Black-Litterman Portfolio Optimization with Asset Universe Given by Large Language Models

Xiangxi Kong, Liangyu Min*, *Member*, *IAENG*, Dijia Lin, Zhen Li

***Abstract*—Generally, the Black-Litterman portfolio model relies heavily on the expert investment opinions, where high-quality investment opinions would improve model performance whereas low-quality investment opinions would result in poor model performance. Essentially, expert investment opinions are highly subjective, sourcing from these financial experts' information and knowledge. ChatGPT, as an advanced generative AI model, could extract and analyze information from huge multi-modal data, which is beneficial to build AI investment opinions for Black-Litterman portfolio model. In this study, we construct and analyze the large language model-based Black-Litterman portfolios, ChatGPT-BL and BARD-BL, where the goals of minimum variance and mean-variance trade-off are taken into account. Computational results show that the ChatGPT-based portfolios tend to be conservative, which is suitable for risk-averse investors, while the BARD-based portfolios are aggressive, which is appropriate for risk-seeking investors. Also, the superiority of Black-Litterman model using investor views generating from gradient boosting regression and GJR-GARCH algorithm is illustrated by the efficient frontiers.**

***Index Terms*—Portfolio selection, Machine learning, ChatGPT, Black-Litterman**

I. Introduction

WITH the development of generative artificial intelligence (AI) technique, some financial institutions have started to apply or develop vertical large models to improve service efficiency. ChatGPT[1], as one of the most successful large language model (LLMs), has been demonstrated to be able to generate detailed and informative contents based on the analysis of massive multi-modal data. Motivated by the huge success of ChatGPT, some scholars have made attempts to build financial models based on some LLMs such as ChatGPT/GPT-4 from OpenAI, Bard[2] from Google, and Claude2[3] from Anthropic. However, only a small number of model test results are released, and whether investors can trust these GPT-based strategies is still questionable. This paper aims to sort out the existing research results and investigate the possibility of developing the portfolio model based on some LLMs such as ChatGPT and Bard (Gemini).

Romanko, Narayan, and Kwon[4] explore the potential uses of ChatGPT for dealing with portfolio selection problem, where the universe of stocks are given by ChatGPT while the investment weights are calculated by the associated quantitative optimization techniques. The key finding of their research is that ChatGPT may be effective in stock selection, but it is not good at assigning asset weight because the GPT model is lacking in the support of conceptual framework. As a result, our paper tries to simulate an investment expert using GPT models to provide available stocks from S&P 500 constituents for the sequel portfolio modelling.

Kim[5] investigates the effectiveness of making ChatGPT as an asset manager, where they ask ChatGPT to recommend asset classes considering different economic conditions. Their study defines eight different economic conditions based on interest rate, inflation, and US dollar. Then, they ask ChatGPT to recommend three assets assuming it is a investment manager under the designed economic background using the prompt of "Suppose you are an asset manager. Pick at most three asset classes that are good investment under the following economic conditions: rising interest rate, rising inflation, and strengthening dollar". Based on that, some classical portfolio models including equal-weighted, mean-variance, global minimum variance, and risk parity are constructed and tested. The provided out-of-sample numerical experiments (October 2022 to July 2023) show that the ChatGPT-based optimal portfolio models achieve higher returns than benchmarks, but no dominance in volatility and downside risk could be observed from the ChatGPT-based optimal portfolio performance.

Saggu & Ante[6] use the synthetic difference-in-difference method to analyze the influence of ChatGPT on AI-related crypto assets, whose computational results show that after the ChatGPT launch, the one-month of crypto assets average returns ranging between 10.7% and 15.6%, suggesting the potential for managing crypto assets aiding by some innovative technologies such as ChatGPT, which also provides some empirical foundations for this research.

Considering the relevant laws and regulations, GPT models can not provide specific investment recommen-

Manuscript received Jan 18, 2024; revised Jun 15, 2024. This work is supported by the National Natural Science Foundation of China (No. 71971132), the Fundamental Research Funds for the Central Universities (No. 2020110552). This work is also supported by the special fund for spiritual research of the 20th CPC national Congress from Shanghai Business School (No. AX-10999-1103-1102) and the College Young Teacher Training Subsidy Plan of Shanghai (No. AG24-33811-3301). And, this study is supported by the Startup Fund for Advanced Talents of Putian University (2020008).

Xiangxi Kong is a lecturer in the College of Business Administration, Ningbo University of Finance & Economics, Ningbo Zhejiang 315175, China (E-mail: kongxiangxikxx@163.com).

Liangyu Min is a lecturer in the Faculty of Business Information,Shanghai Business School, Shanghai 201400, China (Corresponding author to provide email: 21220001@sbs.edu.cn).

Dijia Lin is a PhD candidate of School of Information Management & Engineering, Shanghai University of Finance & Economics, Shanghai 200433, China (E-mail: dijialin@stu.sufe.edu.cn).

Zhen Li is an associate professor in the School of Business, Putian University, Putian Fujian 351100, China (E-mail: 408053275@qq.com).

dations. As a result, we design the hybrid algorithm to generate investment opinions for constructing the Black-Litterman portfolio[7], [8], [9], [10], [11], [12], where the ensemble machine learning method such as Gradient Boosting Regression[13] for return and GJR-GARCH[14] for volatility are involved. To combine the GPT model and machine learning algorithms into one unified framework, this paper selects the Black-Litterman portfolio model, which could fix the prior information observing from historical samples by some subjective inferences (e.g. investment experts opinions) according to the Bayesian rule. Therefore, the superiority of GPT and the predictive ability of machine learning algorithms can be fully utilized under this portfolio framework.

The rest of this paper is structured as follows. Section II gives the general description of the Black-Litterman theoretical framework. In section III, we use two LLMs, ChatGPT and Bard, to recommend assets from the S&P 500 for retail investors. Then, the ensemble machine learning algorithm, Gradient Boosting Regression (GBR), is called to generate investor opinions, and the volatility is modelled by the GJR-GARCH to give the confidence level of the algorithm generated investor opinions. The associated computational results are presented and analyzed in section V. Section VI concludes the paper and points out the further research direction.

II. Black-Litterman Portfolio Framework

Define the associated variables as follows. $N \in \mathbb{R}$ is the number of assets, and $\Sigma \in \mathbb{R}^{N\times N}$ is the asset covariance matrix. Generally, it is reasonable to assume that the asset return vector r follows the multivariate Gaussian distribution, i.e., $r \sim \mathcal{N}(\mu, \Sigma)$, where μ is the expected return vector for the involved asset. Under the scenario of rational individuals, the implied market equilibrium return is $\Pi = \lambda\Sigma x$, where each investor hold the identical maximization utility objective of $\max x^T\Pi - \lambda x^T\Sigma x$. $x \in \mathbb{R}^n$ represents the vector of portfolio weight, $\lambda \in (0, 1)$ indicates the coefficient risk-aversion, where the extreme risk-seeking investor has $\lambda = 0$ while the total risk-averse investor has $\lambda = 1$.

Black-Litterman framework assumes the return vector $\mu = \Pi + \varepsilon_\mu, \varepsilon_\mu \sim \mathcal{N}(0, \tau\Sigma)$, where τ denotes the uncertainty level of the estimated Σ from the historical sample points. Actually, there still existing some discussions about the appropriate value of τ, one can refer to the related papers [15], [16], [8], [17], [11].

Subjective investor opinions can be integrated into the portfolio model by the method of Bayesian. Define K as the number of investor opinions about the M assets, and the pick matrix $P \in \mathbb{R}^{K\times M}$, quantitative vector $Q \in \mathbb{R}^K$. Therefore, we have the following equations hold:

$$\begin{cases} P\mu = Q + \varepsilon_\mu \\ \varepsilon_\mu \sim \mathcal{N}(0, \Omega) \\ P\mu \sim \mathcal{N}(Q, \Omega) \end{cases}$$

And, the posterior distribution for the return vector μ is as follows:

$$\mathcal{N}([(\tau\Sigma)^{-1} + P'\Omega P]^{-1}[(\tau\Sigma)^{-1}\Pi + P'\Omega Q], [(\tau\Sigma)^{-1} + P'\Omega P]^{-1})$$

TABLE I
Asset universe given by ChatGPT and Bard.

LLM	Ticker 1	Ticker 2	Ticker 3	Ticker 4	Ticker 5
ChatGPT	AAPL	GOOG	JNJ	V	HD
Bard	PYPL	CMG	LULU	FTNT	AZO

Note that the requests to the LLMs are repeated 10 times, and we select the most frequently results as the recommended stocks to construct the asset universe. The companies represented by the symbols are as follows. AAPL: Apple Inc.; G.OOG: Alphabet Inc.; JNJ: Johnson & Johnson; V: Visa Inc.; HD: The Home Depot Inc.; PYPL: PayPal Holdings; CMG: Chipotle Mexican Grill; LULU: Lululemon Athletica Inc.; FTNT: Fortinet; AZO: AutoZone.

Thus, by optimizing on the posterior parameters achieved above, the Black-Litterman model would obtain the portfolio weight different from classical Markowitz portfolio[18], which only relays on the moment information from the historical samples. However, high-quality expert opinions are hard to get, hence, we use machine learning algorithms to generate the information about investor opinions. This strategy is reasonable, due to lots of financial engineering studies[16], [19], [20], [21], [22], [23], [24], [25] have demonstrated the effectiveness and efficiency of machine learning algorithm applied in financial modelling.

III. Asset Universe From LLM

ChatGPT and Bard are the two LLMs to generate the asset universe for building the Black-Litterman portfolio model. The prompt used in this study sourcing from Romanko et al[4]: "You are an investment expert, use a range of investing principles taken from leading funds, create a theoretical fund comprising of at least 5 stocks (mention their tickers) from the S&P 500 with the goal to outperform the S&P 500 index."

Table II summarizes the fundamental financial information of the recommended companies, where the market capital, β, and PE are presented. It can be observed that ChatGPT prefers the companies with high market capital (all of the involved companies have market capital over 300 billion dollars), whereas Bard is interested in the companies with relative small market capital (all of its recommended companies have market capital lower than 100 billion dollars). From the perspective of β, ChatGPT tends to give stock with β around 1, indicating the recommended assets show similar risk-level with the market. On the contrary, Bard pays attention on the assets with β significantly deviating from 1, suggesting the selected companies have specific behaviors different from the market. In addition, PE ratio is a financial metric that is widely used to assess the relative valuation of a company's stock, which can be calculated by $\frac{\text{Market Price per Share}}{\text{EPS}}$, where EPS is short for earnings per share. For the stock with high PE ratio indicating the investors have high expectations for future earnings growth, and the stock might be overvalued. For the stock with low PE ratio suggesting the stock might be undervalued. It can be found that some stocks with obviously high PE ratio are selected by Bard, whereas the PE ratios of the stocks given by ChatGPT are relatively stable, ranging from 20 to 32.

According to the recommendation results, ChatGPT shows a clear conservative investment style, while Bard

TABLE II
FUNDAMENTAL FINANCIAL INFORMATION OF THE INVOLVED COMPANIES.

Portfolio	Ticker	Market Cap	Beta	PE Ratio
ChatGPT-based	AAPL	2.829T	1.29	29.63
	GOOG	1.72T	1.05	26.44
	JNJ	386.68B	0.53	30.14
	V	533.55B	0.95	31.35
	HD	336.98B	0.98	21.72
Bard-based	PYPL	63.02B	1.38	17.45
	CMG	60.94B	1.33	52.59
	LULU	62.59B	1.38	63.02
	FTNT	44.83B	1.09	40.55
	AZO	44.22B	0.67	18.62

tends to be more aggressive. To calculate the market implied return Π, we use the S&P 500 return as the market, and the coefficient $\lambda = \frac{R-r_f}{\sigma^2}$ based on the standard theory, where $r_f = 3\%$ representing the annual risk-free rate.

IV. INVESTOR VIEWS FROM ALGORITHMS

Based on the selected stocks by ChatGPT and Bard, the investment views forming the matrix Q are given by the Gradient Boosting Regression(GBR), which can be used to predict continuous output variable. Then, the confidence level of the generating opinions are gauged by the customized algorithm, where the GJR-GARCH is used to model the asset volatility. In general, for the asset with higher volatility, the associated investment opinion given by GBR has lower confidence level. Accordingly, the customized algorithmic framework adjusts the confidence parameter based on the GJR-GARCH forecasting volatility.

A. Gradient Boosting Regression

As an ensemble machine learning model, GBR builds trees sequentially, with each tree aiming to correct the errors of the previous ones. Assume the continuous variable y can be predicted by the input feature space X. The goal of GBR is to minimize the loss function $L(y, F(X))$, at the ith iteration, GBR would fit a weak learner (decision tree) h_i for the negative gradient of the loss function, $h_i = -\frac{\partial L(y,F(X))}{\partial F(X)}$. Based on that, the model $F(X)$ can be updated as follows:

$$F_{i+1}(X) = F_i(X) + \nu \cdot h_i(X)$$

where ν is the hyperparameter of learning rate, controlling the contribution of each weak learner to the GBR. Therefore, the final prediction of GBR is the sum of the predictions from all the weak learners:

$$F(X) = \sum_{i=1}^{m} \nu \cdot h_i(X)$$

where m is the number of weak learners.

B. GJR-GARCH

GJR-GARCH method is an extension of the classical GARCH model, considering different volatility dynamics during the periods of positive and negative returns. Essentially, the GJR-GARCH is asymmetric, which can be expressed as follows:

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1} r_{t-1}^2$$

where σ_t^2 represents the conditional variance of the asset returns at time t, ω is the constant term denoting the long-run average conditional variance, I_{t-1} is the indicator function that takes value of 1 if r_{t-1} is negative, and 0 otherwise, which is helpful to describe the asymmetry of returns.

C. Proposed Algorithmic Framework

The proposed algorithmic framework is called BL-LLM, which combines the recommended asset universe by LLMs and the investor opinions generated from GBR. Based on the work of [15], the confidence level of the generated investor opinions are quantified by the designed method. Although, different from that, the corresponding algorithm in this study considers the asymmetric reactions of investors when facing bull and bear market conditions.

Algorithm 1 BL-LLM portfolio framework.

Input: Asset universe given by LLM; Historical stock sample D.

Output: ChatGPT-BL and BARD-BL portfolios; Investor views Q; Confidence matrix Ω.

1: On the sample data D, separate the training set D_{train}, on which the BL-LLM portfolios and benchmark strategies would be constructed. The residual is testing set D_{test}, on which the involved portfolios would be tested and evaluated.
2: Calculate the market implied return Π by the formula of $\lambda\Sigma x_{mkt}$, where x_{mkt} can be obtained by the weighted average of stock market capital.
3: According to the price & volume information in D_{train}, some technical indicators can be calculated with the help of TA-Lib[26].
4: Obtain the prediction $q_i \in Q$ by the GBR for each stock.
5: Construct the confidence matrix Ω using the forecasting volatility given by GJR-GARCH.
6: Build the ChatGPT-based Black-Litterman portfolio model.
7: Build the Bard-based Black-Litterman portfolio model.
8: **return** ChatGPT-BL, BARD-BL.

Algorithm 1 gives the basic logic of the proposed portfolio framework, where Ω is a diagonal matrix describing the variance of investor opinion. One of the heuristic method is to set $\Omega = \tau P \Sigma P'$, that is, Ω is proportional to the priors and $\tau = 1/N$ is a reasonable value. Alternatively, the Idzorek method[27] is also feasible, which supports to specify the uncertainties of the investor opinions.

Using GBR to generate investor opinions in the proposed framework is reasonable. As an ensemble learning method, GBR could combine the predictions of multiple weak learners to make a strong predictive model, by which the non-linearity and complexity attributes of the financial data could be captured. Over-fitting is a common discussed topic in machine learning algorithm, especially in some tree-based models. Fortunately, some regularization parameters such as learning rate and depth provide a certain

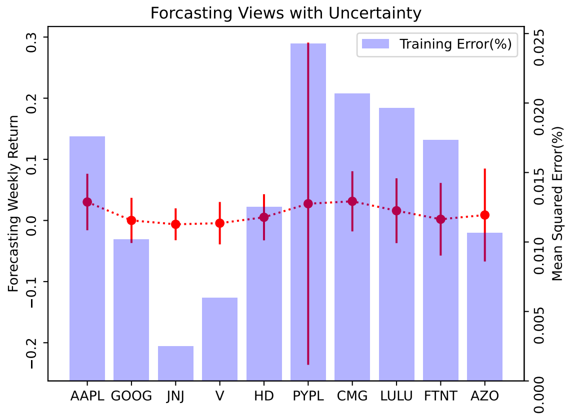

Fig. 1. The Generating Views.

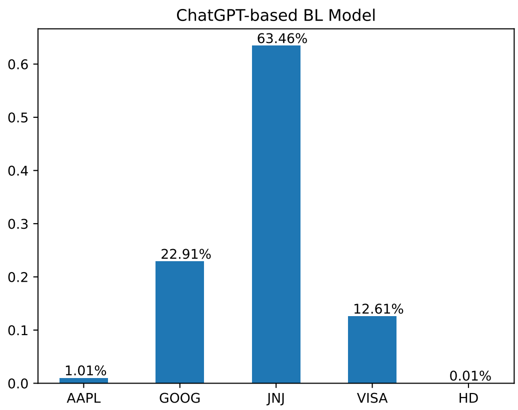

Fig. 2. Portfolio Weights of the ChatGPT-based BL Model.

level of robustness for GBR. Moreover, GBR can identify the importance of different features, which contributes to find the effective driving factors for the stock returns.

V. Computational Results

In this section, we report the performances of the developed Black-Litterman models, as well as some benchmark portfolios. $1/N$ portfolio (EW) is the important model for

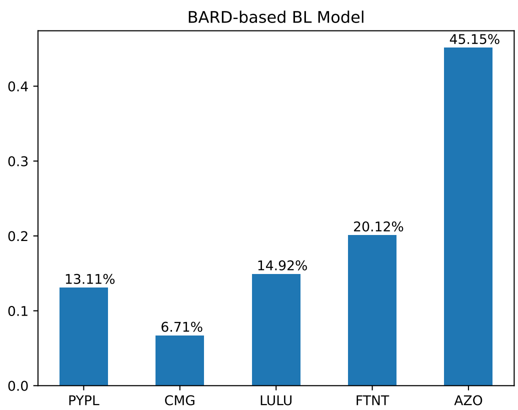

Fig. 3. Portfolio Weights of the BARD-based BL Model.

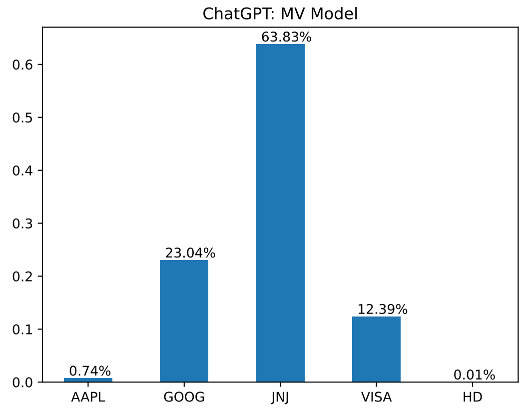

Fig. 4. Portfolio Weights of the ChatGPT-based MV Model.

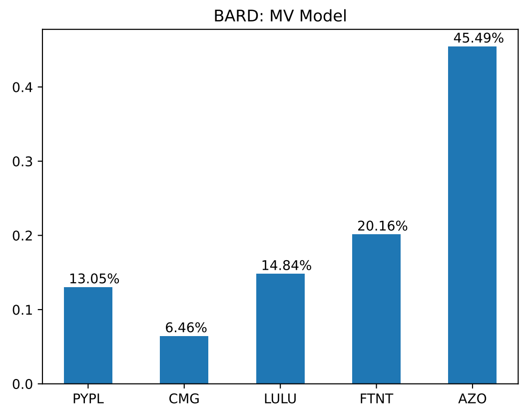

Fig. 5. Portfolio Weights of the BARD-based MV Model.

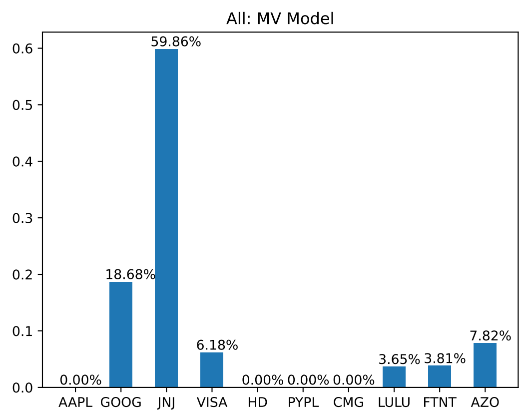

Fig. 6. Portfolio Weights of MV Model.

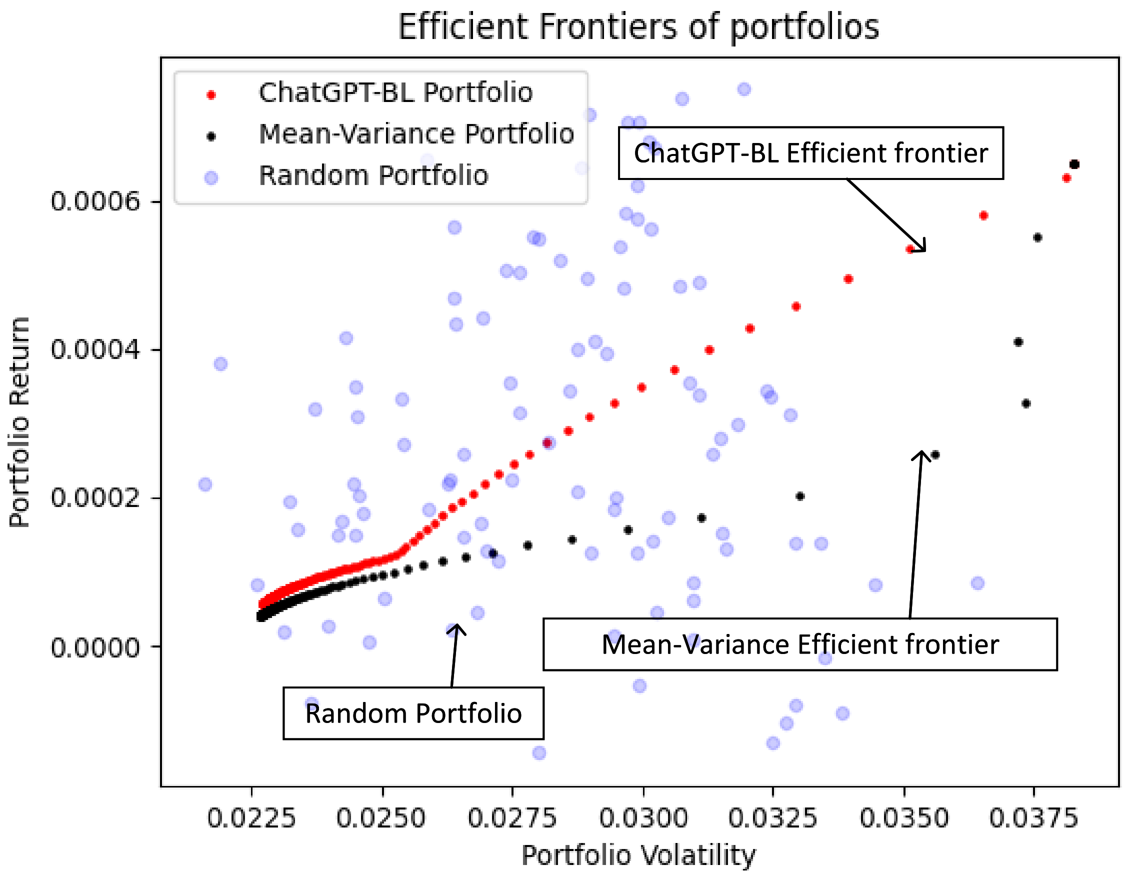

Fig. 7. Efficient Frontiers of the ChatGPT-based Black-Litterman portfolios.

evaluating as the baseline, which has been demonstrated to be quite robust on the out-of-sample performance[28]. Minimum-variance (MV) is the optimized portfolio model aiming to minimize the model risk level. Since some existing researches have illustrated that the classical mean-variance does not show satisfying out-of-sample portfolio performance due to the sensitivity to the estimated parameters from the historical data points, we omit this model in the out-of-sample tests.

A. Data Set

The computational results use the data provided by the Yahoo Finance, where the weekly samples ranging from Jan. 01, 2018 to Jan. 05, 2024 are involved, and the last 100 observations are set as the testing set. Table III gives the descriptive statistics of the geometric return rate, which can be calculated by the formula of $\log(r_t) - \log(r_{t-1})$. According to the descriptive statistics, FTNT has the largest return range, indicating it is volatile during the sampling period, which could also be verified by the results of variance. The Markowitz basic assumption is also reflected by the statistical results, since stocks with higher average returns tend to show higher volatility.

Note that the stock return rates do not follow the Gaussian assumption strictly, most of them are left-skewed (negative skewness) with thick tails (positive kurtosis). Hence, classical Markowitz portfolio theory may not cope

TABLE III
DESCRIPTIVE STATISTICS OF THE STOCKS RETURNS.

Stock	Range	Mean	Variance	Skewness	Kurtosis
AAPL	0.3302	0.0047	0.0016	-0.4045	2.2699
GOOG	0.2623	0.0029	0.0016	-0.0013	0.7702
JNJ	0.1991	0.0009	0.0007	-0.4047	1.8879
V	0.3146	0.0026	0.0012	-0.4417	3.8743
HD	0.5198	0.0023	0.0017	-0.8293	10.3231
PYPL	0.4669	-0.0010	0.0032	-0.4935	3.2109
CMG	0.4514	0.0063	0.0026	0.3415	4.0128
LULU	0.3623	0.0059	0.0028	-0.3342	1.3996
FTNT	0.5743	0.0060	0.0033	-0.5225	4.4085
AZO	0.5039	0.0038	0.0017	-1.3064	13.9337

with this scenario very well, because it heavily relies on the implied assumption of normal distributed return rates. Black-Litterman model could fix this issue to some extent due to the given investor opinions could be highly subjective and free from any specific statistical assumption.

B. Technical Indicators

In the classical CAPM[29] and Factor[30] theories, macroeconomic variables are indispensable and are claimed to be able to interpret the equity return. However, empirical studies suggest that combine fundamental variables and technical indicators is a promising method to forecast equity risk premium and predict market direction[31], [32]. Table IV summaries the technical

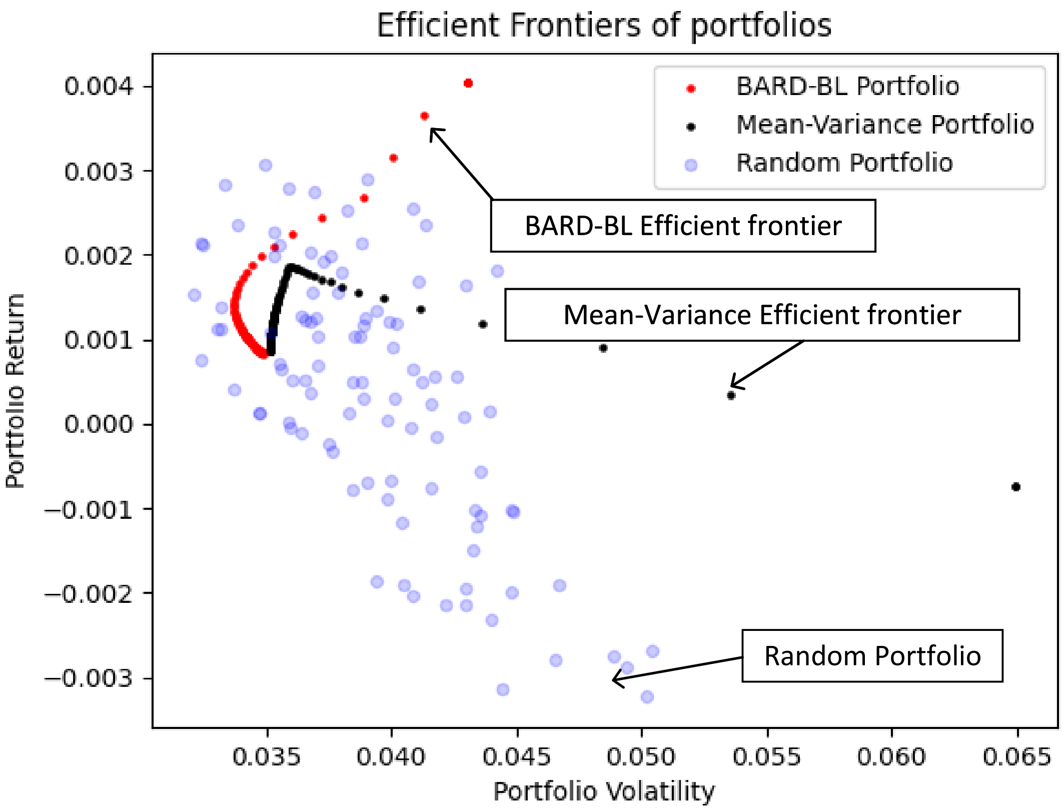

Fig. 8. Efficient Frontiers of the BARD-based Black-Litterman portfolios.

indicators used in this study. Note that the setting of indicator has a certain impact on the modelling effect, readers interested in it can refer to [32].

Besides, the only used macroeconomic variable is the market index, which can be reflected by the S&P price. The theoretical basis for this simplified method is the efficient market hypothesis[33], in which all of the valid information is shown in the change of price. Although such a hypothesis has presented its non-negligible theoretical value, few markets could achieve the ideal condition, and the price sometimes lags behind information.

C. *Investor Views*

To circumvent the issue of data snooping, the investor views are merely generated based on the training set. The resampling technique for the training data is used to simulate the unobservable stock condition in the testing period, that is, we randomly sample an equal amount of data points as the testing set, and take the average over them as the representative observation for generating investor view.

Table V reports the details of the generated investor views, where the training error is measured by the mean-squared-error and uncertainty of an investor view is measured by the forecasting volatility of the stock return. Fig. 1 visualizes the main results shown in Table V, from which we can find that JNJ has the smallest training error while PYPL has the largest training error. As regarding of the forecasting returns, CMG is the most promising stock recommended by GBR, and AAPL ranks second with the forecasting return of 0.0302. On the contrary, JNJ is the least favored stock by GBR, whose forecasting return is -0.0063. Apart from that, V and HD are predicted to show negative returns even though they have positive returns in the training period. In terms of volatility, PYPL exhibits significantly higher forecasting volatility compared to other stocks, whereas JNJ is assigned the highest level of confidence, due to its minimal forecasting volatility of 0.0261.

D. *Confidence Matrix*

The diagonal matrix Ω plays an important role in adjusting the Black-Litterman portfolio weight. To construct the user-specified confidence matrix Ω, the Idzorek's method[27] is used, that is, $T_i = (x_{100\%} - x_{mkt}) \times C_i$, where T_i is the approximate tilt caused by the ith investor view, and C_i is the associated confidence level for the ith investor view, $x_{100\%}$ is the portfolio weight with fully confidence, x_{mkt} is the portfolio weight using the market implied return. And, the corresponding target weight is approximated to $x_{i,\%} \approx x_{mkt} + T_i$.

To build the linkage between the forecasting volatility and the confidence level of the investor view, we use the heuristic method as [15], that is, $c_i = 50\% + \frac{1/v_i}{\sum 1/v_i}$, where v_i is the stock forecasting volatility, c_i is the corresponding confidence level for this stock. The constant 50% represents the naive confidence level for the used machine learning algorithm to generate investor views. In

TABLE IV
TECHNICAL INDICATORS.

Item	Detail	Parameter
EMA	Exponential Moving Average	time period=10
ADX	Average Directional Movement Index	time period=10
ADOSC	Chaikin A/D Oscillator	fast period=5, slow period=10
NATR	Normalized Average True Range	time period=10

TABLE V
INVESTOR VIEWS WITH THE ASSOCIATED UNCERTAINTY.

Portfolio	Ticker	Training Error	Forecasting Volatility	View(Forecasting Return)	Confidence Level
ChatGPT-based	AAPL	0.0002	0.0462	0.0302	60.05%
	GOOG	0.0001	0.0369	0.0001	62.57%
	JNJ	0.0000	0.0261	-0.0063	67.80%
	V	0.0001	0.0347	-0.0044	63.38%
	HD	0.0001	0.0377	-0.0006	62.31%
Bard-based	PYPL	0.0002	0.2636	0.0145	51.76%
	CMG	0.0002	0.0492	0.0333	59.43%
	LULU	0.0002	0.0530	0.0156	58.76%
	FTNT	0.0002	0.0594	0.0170	57.82%
	AZO	0.0001	0.0760	0.0126	56.11%

Table V we present the confidence levels for the investor views about the involved stocks.

E. Portfolio Performance

The two LLM-based portfolios, ChatGPT-BL and BARD-BL are constructed on the training data and tested on the testing data. Also, the two baseline portfolios, EW portfolio and MV portfolio are evaluated for comprehensive comparison.

1) Global Minimum Variance: Achieving the posterior parameters, $\hat{\mu}$ for expected return and $\hat{\Sigma}$ for covariance matrix, one can construct the Markowitz-based portfolios on them. Firstly, we consider the global minimum variance portfolio, which is applauded by the risk-averse investor and only $\hat{\Sigma}$ is used (the objective function is to minimize portfolio variance, that is, $x^T\hat{\Sigma}x$). Table VI summarizes the performances of the involved portfolio models on the testing data, where the annual return, annual volatility, Sharpe Ratio[34], Maxdrawdown, and VaR(5%) are presented. In terms of benchmarks, we give three types of EW and MV, where the all of the recommended stocks, the stocks recommended by ChatGPT, and the stocks recommended by Bard are considered for detailed comparison. In these tested portfolio models, the budget constraint $\mathbf{1}'x = 1$ and the non-shorting constraint $x \geq 0$ are added, and the optimized weights solved by calling Gurobi 10.02 on the 64-bit windows platform are reported in Figs. 2$\sim$6.

According to the results shown in Table VI, the stocks recommended by ChatGPT have the conservative style, where both the ChatGPT-EW (equal weighted portfolio built on the stocks recommended by ChatGPT) and ChatGPT-MV (minimum variance portfolio built on the stocks recommended by ChatGPT) have clear lower annual returns and lower annual volatility than those built on the stocks recommended by Bard, i.e., Bard-EW and Bard-MV, respectively. As the identical strategy to purse minimum variance using the stocks recommended by ChatGPT, ChatGPT-BL has better performance than ChatGPT-MV, due to ChatGPT-BL has higher annual return, higher Sharpe Ratio, while lower maxdrawdown than ChatGPT-MV. Also, BARD-BL has slightly better performance than Bard-MV, since they have the same annual return, while BARD-BL shows a bit lower annual volatility than Bard-MV (0.2533 vs 0.2534). However, when there are more assets could be chosen, the risk level of MV portfolio is decreased, because the annual volatility, maxdrawdown, and VaR(5%) reach to the lowest level among all of the tested portfolios. Significantly, the optimization process could exacerbate the difference between the ChatGPT-based portfolio and the Bard-based portfolio, where the ChatGPT-based portfolio aims to achieve low risk indicators such as annual volatility, maxdrawdown, and VaR(5%), whereas the Bard-based portfolio prefer the high annual return.

From the analysis of portfolio weights given in the Figs.2$\sim$6, the generated investor opinions in Black-Litterman framework could affect the weight distribution to some extent. Although the impact might not be very significant due to it is a single-period model, it would be accumulated and enlarged in the scenario of multi-period model, which would be considered in our future work. The portfolio weight of the MV model considering all of the involved assets suggests that it combines the main stocks recommended by ChatGPT and Bard, but the stocks chosen by ChatGPT (JNJ, GOOG) are the most favored by this investment strategy, which is consistent with the style of minimum variance model.

2) Efficient Frontier: The mean-variance strategy considers the trade-off between risk and return expressed as follows, which can be used to provide the efficient frontier.

$$\begin{aligned} &\max \hat{\mu}^T x - \lambda x^T \hat{\Sigma} x \\ s.t. \quad &\mathbf{1}'x = 1, x \geq 0, \lambda \geq 0 \end{aligned}$$

where λ indicates the coefficient of risk-averse level. In the computational results, we take 200 values of $\lambda \in [0, 100]$ uniformly for illustrating the efficient frontier. Also, we generate 100 fictitious weights randomly and calculate the corresponding return and volatility for comparison. Because the goal of this section is to check the effectiveness of the inputted opinions in the Black-Litterman

TABLE VI
BLACK-LITTERMAN PORTFOLIO CONSIDERING GLOBAL MINIMUM VARIANCE.

Item	Annual Return	Annual Volatility	Sharpe Ratio	Maxdrawdown	VaR(5%)
ChatGPT-BL	0.0021	0.1627	0.0129	18.56%	0.033
BARD-BL	0.0396	0.2533	0.1564	20.86%	0.057
Benchmarks: Use all of the stocks recommended by ChatGPT and Bard					
EW	0.0177	0.2204	0.0803	19.43%	0.048
MV	0.0135	0.1626	0.0828	16.89%	0.033
Benchmarks: Use the stocks recommended by ChatGPT					
EW	0.0161	0.1912	0.0842	19.19%	0.042
MV	0.0018	0.1627	0.0111	18.60%	0.033
Benchmarks: Use the stocks recommended by Bard					
EW	0.0193	0.2730	0.0707	22.93%	0.062
MV	0.0396	0.2534	0.1562	20.86%	0.057

Note: The maxdrawdown(MDD) measures the maximum loss from the peak to the trough, which can be calculated by the formula of MDD $= \frac{V_{\text{trough}} - V_{\text{peak}}}{V_{\text{peak}}}$ and we report the absolute value of MDD here for clear comparison. VaR(%5) measures the tail risk of an investment, the absolute value represents the potential loss for the scenario occurs at the possibility of 5%.

framework, only the out-of-sample data points are used in this numerical experiments. Fig. 7 & Fig. 8 show the efficient frontiers of the portfolios on the ChatGPT-based and Bard-based stocks, respectively.

On the stocks recommended by ChatGPT, the efficient frontier of ChatGPT-BL portfolio locates higher than the classical mean-variance portfolio, suggesting the ChatGPT-BL portfolio could achieve higher return while taking the same risk as the classical mean-variance portfolio on the testing set. However, some random portfolios show better out-of-sample performance than the ChatGPT-BL strategy. BARD-BL portfolio is also superior to the classical mean-variance portfolio regarding to the out-of-sample performance, even, few random portfolios could approach to the BARD-BL efficient frontier when the investment style is aggressive. Hence, using the investor views generating from GBR and GJR-GARCH via Black-Litterman framework in portfolio formation could be supported from the computational results.

VI. CONCLUSIONS

This paper presents a feasible method to build individual portfolio model with the help of some LLMs such as ChatGPT and Bard. From the recommendation results of the designed prompt, ChatGPT acts as a conservative investment robo-advisor, whose preference for selection is mainly for large market value companies. Quite the contrary, Bard favors stocks with relative small market capital while high PE ratio and high β. GBR and GJR-GARCH are also implementable tools for generating investor views, which can be used to calculate the posterior distribution via Black-Litterman framework. Out-of-sample analysis suggests that ChatGPT-BL and BARD-BL could achieve better efficient frontier than the classical MV portfolio to some extent.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the associated professor from the School of Information Management and Engineering, Shanghai University of Finance & Economics, Jianjun Gao, for his professional guidance. The authors also thank the professor from the School of Information Management and Engineering, Shanghai University of Finance & Economics, Dongmei Han, for her kindly help.

REFERENCES

[1] P. P. Ray, "Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, 2023.

[2] S. K. Singh, S. Kumar, and P. S. Mehra, "Chat gpt & google bard ai: A review," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*. IEEE, 2023, pp. 1–6.

[3] S. Wu, M. Koo, L. Blum, A. Black, L. Kao, F. Scalzo, and I. Kurtz, "A comparative study of open-source large language models, gpt-4 and claude 2: Multiple-choice test taking in nephrology," *arXiv preprint arXiv:2308.04709*, 2023.

[4] O. Romanko, A. Narayan, and R. H. Kwon, "Chatgpt-based investment portfolio selection," in *Operations Research Forum*, vol. 4, no. 4. Springer, 2023, pp. 1–27.

[5] J. H. Kim, "What if chatgpt were a quant asset manager," *Finance Research Letters*, vol. 58, p. 104580, 2023.

[6] A. Saggu and L. Ante, "The influence of chatgpt on artificial intelligence related crypto assets: Evidence from a synthetic control analysis," *Finance Research Letters*, p. 103993, 2023.

[7] F. Black and R. Litterman, "Global portfolio optimization," *Financial Analysts Journal*, vol. 48, no. 5, pp. 28–43, 1992.

[8] S. Satchell and A. Scowcroft, "A demystification of the black–litterman model: Managing quantitative and traditional portfolio construction," *Journal of Asset Management*, vol. 1, pp. 138–150, 2000.

[9] A. Meucci, "Beyond black-litterman: Views on non-normal markets," *Available at SSRN 848407*, 2005.

[10] ——, "The black-litterman approach: Original model and extensions," *Shorter version in, THE ENCYCLOPEDIA OF QUANTITATIVE FINANCE, Wiley*, 2010.

[11] W. Cheung, "The black–litterman model explained," *Journal of Asset Management*, vol. 11, pp. 229–243, 2010.

[12] R. O'Toole, "The black–litterman model: active risk targeting and the parameter tau," *Journal of Asset Management*, vol. 18, no. 7, pp. 580–587, 2017.

[13] R. Zemel and T. Pitassi, "A gradient-based boosting algorithm for regression problems," *Advances in Neural Information Processing Systems*, vol. 13, 2000.

[14] X. Ma, R. Yang, D. Zou, and R. Liu, "Measuring extreme risk of sustainable financial system using gjr-garch model trading data-based," *International Journal of Information Management*, vol. 50, pp. 526–537, 2020.

[15] Z. Li, C. Li, L. Min, and D. Lin, "Black-litterman portfolio optimization using gaussian process regression." *IAENG International Journal of Applied Mathematics*, vol. 53, no. 4, pp. 1471–1476, 2023.

[16] L. Min, J. Dong, D. Liu, and X. Kong, "A black-litterman portfolio selection model with investor opinions generating from machine learning algorithms," *Engineering Letters*, vol. 29, no. 2, pp. 710–721, 2021.

[17] P. N. Kolm, G. Ritter, and J. Simonian, "Black–litterman and beyond: The bayesian paradigm in investment management," *The Journal of Portfolio Management*, vol. 47, no. 5, pp. 91–113, 2021.

[18] H. M. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, p. 77, 1952.

[19] L. Min, J. Dong, J. Liu, and X. Gong, “Robust mean-risk portfolio optimization using machine learning-based trade-off parameter,” *Applied Soft Computing*, vol. 113, p. 107948, 2021.

[20] L. Min, Y. Han, and Y. Xiang, “A two-stage robust omega portfolio optimization with cardinality constraints.” *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp. 86–93, 2023.

[21] L. Min, D. Liu, X. Huang, and J. Dong, “Worst-case mean-var portfolio optimization with higher-order moments.” *Engineering Letters*, vol. 30, no. 1, pp. 266–275, 2022.

[22] T. Hai and L. Min, “Hybrid robust portfolio selection model using machine learning-based preselection,” *Engineering Letters*, vol. 29, no. 4, pp. 1626–1635, 2021.

[23] R. Barua and A. K. Sharma, “Dynamic black litterman portfolios with views derived via cnn-bilstm predictions,” *Finance Research Letters*, vol. 49, p. 103111, 2022.

[24] W. Wang, W. Li, N. Zhang, and K. Liu, “Portfolio formation with preselection using deep learning from long-term financial data,” *Expert Systems with Applications*, vol. 143, p. 113042, 2020.

[25] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, “Decision-making for financial trading: A fusion approach of machine learning and portfolio selection,” *Expert Systems with Applications*, vol. 115, pp. 635–655, 2019.

[26] M. Duvinage, P. Mazza, and M. Petitjean, “The intra-day performance of market timing strategies and trading systems based on japanese candlesticks,” *Quantitative Finance*, vol. 13, no. 7, pp. 1059–1070, 2013.

[27] T. Idzorek, “A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels,” in *Forecasting expected returns in the financial markets*. Elsevier, 2007, pp. 17–38.

[28] V. DeMiguel, L. Garlappi, and R. Uppal, “Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?” *The review of Financial studies*, vol. 22, no. 5, pp. 1915–1953, 2009.

[29] H. Levy, “The capm is alive and well: A review and synthesis,” *European Financial Management*, vol. 16, no. 1, pp. 43–71, 2010.

[30] E. F. Fama and K. R. French, “A five-factor asset pricing model,” *Journal of Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[31] C. J. Neely, D. E. Rapach, J. Tu, and G. Zhou, “Forecasting the equity risk premium: the role of technical indicators,” *Management Science*, vol. 60, no. 7, pp. 1772–1791, 2014.

[32] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, “Forecasting price movements using technical indicators: Investigating the impact of varying input window length,” *Neurocomputing*, vol. 264, pp. 71–88, 2017.

[33] A. Timmermann and C. W. Granger, “Efficient market hypothesis and forecasting,” *International Journal of forecasting*, vol. 20, no. 1, pp. 15–27, 2004.

[34] C. L. Israelsen *et al.*, “A refinement to the sharpe ratio and information ratio,” *Journal of Asset Management*, vol. 5, no. 6, pp. 423–427, 2005.