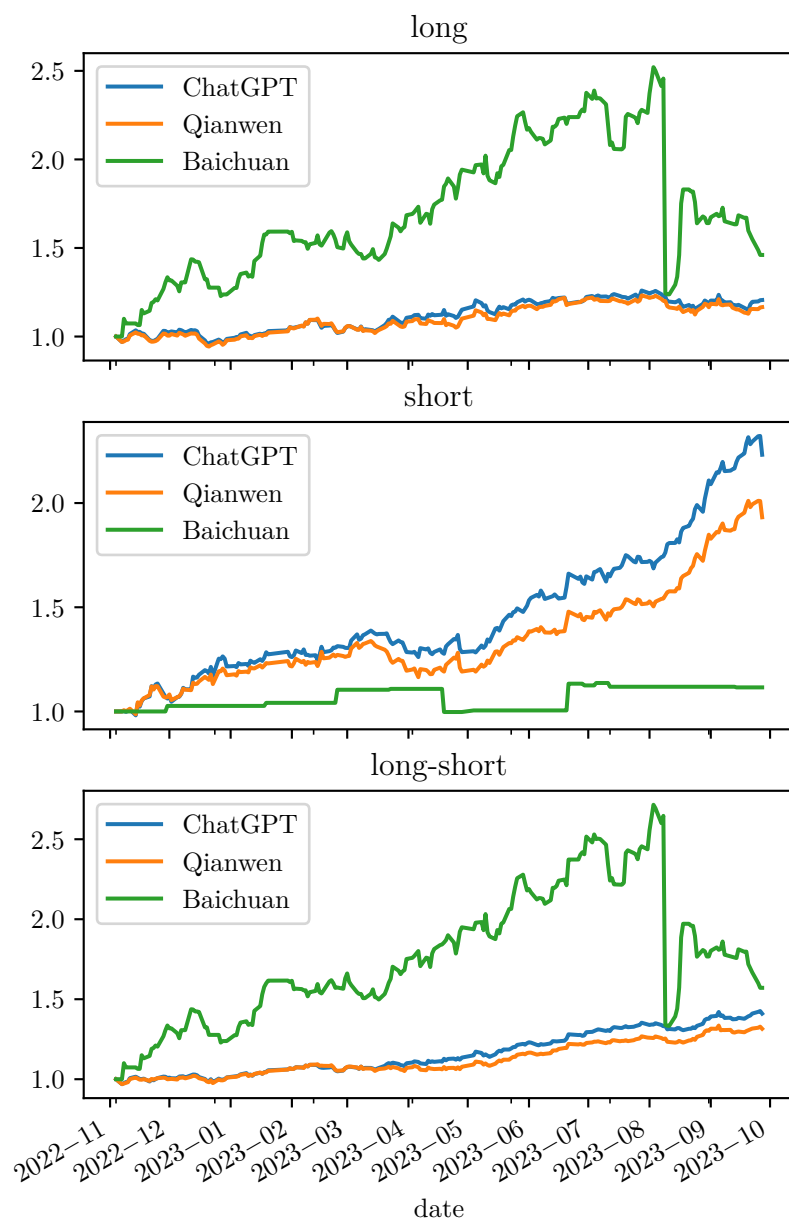


Figure 6: Market News

## References

- [1] Boris Kovalerchuk and Evgenii Vityaev. “Data mining for financial applications”. In: *Data Mining and Knowledge Discovery Handbook* (2005), pp. 1203–1224.
- [2] Arash Bahrammirzaee. “A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems”. In: *Neural Computing and Applications* (2010), pp. 1165–1195.
- [3] Yuan Qi and Jing Xiao. “Fintech: AI powers financial services to improve people’s lives”. In: *Commun* (2018), pp. 65–69.
- [4] Tom Brown et al. “Language models are few-shot learners”. In: *Advances in neural information processing systems* (2020), pp. 1877–1901.
- [5] Tyler Cowen and Alexander T. Tabarrok. “How to Learn and Teach Economics with Large Language Models, Including GPT”. In: *SSRN Electronic Journal* (Mar. 2023). ISSN: 1556-5068. DOI: <https://doi.org/10.2139/SSRN.4391863>.
- [6] Anne Lundgaard Hansen and Sophia Kazinnik. “Can ChatGPT Decipher Fedspeak?” In: *SSRN Electronic Journal* (Mar. 2023). ISSN: 1556-5068. DOI: <https://doi.org/10.2139/SSRN.4399406>.
- [7] Arun Jambulapati et al. *Testing Causality for High Dimensional Data*. 2023. arXiv: 2303.07774 [cs.LG].
- [8] Anton Korinek. “Language Models and Cognitive Automation for Economic Research”. In: (Feb. 2023). DOI: <https://doi.org/10.3386/W30957>.
- [9] Alejandro Lopez-Lira and Yuehua Tang. “Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models”. In: *Social Science Research Network* (Apr. 2023). DOI: <http://dx.doi.org/10.2139/ssrn.4412788>.
- [10] Shakked Noy and Whitney Zhang. “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”. In: *SSRN Electronic Journal* (Mar. 2023). DOI: <https://doi.org/10.2139/SSRN.4375283>.