Author(s): Heikkinen, Risto; Karvanen, Juha; Miettinen, Kaisa

Title: A Bayesian model for portfolio decisions based on debiased and regularized expert predictions

Year: 2024

Version: Published version

Journal of Business Economics
https://doi.org/10.1007/s11573-024-01208-5

ORIGINAL PAPER

A Bayesian model for portfolio decisions based on debiased and regularized expert predictions

Risto Heikkinen[1] · **Juha Karvanen**[2] · **Kaisa Miettinen**[1]

Accepted: 17 September 2024

Abstract

Expert predictions of future returns are one source of information for educated stock portfolio decisions. Many models for the mathematical aggregation of expert predictions assume unbiased predictions, but in reality, human predictions tend to include biases, and experts' competence may vary. We propose a Bayesian aggregation model that includes a regularization process to eliminate the influence of experts who have not yet shown competence. The model also includes a debiasing process that fits a linear model to predicted and realized returns. We applied the proposed model to real experts' stock return predictions of 177 companies in the S&P500 index in 37 industries. We assumed that the decision-maker allocates capital between the industry index and the most promising stock within the industry with the Kelly criterion. We also conducted a simulation study to learn the model's performance in different conditions and with larger data. With both the real and simulated data, the proposed model generated higher capital growth than a model that ignores differences between experts. These results indicate the usefulness of regularizing incompetent experts. Compared to an index investor, the capital growth was almost identical with real data but got higher when applied only to industries that were estimated to have multiple competent experts. The simulation study confirmed that more than two competent experts are necessary for the outstanding performance of the presented model.

Keywords Portfolio optimization · Stock returns · Biased judgments · Expertise aggregation · Horseshoe prior · Investing

JEL Classification C11 · G11

✉ Risto Heikkinen
risto.heikkinen@statisti.fi

1 Faculty of Information Technology, University of Jyvaskyla, P.O. Box 35, 40014 University of Jyvaskyla, Finland

2 Department of Mathematics and Statistics, University of Jyvaskyla, P.O. Box 35 (MaD), 40014 University of Jyvaskyla, Finland

Published online: 12 December 2024

1 Introduction

A portfolio is a collection of risky assets that a decision-maker (DM) selects to invest in. For an optimal portfolio allocation, the DM needs the probability distribution of future returns of the assets. Forecasting this distribution may rely on expert information from stock analysts. Analysts publish target prices, which are their predictions of stock prices after some investment horizon, typically 12 months (Hao and Skinner 2023). Predictions on future returns are then obtained by comparing the target prices and the current prices. As these predictions usually differ between the analysts, the DM faces the challenge of forecasting future returns based on multiple expert predictions.

According to the efficient market hypothesis (Fama 1970), all relevant information should be included already in the stock prices, but Grossman and Stiglitz (1980) argue that there has to be some compensation available for the investors who allocate resources in gathering information. The studies about the usefulness of analysts' judgments to investors report inconsistent conclusions. Some studies (Barber et al 2001; Brav and Lehavy 2003; Womack 1996) found that updates in analyst judgments affected the stock prices within a couple of days, and prices showed a slow drift towards the direction of analyst judgment updates during the upcoming months. Boni and Womack (2006) and Da and Schaumburg (2011) observed that it is possible to make abnormal returns based on analyst target prices even without immediate reactions by investing in the companies with the highest expectations within industries. However, Cvitanic et al (2006) found that in 1993–2003, the average of analysts' judgments did not appear to be very valuable, but the usefulness increased when many analysts were analyzing a stock. In addition, Bradshaw et al (2013) found analyst target prices only weakly informative in 2000–2009 as the prediction error was very high. Clement (1999) claimed that there are substantial differences between analysts' abilities due to experience, size of the employer (i.e. company's resources) and the number of stocks they are following (i.e. the more they have to pay attention to different stocks, the less accurate their predictions are).

Human judgements and predictions are prone to biases (Tversky and Kahneman 1974). In a laboratory stock market experiment (Anderson and Sunder 1995), professional traders were less prone to known human biases than students. Regarding actual market observations, McNichols and O'Brien (1997) showed a systematic positive bias in stock analyst predictions and discussed a motivation to analyze and publish favourable judgments on stocks and not publish if judgments change to a negative direction. According to Bradshaw et al (2013), in 2000–2009, target prices indicated on average 15% higher returns than the realizations. These findings support the idea that biases and differences between experts should be handled during the modelling process.

In addition to systematic bias in forecasted future returns, random variation in forecasts is a source of bias in portfolio decisions, often referred to as optimizer's curse (see Smith and Winkler 2006). More specifically, stocks with the highest (lowest) return forecasts tend to have positive (negative) random errors in the forecast, even though random errors of all assets are expected to sum up to zero. Because only the stocks with the highest return potential (which tends to be partly caused by positive random error) are selected for a portfolio, the optimizer's curse leads to a systemically lower portfolio performance than was expected at the decision time. Thus, more accurate forecasts are valuable.

The traditional stock portfolio model with expert predictions, the Black-Litterman model (Black and Litterman 1992), assumes that all expert predictions are well calibrated without systematic biases, and a DM must quantify the competence of each expert a priori manually. Later, the Black-Litterman model was generalized by Chen and Lim (2020), who included the possibility for systematic additive bias in expert predictions. However, the generalized Black-Litterman model does not consider systematic over/under-reactions of the experts, i.e., it assumes that the regression coefficient of the predicted returns and actual returns equals one. Black-Litterman models are based on the assumption of normally distributed stock returns. However, a normal distribution has shown to be a weakly performing model with fat-tailed stock prices (see, e.g., Hu and Kercheval 2010).

In this paper, we focus on improving the following gaps in the existing models:

1) crude assumptions about expert biases,
2) challenges in quantifying the informativeness of different experts and
3) strict assumptions about stock return distribution.

We introduce a Bayesian model for aggregating (i.e. combining) experts' predictions to support stock portfolio decisions. We call the model a selective Bayesian expert debiasing (SBEDE) portfolio model. The core of the SBEDE model is a linear regression fit between an expert's past predictions and reality. The SBEDE model can be used when multiple experts produce point predictions of future returns and data about the historical accuracy of the experts is available. Compared to earlier portfolio models, a more flexible debiasing process and an automatic selection of competent experts are the main improvements. The SBEDE model includes an automatic regularization process that eliminates the influence of experts who have not yet proved their competence. The process is executed with a horseshoe prior distribution (Carvalho and Polson 2010; Piironen and Vehtari 2017). This is a major improvement compared to manually quantifying confidence towards different experts like in the Black-Litterman model or doing a separate expert classification study like with the Cooke's method (Cooke 1991).

The SBEDE model has debiasing elements similar to recently published expert models for various forecasting purposes but not specifically stock markets. These are the BIN (Bias, Information, Noise) model by Satopää et al (2021) for a binary target variable and the model by Merkle et al (2020) (Merkle model) for a continuous target variable. SBEDE includes parameters for systematic bias ('shifting bias' in the Merkle model), responsiveness to reality ('information' in the BIN model and 'scaling bias' in the Merkle model) and inaccuracy ('noise' in the BIN model and 'incompetence' in the Merkle model). One can consider SBEDE as an extension of the Merkle model for portfolio optimization purposes with the following additions: 1) horseshoe prior or hierarchical structure for parameters for avoiding identification or overfitting problems with a limited amount of data, 2) correlations between experts and 3) formulation that allows fat-tailed unpredictable shocks.

As experts tend to have a poor performance in predicting fat-tailed economical fluctuations (Makridakis and Taleb 2009), we estimate experts' abilities to predict an expected value of a return distribution and treat the modeling of unpredictable shocks as a separate problem. One benefit of the proposed SBEDE model is that it can be applied together with many different asset pricing models. For demonstration purposes, we apply SBEDE with an asset pricing model that allows fat-tailed t-distributed random shocks.

The contributions of the paper are:

- proposing a novel expert prediction aggregation model SBEDE that includes automated expert selection by regularizing the influence of the experts without convincing history and a flexible debiasing process,
- as a proof of concept, showing the potential of the SBEDE model with real data of Standard and Poor's 500 (S&P500) companies, considering forecasting accuracy and decision-making performance and
- conducting a simulation study for better understanding of the conditions where the SBEDE model is useful.

Our application uses stock analyst target prices for 177 companies in the S&P500 index in 37 industries. For simplicity, we assume that the DM is risk-neutral for short-term volatility and aims to maximize the expected capital growth rate in the long term. This strategy is called the Kelly criterion (Kelly 1956; Thorp 1971; Peters 2011). Nevertheless, the SBEDE model could also be applied together with other investment strategies, e.g. "mean-variance" (Markowitz 1952) as we do for comparison in the simulation study.

We measure the benefit of the SBEDE model for decision-making with realized capital growth of optimized portfolios during the year. For the sake of the model comparison, the portfolio decisions are made separately for 37 industries. In the application, we compare our model's performance to an index investor's (who allocates capital equally to every stock within an industry) performance. If the efficient market hypothesis (Fama 1970) is correct, portfolio decisions based on analyst target prices should not add any value compared to an index investor. For studying the value of modelling differences between analysts, we also compare the performance of the model to a simplified Bayesian model, called an exchangeable expert (EE) model, where all experts are assumed to be equally competent.

The rest of the paper is organized as follows: we first introduce notations, the background of expert prediction aggregation methods and details of Kelly portfolio optimization in Sect. 2. After that, we introduce our SBEDE model and a separate asset price model in Sect. 3. In Sect. 4, we describe our practical application and the data used. We present the empirical results in Sect. 5 and the simulation results in Sect. 6. The results are discussed in Sect. 7 and the conclusions are presented in Sect. 8.

2 Background

This section introduces notations and our assumptions about portfolio optimization. It also presents earlier expert prediction aggregation methods and Kelly portfolio optimization.

2.1 Notations

We consider a DM that allocates capital between N stocks in one industry that are followed by J stock analysts (experts). We assume that the DM aims to invest in the most promising (the highest expected capital growth rate) stock within the industry but can add diversification by allocating part of the capital to the industry index, which weighs all N stocks within the industry equally. By selecting only one stock, DM does not have to model a correlation structure between individual stocks, and index funds are economical tools for diversification in practice. Point predictions of future returns are derived from experts' target prices.

Historical data are available from time points $t \in \{1, 2, \ldots, H\}$ that present ends of quarters. A quarter, i.e., a period between consecutive time points $(t;t+1]$, is denoted by its starting point t within brackets as $[t]$. We assume that at time point H, a DM forecasts stock returns for a one-year horizon, including four quarters. This multiple-period return is denoted by a subscript $[H, H+4]$, including the starting and ending points. The DM utilizes information from target prices from an expert j ($j \in \{1, 2, \ldots, J\}$) for a stock i ($i \in \{1, 2, \ldots, N\}$), available at the time point H, denoted by $\text{TargetPrice}_{ij(H)}$. The target prices are assumed to be predictions of the stock prices at the time point $H+4$. All stock returns are transferred to a logarithmic scale. Table 1 shows the notations used within this study.

Table 1 Notations and their definitions used in this paper

Notation	Definition
$x_{i[t,t+4]}$	$\log\left(\frac{\text{MarketPrice}_{i(t+4)}}{\text{MarketPrice}_{i(t)}}\right)$, return of stock i, based on price change during $(t;t+4]$
$\mu_{i[t,t+4]}$	$\mathbf{E}(x_{i[t,t+4]})$, expected return for period $(t;t+4]$ with the information available at the time point t
$y_{i[t]}$	The total return of stock i including price change and the dividend during time period $[t]$
$y_{0[t]}$	The total return of industry index during time period $[t]$
w_i	The weight in the portfolio for stock $i (i \in \{1, 2, \ldots, N\})$
w_0	The weight in the portfolio for the related industry index
H	The time point when a decision on the weights w_i of the different assets in the portfolio is made
$M_{ij[t,t+4]}$	$\log(\text{TargetPrice}_{ij(t)}/\text{MarketPrice}_{i(t)})$, expert j's point prediction of $x_{i[t,t+4]}$, i.e. target price implied expected return
b_{ij}	The systematic additive bias of expert j analyzing stock i, so that $\mu_{i[t,t+4]} = 0$ implies $\mathbf{E}(M_{ji[t,t+4]}) = b_{ij}$

2.2 Expert prediction aggregation

Here, we briefly summarize earlier multiple expert aggregation models and findings that are the most relevant to our model but are not specifically designed for stock markets. Clemen and Winkler (1999) reviewed expert prediction aggregation methods in risk analysis and concluded that the best forecasts had been achieved by combining judgments from multiple experts. According to them, weighting every expert's predictions equally in the aggregate forecast tended to perform quite well. Later, Cooke and Goossens (2008) reviewed results from 45 expert panels in different areas and showed a clear advantage of setting a higher weight to experts who have performed well in the past. They used Cooke's method (Cooke 1991), where the experts' weights are based on seed questions and a scoring rule. However, the method was criticized by O'Hagan et al (2006) because it does not take into account the correlations between the experts.

The predictions from different experts may be correlated because of the same information sources and experts can read each other's analyses. Clemen and Winkler (1985) studied the effect of correlated experts and later experimented in practice (Winkler and Clemen 2004) that after a couple of good experts, the value of adding new predictions for the same problem diminishes. Jaspersen (2022) demonstrated that the averaging of the predictions may put too much weight on some widely used information source and introduced a method for sophisticated weighting of the experts.

The main tool for aggregating multiple experts' predictions is the multivariate normal model (see, e.g., Clemen and Winkler 1985)

$$\mathbf{M}_{[H]} \sim \mathrm{MVN}(x_{[H]}, \Sigma), \tag{1}$$

where $\mathbf{M}_{[H]}$ is a vector of experts' point predictions of an unknown value of the target variable $x_{[H]}$. A covariance matrix Σ includes experts' individual variances and covariances between experts. The experts' performance in predicting target variable values earlier in history, $[x_{[1]}, \dots, x_{[H-1]}]$, can be used for estimating the parameters in this model.

The methods mentioned above do not consider systematic biases or model different characteristics of the experts. On the contrary Merkle et al (2020) presented a model for the characteristics of different experts, influenced by Cultural Consensus Theory (Batchelder and Romney 1988), as

$$M_{[H]} \sim \mathrm{N}(b_j + \phi_j x_{[H]}, \sigma_j^2), \tag{2}$$

where b_j stands for expert's individual systematic bias, ϕ_j for individual "scaling bias" and σ_j^2 for individual variance. The model does not consider correlations between experts.

2.3 Kelly portfolio optimization

In this paper, the DM optimizes the portfolio using the Kelly criterion as an objective function. The Kelly criterion (Kelly 1956) maximizes investor's long-term logarithmic growth rate

$$\lim_{T\to\infty} \frac{1}{T} \log \frac{V_{(H+T)}}{V_{(H)}},$$

where V is the investor's capital, H is the decision time, and T is the number of time points in the investment horizon. Later, Breiman (1961) and Thorp (1971) showed that the recipe for reaching this criterion with repeated investment decisions is to maximize the expected value of logarithmic capital at the next time point $\mathbf{E}[\log(V_{(H+1)})]$.

The Kelly criterion is especially relevant with an infinite investment horizon (relevant assumption, e.g., for foundations). The strategy based on $\mathbf{E}[\log(V_{(H+1)})]$ recipe also has good properties in the shorter term, such as avoiding losing the whole capital in bankruptcy or minimizing time to reach some preassigned goal for capital size (see, e.g., Thorp 1971). Because an investor's capital growth is not an ergodic process, Peters (2011) suggests using the Kelly criterion as an investment object instead of the arithmetic mean of the future scenario distribution for the portfolio. The main disadvantage of the Kelly criterion is that, in the short-term, it is a too volatile strategy for a risk-averse investor (MacLean et al 2011). However, if the logarithmic utility function describes investor's risk preferences, the Kelly criterion is optimal also as a short-term strategy (Thorp 1971).

In this paper, a portfolio consists of two risky assets, namely an index fund and one individual stock i and the portfolio is assumed to be updated once a year. As presented earlier, next year's unknown total returns (including dividends) of these assets are denoted by $y_{0[H,H+4])}$ and $y_{i[H,H+4])}$, respectively, and the vector of these returns is $\mathbf{y}_{[H,H+4]}$. Decision variables are weights w_0 and w_i. Because $w_0 + w_i = 1$, we have $w_0 = 1 - w_i$. The decision of the value for w_i is based on maximizing the expected value of logarithmic capital with a one-year investment horizon (recipe for maximizing long-term capital growth). We have $\log(V_{(H+4)}(w_i)) = \log(V_{(H)}) + g_{[H,H+4])}$, where $g_{[H,H+4])}$ is a logarithmic growth rate during the next year. As the current capital $V_{(H)}$ is deterministic, we have to maximize the expected value for

$$g_{[H,H+4]}(w_i, \mathbf{y}_{[H,H+4]}) = \log[w_i e^{y_{i[H,H+4]}} + (1 - w_i) e^{y_{0[H,H+4]}}]. \tag{3}$$

3 Proposed models

Next, we introduce the main contribution of this study, the SBEDE model for stock portfolio optimization. To study the value of modeling the differences between experts, we also introduce the simplified version, the EE model, where all experts are assumed to be similar.

3.1 Arrangement

As mentioned earlier, we assume that the experts set their target prices based on expected price changes within a one-year investment horizon. Dividends are not included but are considered separately in the portfolio optimization phase. We also assume that the experts update their judgments at least once a quarter, as companies tend to report quarterly. We use quarterly measured data, where the target variable is the return during four quarters. We model the returns with a modification of the single index asset model (see, Elton et al 2014), where the time dependence of annual returns, measured quarterly, is modeled with unpredictable quarterly shocks. The effect of these shocks is assumed to be similar to each company within an industry. The shocks are modeled by a t-distribution (probably fat-tailed) as suggested by Hu and Kercheval (2010) and Praetz (1972).

The asset and expert models are connected with the annual expected return $\mu_{[t,t+3]}$, and one may use any asset price model where a parameter comparable to $\mu_{[t,t+3]}$ is present. For example, one can use CAPM (Capital Asset Pricing Model; Sharpe 1964) for modelling $\mu_{[t,t+3]}$ with a risk-free rate and stock's sensitivity to the market risk premium. We are only interested in the experts' ability to predict future returns and not in the features of stocks that CAPM estimates.

The challenge with a flexible expert model is that estimating a high number of parameters in the model needs a lot of data. However, it is common that there is little data available about past expert predictions. The first tool to solve a small data challenge is a hierarchical structure of the parameters, so that information about the experts' characteristics is shared. This structure reduces overfitting when data about individuals is limited (Gelman et al 2013). Hartley and French (2021) used a hierarchical model for estimating experts' biases, but instead of point predictions, they modeled judgements given as quantiles. The second tool for avoiding overfitting is to regularize regression parameters towards zero if there is not much evidence of deviating from zero (see James et al 2021). Regularization is advantageous if it is known that only a proportion of the experts are competent enough, but we cannot identify them a priori.

3.2 Asset model

With quarterly measured data, the annual (four quarters) return of asset i in one selected industry starting at the beginning of a time point t is modeled as a sum of predictable expected value and four unpredictable shocks. The model is formulated as

$$\begin{aligned} x_{i[t,t+4]} &= \mu_{i[t,t+4]} + \sum_{k=t}^{t+3}(\eta_{[k]} + \epsilon_{i[k]}), \text{ where } i \in 1,2,\dots,I \text{ and } t \in 1,2,\dots,H \\ \eta_{[t]} &\sim t_{\nu_0}(0,\psi_0) \text{ and } \epsilon_{i[t]} \sim t_{\nu}(0,\psi_i), \end{aligned} \tag{4}$$

where $\eta_{[t]}$ stands for an unpredictable industry shock for the considered industry and $\epsilon_{i[t]}$ is an asset-specific shock at the quarter t. Asset-specific scale parameters ψ_i have a hierarchical structure

$$\psi_i = \frac{\psi^*}{\sqrt{c_i}}, \quad c_i \sim Gamma\left(\frac{1}{\omega^2}, \frac{1}{\omega^2}\right),$$

where ψ^* is an average scale across assets, and the random effect coefficients c_i have an expected value of one and variance ω^2. The gamma distribution is selected because it is flexible with positive support ($c_i > 0$). With this selection, ψ^2 follows an inverse gamma distribution, which has a history as a justifiable distribution for heterogeneous variance (see Praetz 1972). Degrees of freedom parameters, ν_0 for industry shocks and ν for asset-specific shocks, define the fat-tailness of a total return distribution. Prior distributions are needed for parameters $\mu_{i[t,t+4]}$, ψ_0, ψ^*, ν_0, ν and ω. Our selections for this study's application will be introduced in Sect. 4.2 and Appendix B.

3.3 Selective Bayesian expert debiasing model

By definition, experts cannot predict unpredictable shocks. In addition, a multivariate normal model for return predictions would not be justified if the prediction errors include fat-tailed shocks. Thus, unpredictable shocks are not included in the expert model, and the connection between the asset model and the expert model is based on the latent expected annual return $\mu_{i[t]}$ (see (4)). This is one difference compared to the Black-Litterman model, the multivariate normal aggregation model (1) and the Merkle model (2), where the expected values of the expert predictions are functions of $x_{i[t,t+4]}$. Here we introduce the full SBEDE model as

$$\mathbf{M}_{i[t,t+4]} \sim \mathrm{MVN}(\mathbf{b}_i + \boldsymbol{\phi}\mu_{i[t,t+4]}, \Sigma),$$

where parameters $\boldsymbol{\phi}$, $\mathbf{b}_i$ and Σ are explained below.

The SBEDE model has a parameter vector $\boldsymbol{\phi}$ to model experts' responsiveness to reality. The responsiveness parameters ϕ_j in the vector $\boldsymbol{\phi}$ can be used for classifying experts as follows: over-responsive if $\phi_j \gg 1$, balanced if $\phi_j \approx 1$, under-responsive if $0 \ll \phi_j \ll 1$, uninformative if $\phi_j \approx 0$ and contra-indicative if $\phi_j \ll 0$. This formulation allows us to ignore uninformative experts in future predictions by using a regularizing prior. For this purpose, a horseshoe prior (Piironen and Vehtari 2017) is used for the responsiveness parameter as

$$\begin{aligned} \phi_j | \lambda_j, \tau &\sim \mathrm{N}(0, \tau^2\lambda_j^2) \\ \lambda_j &\sim \mathrm{C}^+(0,1), j \in 1, \dots, J, \end{aligned}$$

where C^+ is a half-Cauchy distribution and τ is a global shrinkage factor. A prior distribution for τ has to be set based on expectations about the number of informative experts (ϕ_j differs significantly from zero). The details are demonstrated in Appendix B.

For biases b_{ij} in vector $\mathbf{b}_i$ we set a prior distribution $b_{ij} \sim N(\kappa, \xi^2)$ where κ is the average bias across all experts and all assets and ξ is a prior view of the standard deviation. Based on earlier studies (e.g., McNichols and O'Brien 1997), κ is probably something positive, but the DM can use his expertise to set a prior for it. The prior distribution selections for our application are introduced in Sect. 4.2 and Appendix B.

The covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_J \end{bmatrix} \Omega \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_J \end{bmatrix}$$

includes inaccuracy parameters (standard deviation of a noise, σ_j) for each expert j, and Ω is the correlation matrix of the experts. In addition, parameters σ_j have a hierarchical structure as

$$\sigma_j = \frac{\sigma^*}{\sqrt{d_j}}, \quad d_j \sim Gamma\left(\frac{1}{\iota^2}, \frac{1}{\iota^2}\right), \tag{5}$$

where σ^* is an average inaccuracy and the random effect coefficients d_j have an expected value 1 and a variance ι^2. Parameters σ^* and Ω need prior distributions.

For illustration purposes, Fig. 1 shows simulated return predictions from an expert with $b = 0.1$, $\phi = 0.5$ and $\sigma = 0.075$ (solid line). One interpretation for the bias b is "expert's expected return prediction when the real expected return is zero". Because the expert is not balanced ($\phi \neq 1$), the SBEDE model with the responsive parameter is needed for debiasing. A directed acyclic graph in Fig. 2 summarizes the structure of the SBEDE model. The key elements of the experts' return predictions for stock i, predictable reality $\mu_{i[H,H+4]}$, additive biases $\mathbf{b}_i$, responsivenesses $\boldsymbol{\phi}$, in-accuracies $\boldsymbol{\sigma}$ and correlations between experts Ω, are at the third row in the figure.

3.4 Exchangeable expert model

For estimating the value of modeling differences between experts, we introduce a simplified model, where all experts are assumed to be similar. We assume the same systematic bias, sensitivity and inaccuracy for all experts, i.e., $b_{ij} := b_i$ (all experts have the same bias when analyzing company i), $\phi_j := \phi^*$ and $\sigma_j = \sigma^*$ (compare with (5)) for all $j \in 1, 2, \ldots, J$. We also assume that there is one common correlation coefficient ρ between all expert. The model is formulated as

$$\mathbf{M}_{i[t,t+4]} \sim MVN(\mathbf{b}_i + \phi^* \mu_{i[t,t+4]}, \Sigma),$$

where

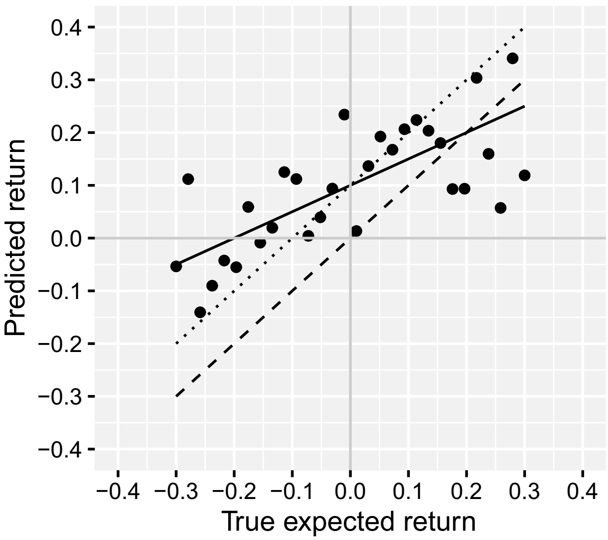

Fig. 1 Illustration of the advantages of the SBEDE model compared to the Black-Litterman models. The dots represent simulated pairs of expected returns (not observable in reality) and the predicted returns of an expert. The solid line is the expert's personal "regression line" (the SBEDE model estimates), and the dashed line is the "regression line" of a perfectly calibrated expert (as the original Black-Litterman model assumes) that is rare in reality. The dotted line estimates the additive bias but assumes the regression coefficient of one (generalized Black-Litterman model estimates)

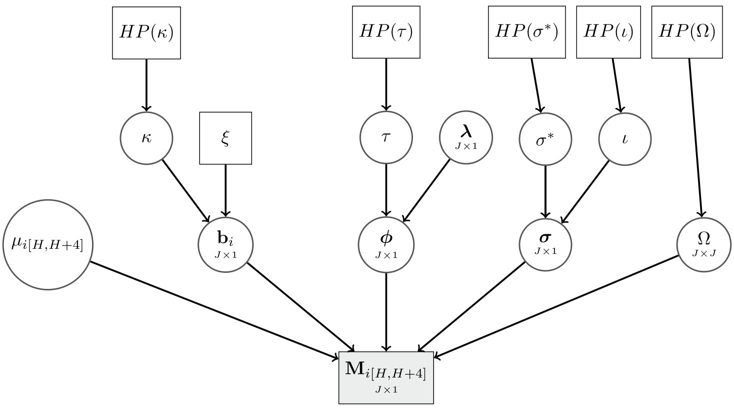

Fig. 2 A directed acyclic graph for visualizing parameters in the SBEDE model. Circles are unknown parameters, and boxes are hyperparameters whose values the DM has to set (white background) or observed data (grey background). The abbreviation HP refers to the hyperparameters of the parameter in parentheses. They depend on prior distribution selections. Parameter $\mu_{i[H,H+4]}$ is modeled with a separate asset model

$$\Sigma = \underset{J\times J}{\begin{bmatrix} \sigma^* & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^* \end{bmatrix}} \underset{J\times J}{\begin{bmatrix} 1 & \rho & \rho \\ \rho & \ddots & \rho \\ \rho & \rho & 1 \end{bmatrix}} \underset{J\times J}{\begin{bmatrix} \sigma^* & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^* \end{bmatrix}}.$$

Instead of a horseshoe prior, the prior distribution has to be set only for the common responsiveness ϕ^*, and the correlation matrix includes only one stochastic parameter ρ. The prior distributions used in our application are presented in Appendix B.

4 Application to S&P500 data

To demonstrate the models presented in Sect. 3, we apply them with real data. The aim is to compare the performances of the models to each other and to an index investor's performance. First, we describe the data used. After that, we present prior distributions and other practical selections for Bayesian modeling, and, finally, define the optimization process for portfolio decisions.

4.1 Data

We collected historical stock market prices and experts' target prices of companies listed in the S&P500 index on 2021–09-30 from the Refinitiv-Eikon database[1] (Institutional Brokers' Estimate System, IBES). Individual experts' target price data was very sparse before the mid-2010s, and that caused limitations to the time scale of the analysis. As mentioned earlier, we assume that experts usually update their judgments at least once a quarter. Nevertheless, some experts had long periods without any updates (or without sending the updates to the database), and we decided to consider target prices on idle periods as missing data. The presented Bayesian models have many parameters, and for reliable estimation, there should be multiple observations across stocks and experts. All these considerations lead to the following data collecting and filtering process before analysis.

1. Set a date for $t = 1$: 2013–12-31 (beginning of the relevant history).
2. Set a date for $H = 29$: 2020–12-31 (decision time, so that performance in the year 2021 is used for comparing models).
3. Exclude target prices of idle experts. Definition of an idle expert j at a time point t for the company i is the following:
    - time lapsed after expert j's last update of company i is more than 185 days (2 quarters + flexibility for release delays) and time until the next update is more than 185 days (2 quarters + flexibility for release delays)

[1] Refinitiv-Eikon Datastream, 2022.

- or time lapsed after expert j's last update of company i is more than 370 days (a year + flexibility for release delays).

4. Exclude companies that do not have full stock price history and at least one non-idle expert target price starting from $t = 1$.
5. Collect quarterly data of stock prices and non-idle expert target prices from the time $t = 1$ until $t = H$ and stock prices at $t = H + 4$ (a year after the decision for measuring performance). These time points define the ends of the quarter.
6. Set the minimum sample size for a company+expert-specific target price history: 16 (= 4 years), and exclude target histories where a number of qualified observations is less.
7. Set the minimum number for qualified experts in decision date: $\min(J_{i(H)}) = 3$, and exclude companies with $J_{i(H)} < 3$. (There needs to be multiple experts so that prediction aggregation and estimating differences between experts make sense.)
8. Set the minimum number of companies in one industry: $\min(N) = 3$, and exclude industries where $N < \min(N)$. (There needs to be multiple companies in the industry index to have options for decision-making and to get diversification by investing in the index.)

Because of limited data, we did not set the thresholds in steps 6–8 higher than what is necessary for making a sensible analysis.

The classification of companies into industries is based on the Global Industry Classification Standard[2] sub-industries. After filtering, 177 companies in 37 industries were left for analysis. The industries are listed in Table 5 in Appendix A with the number of qualified companies and experts, followed by descriptive statistics of them in Table 6. The industry index returns are calculated as an equally weighted mean of all stocks in the industry qualified for analysis. For the 177 companies left for the analysis, we also collected the dividend histories from Nasdaq (2022) for forecasting returns with dividends $y_{i[t,t+4]}$ in the optimization phase in Sect. 4.3. Dividends are ignored in modelling experts' predictions because target prices do not include dividends. The relation between target prices and dividends is analyzed further by Hao and Skinner (2023).

4.2 Bayesian modelling

We applied the SBEDE and the EE models, introduced in Sect. 3, with data presented in Sect. 4.1. We combined these models with an asset model specified in Sect. 3.2. Next, we present the steps to estimate a posterior distribution of the parameters.

A prior for expected return is expressed as $\mu_{i[t,t+4]} \sim \mathrm{N}(r_{(t)} + 0.04, 0.1^2)$, where $r_{(t)}$ is the risk-free rate (52 weeks US treasury bill rate) at the time point t. The point estimate for a current risk premium, 0.04, is based on the average difference between the S&P500 index returns (without dividends) and short-term interest rates from

[2] https://www.msci.com/our-solutions/indexes/gics.

2000 to 2020. In addition, 0.1^2 describes our prior uncertainty and variation between industries. We also tested the model with different selections for this parameter (see Sect. 5.3). In a real modeling situation, a DM with knowledge of a specific industry can use a more informative prior distribution to reduce the optimizer's curse, as mentioned in the introduction. Other prior distribution selections and details of applying the MCMC algorithm for fitting the models are presented in Appendix B.

We also tried to fit the Merkle model as a reference model, but it had significant convergence problems (20 of 37 industries did not meet the convergence criteria), possibly due to the sparsity of the data. Thus, we left it out of the analysis.

The result of the modelling is $S = 60000$ simulated values from the posterior distribution of each parameter in the model. The most interesting parameters for the decision-making process are next year returns $\mathbf{x}_{[H,H+4]}$ and responsiveness parameters $\boldsymbol{\phi}$. As an example of regularizing experts, Fig. 3 illustrates the historical performance of the selected different types of experts in one industry and how the performance affects the estimated posterior distribution of $\boldsymbol{\phi}$. Figure shows three experts analyzing five companies. The posterior distributions show that Expert 3 has enough evidence to be labelled informative but under-responsive as ϕ_3 is estimated to be around 0.6–0.8. Based on regression lines in the scatter plot, the expert has performed well in analyzing both companies (Company 2 and Company 6). Expert 2 can most likely be labelled as uninformative ($\phi_j \approx 0$), but there is a small probability of being informative based on the heavy tail of the posterior distribution. Based on regression lines, after excluding positive systematic bias, Expert 2 has performed tolerable in analyzing Company 1 but not in Company 3. Expert 10 can be clearly

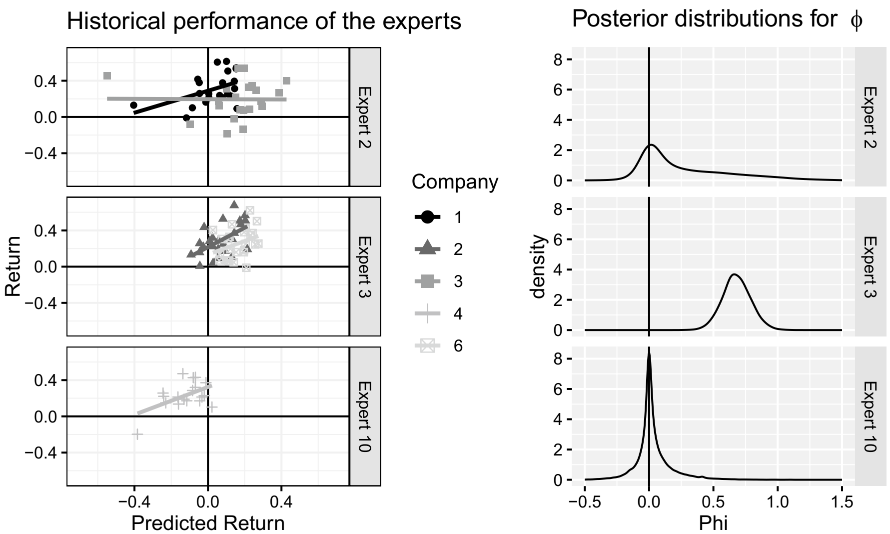

Fig. 3 Historical performance (predictions versus realized returns) of three experts chosen for the illustration and posterior distributions of responsiveness parameters for one example industry. On the left, each facet represents the performance of one expert (realized returns include also unpredictable shocks). Different greyscales and shapes present companies. Regression lines are fitted for each company. Posterior distributions for responsiveness parameters $\boldsymbol{\phi}$ based on the SBEDE model are on the right

labelled as uninformative based on the posterior distribution of ϕ_{10}. At first look, an increasing regression line suggests a good performance of Expert 10. However, this line is influenced by a single outlying observation. The asset model considers extreme returns mainly as t-distributed unpredictable shocks.

4.3 Optimization process

In the asset model presented in Sect. 3.2, we have modeled the dependency between stocks with the industry shock parameter common to all stocks within the industry in question. We did not model pairwise dependencies because the aim is to allocate capital between two assets: one stock with high expectations and an industry index (equally weighted mean return). The DM of the study has to optimize the proportion of capital invested in the most attractive stock and how much to diversify capital to the industry index.

Since the expert target prices do not include dividends, the DM adds last year's dividends as dividend estimates for the following year. The portfolio will be updated once a year; thus, we are considering only the next year's returns. Taxes and trading or holding costs are ignored for simplification. Following the Kelly criterion (3) as the investment strategy, the optimization problem is defined as follows:

$$\text{maximize } \mathbf{E}g_{[H,H+4]}(w_1,\dots,w_N,\mathbf{y}_{[H,H+4]}) \tag{6a}$$

$$\text{subject to } w_i w_k = 0, \text{for all } i \neq k \tag{6b}$$

$$\sum_{i=1}^{N} w_i \leq 1 \tag{6c}$$

$$w_1,\dots,w_N \geq 0 \tag{6d}$$

and

$$g_{[H,H+4]}(w_1,\dots,w_N,\mathbf{y}_{[H,H+4]}) = \log\left[\sum_{i=1}^{N} w_i e^{y_{i[H,H+4]}} + \left(1-\sum_{i=1}^{N} w_i\right)e^{y_{0[H,H+4]}}\right], \tag{7}$$

where the vector of unknown future returns is

$$\mathbf{y}_{[H,H+4]} = [y_{0[H,H+4]}, y_{1[H,H+4]},\dots,y_{N[H,H+4]}].$$

The equality constraint (6b) ensures that only one stock will be selected for the portfolio. Thus, the sums in (7) include only one element that differs from zero.

As mentioned in Sect. 4.2, there are $S = 60000$ simulations of $\mathbf{x}_{[H,H+4]}$ based on a Bayesian model. A simulated realization of $\mathbf{x}_{[H,H+4]}$ from round s is denoted by $\tilde{\mathbf{x}}^s_{[H,H+4]}$. The DM adds dividend estimates to obtain $\tilde{\mathbf{y}}^s_{[H,H+4]}$, where $\tilde{y}^s_{i[H,H+4]} = \log(e^{\tilde{x}^s_{i[H,H+4]}} + D_{i[H-4,H]})$ and $D_{i[H-4,H]}$ is a dividend percentage of stock

i during the latest period prior decision making. The expected value in (6a) can be calculated based on these simulations as (see Gelman et al 2013)

$$\mathbf{E}g_{[H,H+4]}(w_1,\ldots,w_N,\mathbf{y}_{[H,H+4]}) \approx \frac{1}{S}\sum_{s=1}^{S} g_{[H,H+4]}\left(w_1,\ldots,w_N,\tilde{\mathbf{y}}^{s}_{[H,H+4]}\right).$$

The optimization process was conducted with a hybrid of two methods to find a global optimum.

1. Global step: Initial solutions are searched with the differential evolution method (Ardia et al 2011), which has been designed for global optimization.
2. Local step: Final solutions are found with the L-BFGS-B method (Byrd et al 1995) using solutions of step 1) as initial values. L-BFGS-B is a local method but faster (compared to differential evolution) and can guarantee local optimality.

This optimization process was repeated with all 37 industries using information available at the end of the year 2020. Both models, SBEDE and EE, were utilized separately; thus, the process was repeated 74 times.

5 Empirical results

We measure the performance of the forecasts and the decisions made on 2020–12-31 with the real stock returns during 2021. The decisions are presented in Table 9 in Appendix C. We compare the forecasting accuracy of the models in Sect. 5.1 and the performance of the allocation decision in Sect. 5.2. We do sensitivity analyses and compare the performance in Sect. 5.3. Finally, we discuss these results in Sect. 7.

5.1 Forecasting accuracy

In this section, we compare the future return forecasting accuracy in 37 industries based on (1) the SBEDE model, (2) the EE model, and (3) the median of experts' return predictions, which is called a raw consensus forecast. The performance of the models is quantified with a mean absolute error (MAE), measuring the accuracy of the point forecast, and a coverage probability (CP), measuring how well the posterior distributions quantify uncertainty. Here, $x_{i[H,H+4]}$ is a realized logarithmic annual return (without dividends) of company i starting at the time point H, $\tilde{x}^{q}_{i[H,H+4]}$ is a qth quantile of an estimated posterior distribution, and I is an indicator function with values 0 and 1. The performance metrics are defined as

1) MAE, where posterior median $\tilde{x}^{0.5}_{i[H,H+4]}$ is used as a point forecast of the parameter $x_{i[H,H+4]}$. The absolute error AE is defined as $|\tilde{x}^{0.5}_{i[H,H+4]} - x_{i[H,H+4]}|$. The mean is taken over AEs of all 177 companies in 37 industries. Here, a smaller value is better.
2) CP, where coverages C are defined as

$$C^{50} = I[x_{i[H,H+4]} \in (\tilde{x}^{0.25}_{i[H,H+4]}, \tilde{x}^{0.75}_{i[H,H+4]})],$$
$$C^{68} = I[x_{i[H,H+4]} \in (\tilde{x}^{0.16}_{i[H,H+4]}, \tilde{x}^{0.84}_{i[H,H+4]})],$$
$$C^{90} = I[x_{i[H,H+4]} \in (\tilde{x}^{0.05}_{i[H,H+4]}, \tilde{x}^{0.95}_{i[H,H+4]})] \text{and}$$
$$C^{95} = I[x_{i[H,H+4]} \in (\tilde{x}^{0.025}_{i[H,H+4]}, \tilde{x}^{0.975}_{i[H,H+4]})].$$

The average is taken over coverages of all 177 companies in 37 industries, and, e.g., CP 90 measures the proportion of the estimated 90% Bayesian credible intervals that covered the actual observations. If CP 90 markedly differs from 90%, this is a sign of failed estimation of uncertainty.

Table 2 summarizes the forecasting accuracy of the SBEDE models and the raw consensus. It shows that the MAE of the point forecasts (posterior medians or median of the raw forecasts) of the SBEDE and EE models are almost identical and worse than the raw consensus. As overall market returns strongly influence MAE, the table also shows MAE Premium, which is based on predicted returns over the index (predicted with the same method). Based on this metric, SBEDE is the most accurate, and the raw consensus is the least accurate. Both SBEDE and EE models quantified uncertainty equally well. All CP values are slightly negatively biased, indicating more extreme realizations than expected based on the models. Uncertainty quantification is not applicable to raw consensus. Overall, differences in forecasting accuracy are small.

5.2 Decision performance

Instead of accurate stock price forecasting, the investor's main objective in this study is to maximize the capital growth rate. We compare the performance of the allocation decisions for 37 industries based on 1) the SBEDE model, 2) the EE model and 3) an index investor who ignores expert target prices and shares capital equally among all companies within the industry (Index). Boxplots in Fig. 4 show the capital growth performance of the models and the index investor in 37 industries. The results are presented as basis points (BPS), where 100 BPS equals a 1% growth rate. Averages are calculated as geometric mean (GMean) as they describe

Table 2 Forecasting accuracy of the SBEDE model, the EE model and raw consensus median of expert predictions (Raw) with S&P500 data

	Method	MAE	MAE Premium	CP 50	CP 68	CP 90	CP 95
1	SBEDE	0.249	0.142	0.407	0.605	0.859	0.921
2	EE	0.251	0.147	0.424	0.599	0.831	0.915
3	Raw	0.233	0.152				

MAE, MAE Premiun (for predicted returns over the index) and 50%, 68%, 90% and 95% coverage probabilities

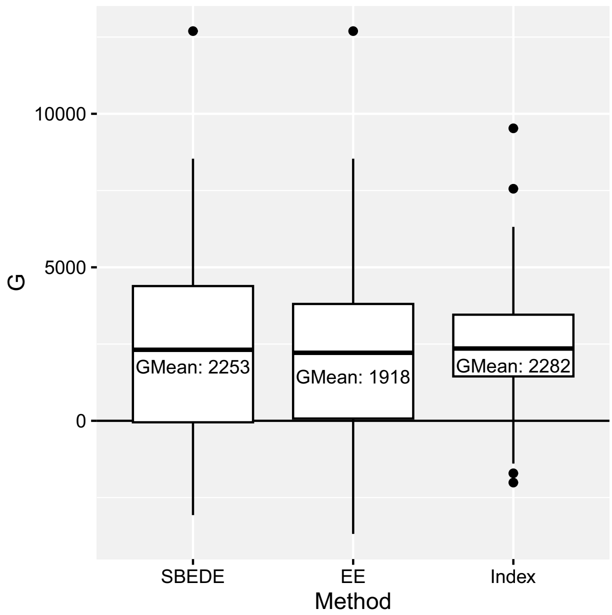

Fig. 4 Observed capital growth rates (BPS) in 37 industries during the experiment year 2021 (S&P500 data) with SBEDE and EE models and Index. Standard boxplots show geometric mean values between the lower quartile and the median lines

the multiplicative nature of capital growth. On average, the SBEDE model offered higher growth rate than the EE model but slightly lower than Index. The variation between industries was lower with Index compared to the SBEDE and EE models.

The growth rates that exceed the Index's growth rates are called premium growth rates. On average, the SBEDE model generated -30 BPS and the EE model -365 BPS premium growth rates. Compared to these averages, premium growth rates had high standard deviations between industries, 2204 BPS for the SBEDE model and 2250 BPS for the EE model.

As mentioned in the introduction, uncertainty in future returns may cause disappointing portfolio performance, known as the optimizer's curse. Here optimizer's curse is measured with expected and realized premium growth rates. Table 3 shows average growth rates with different models and compares them with Index. It also shows the models' average expected growth rates at the decision time. The realized

Table 3 Realized average portfolio growth rates (GMean G), return premiums, premium expectations, and realized optimizer's curse (OC) measured as BPS, with S&P500 data

	Method	GMean G	GMean G Index	Premium	Premium Exp	OC
1	SBEDE	2253	2282	−30	743	−773
2	EE	1918	2282	−365	805	−1170

optimizer's curse is the difference between average premium expectations and average premium realizations. The EE model caused a 1170 BPS disappointment to expectations, while SBEDE's disappointment was milder, 773 BPS.

To understand the role of the chance in these results, we conducted a Bayesian analysis for the advantage of SBEDE compared to EE and Index. The premium logarithmic growth rates (later referred to as premium for short) in industry l between models during the time period $[H, H+4]$ are

$$\Delta_{1l} = g^{\text{SBEDE}}_{[H,H+4]l} - g^{\text{EE}}_{[H,H+4]l}$$
$$\Delta_{2l} = g^{\text{SBEDE}}_{[H,H+4]l} - g^{\text{Index}}_{[H,H+4]l}.$$

We assume that the

$$\Delta_{1l} \sim N(\delta_1, \sigma^2_{\Delta 1})$$
$$\Delta_{2l} \sim N(\delta_2, \sigma^2_{\Delta 2}),$$

where δ_1 and δ_2 are SBEDE's mean premium over EE and Index, respectively. Figure 5 shows the estimated posterior distributions for the mean premiums using uninformative prior distributions. We conclude that there is 89% and 47% probability that SBEDE has a positive premium over EE and Index, respectively, if we apply these models to an additional industry.

5.3 Sensitivity analysis

Many prior distribution selections need some expert knowledge from a DM. In ideal conditions, prior distributions are selected based on knowledge of the particular industry. We did not have that information available and used the same prior distributions for all industries based on general knowledge of the stock markets. Especially, the prior selection

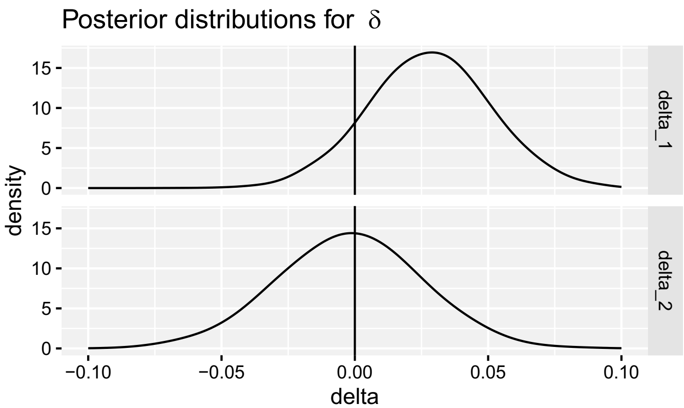

Fig. 5 Posterior distributions for the mean premium parameters $\boldsymbol{\delta}$. The areas on the positive side describe the probability of SBEDE having an advantage against EE (δ_1) and Index (δ_2)

$$\mu_{i[t,t+4]} \sim \mathrm{N}(r_{(t)} + 0.04, 0.1^2)$$

is critical as the standard deviation defines how much we allow predictable differences between stocks. To understand the results' sensitivity to this selection, we calculated average growth rates and optimizer's curse for alternative standard deviations (std) 0.05 and 0.2. We found (see Tables 10 and 11 in Appendix D) that the optimizer's curse reduced from −773 to −290 for SBEDE and from −1179 to −513 for EE with very restrictive std = 0.05. With std = 0.2, the curse increased to − 1473 and − 1787 for SBEDE and EE, respectively. The average growth rates of SBEDE were close to the index investor, and EE had the lowest growth with both of these alternative std selections. In addition, with std = 0.2, the decisions were courageous, as in almost all industries and with both methods, the weight of the selected stock was 1, indicating no diversification to the index. Stock selections and weights in the portfolio with all tested prior selections and both models are shown in Table 9 in Appendix C.

One benefit of the SBEDE model is the ability to label experts as informative and uninformative based on the parameters $\boldsymbol{\phi}$. We label the expert j as informative if the absolute value of the median of the posterior distribution of ϕ_j is larger than 0.2, as defined in Sect. 4.2. With this criterion, 190 of 384 experts were labelled as informative. If some industry does not have informative experts, an advantage over Index is not expected. We re-calculated premium performances based on the industries where the number of informative experts meets a specific minimum requirement. Table 4 shows, for example, that with 16 industries with at least four informative experts, SBEDE offered a 543 BPS higher growth rate than Index. By repeating the Bayesian analysis for the advantage in Sect. 5.2 but only for these 16 industries, we get a 79% probability that SBEDE has a positive premium over Index.

6 Simulation study

As the available real data was limited and sparse, we conducted a simulation study to understand better the SBEDE model's performance in different conditions and with larger data. In the study, we are altering the following conditions:

Table 4 Minimum number of informative experts, the methods, number of industries and premium growth rates over Index (S&P500 data)

Minimum	Method	Industries	Premium
0	SBEDE	37	−30
0	EE	37	−365
1	SBEDE	31	−131
1	EE	31	−296
2	SBEDE	24	32
2	EE	24	45
3	SBEDE	18	353
3	EE	18	197
4	SBEDE	16	543
4	EE	16	91

- number of historical data to fit model: $n_t = 20$ (5 years) and $n_t = 40$ (10 years)
- number of experts: $J = 4$ and $J = 8$.

We fixed the number of assets to 5 and assumed that half of the experts are uninformative ($\phi = 0$) and half balanced ($\phi = 1$). We simulated 500 data sets, including asset returns and expert predictions, for each of the 4 (n_t, J) conditions. More detailed assumptions and selected hyperparameter values for simulation are presented in Appendix E.

In addition to the SBEDE and EE models, we use the Merkle model, discussed earlier in this paper, as an alternative benchmark model. As many elements are the same as in the SBEDE model, we use the same prior distributions, presented in Appendix B with the following exceptions. Instead of horseshoe prior

$$\phi_j \sim \mathrm{N}(0, 1), j \in 1, \ldots, J,$$

which is similar to EE model (see Appendix B) but allows more variation because ϕ_j:s are not assumed to be identical. Inaccuracies (σ_j) do not have a hierarchical structure, and the prior is the same weakly informative one as used in Merkle et al (2020), namely Inv-Gamma(0.1,0.1). The model assumes independence between experts; thus, we set the correlation parameter ρ to zero. We also modified the model to fit better in the environment of fat-tailed shocks. We modified Equation (2) to a form $M_{[H]} \sim \mathrm{N}(b_j + \phi_j \mu_{[H]}, \sigma_j^2)$, where the expected value of a prediction is now a function of the expected value of a return.

In addition to the Kelly criterion, we tested two other investment objects to see if different personal preferences of DM affect the results. The first alternative object is the Sharpe ratio. (Sharpe 1994)

$$SR = \frac{\mathbf{E}(x)}{\sqrt{\mathrm{Var}(x)}}, \tag{8}$$

where the expected value $\mathbf{E}(x)$ and variance $\mathrm{Var}(x)$ are estimated based on the corresponding posterior distribution. The second alternative object is the expected utility of a risk-averse investor. In the mean-variance utility framework (Levy and Markowitz 1979), the objective to maximise is

$$U = \mathbf{E}(x) - \frac{RRA}{2}\mathrm{Var}(x),$$

where RRA is a DM's relative risk aversion level (Pulley 1981). Decisions made by the Kelly criterion are described by $RRA = 1$. However, for comparison, we select $RRA = 3$, which relates to higher risk aversion that is typical in most countries (Gandelman and Hernandez-Murillo 2015).

Figure 6 shows the accuracy of the point forecasts for each method in each condition. With all conditions and both metrics (MAE and MAE Premium), SBEDE has the best accuracy, but the differences to the benchmark models are small. Coverage probabilities are reported in Tables 13, 14, 15, 16 in Appendix E. In summary, the SBEDE and EE models have more accurate CP values than the Merkle model in every condition.

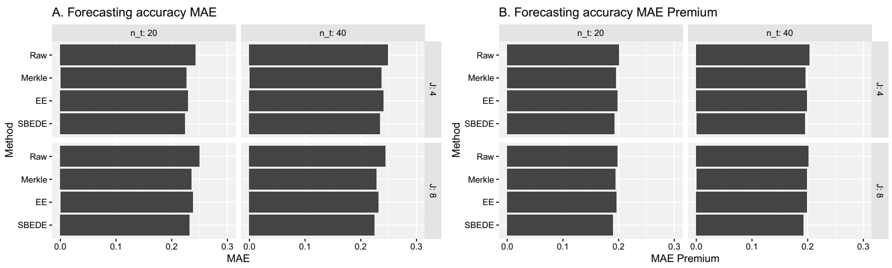

Fig. 6 Forecasting accuracy for return point forecasts of different methods in the simulation study. MAE = Mean Absolute Error. Different numbers of historical data as columns and numbers of experts as rows

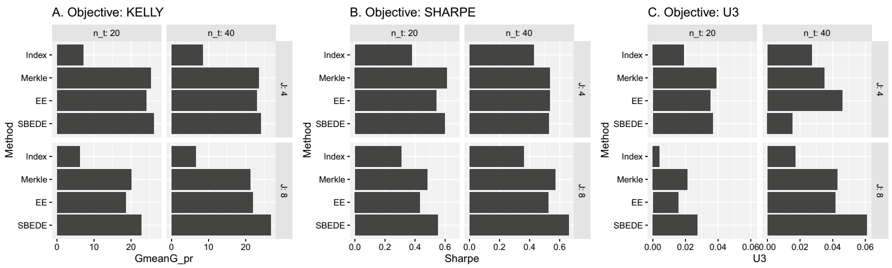

Fig. 7 Performance of different methods in the simulation study. Different investment objectives (performance metric) in panels: A. KELLY = Kelly crterion (geometric average capital growth rate %), B. SHARPE = Sharpe-ratio (realized Sharpe-ratio), C. U3 = Utility with risk aversion coefficient 3 (realized average utility). Different numbers of historical data as columns and numbers of experts as rows within each panel

Figure 7 shows decision performance metrics for all three objectives in all four conditions. Compared to forecasting accuracy, there is more variability in decision performance. The main finding is SBEDE's outperformance always when there are eight experts ($J = 8$), including four informative ones. In addition, SBEDE's performance increases when ten years of data are available. With the Kelly criterion, SBEDE offers the best capital growth rate in each condition but only with a small margin when $J = 4$. Another interesting finding is that methods' ranking based on forecasting accuracy is not always the same as the ranking based on decision performance. Index does not perform very well in this comparison.

7 Discussion

The real data application and the simulation study demonstrated the benefits of Bayesian models in expert prediction aggregation for portfolio decision support. The hierarchical structure and regularizing horseshoe prior distribution enables fitting a model with a high number of parameters with limited data. Applying informative prior distributions based on a DM's knowledge decreases the impact of outliers in the data and reduces the optimizer's curse. MCMC methods enable the use of various probability distributions, including those not defined in a closed form. For example, we did not have to assume normally distributed future returns. In addition, parameter uncertainty is naturally included in the return predictions.

In the real data results, SBEDE and the simpler EE model were equally accurate with the point forecasts. However, the SBEDE model showed a better capital growth performance. The average difference of 335 BPS in growth rates between the models and 89% probability of SBEDE's advantage over EE, combined with the simulation results, indicate some benefits of regularizing uninformative experts. In addition, SBEDE's over-performance against the Merkle model (forecasting and decision making) in all conditions with eight experts indicates the usefulness of lowering over-fitting by the horseshoe prior, hierarchical parameter structure and considering correlations between experts.

The SBEDE model offered, on average, almost identical capital growth during the experiment year with 37 industries compared with the index investor. However, in the 18 industries where at least three experts were labelled as informative, the SBEDE model generated a 353 BPS premium growth. When at least four experts are required to be informative, the premium growth was 543 in 16 industries. There is over 20% probability that this advantage is caused by chance. The simulation study also supported the conclusion that the SBEDE model needs more than two competent experts for outstanding performance. We focused on the capital growth rate, but the simulation study result also holds with alternative investment objectives.

Actually, SBEDE outperfomerd Index in the simulation study in every condition, but the model used to simulate data does not account for all variability and uncertainties of real life. Thus, the simulation results of all tested models are expected to be too optimistic and real-life performances are expected to be closer to Index.

The sensitivity analysis also demonstrated that the DM can manage the optimizer's curse by adjusting the prior distribution of return expectations. The average optimizer's curse of the SBEDE model was reduced from -773 BPS to -290 BPS by modifying the standard deviation of return expectations from 0.1 to 0.05. However, we think that 0.05 is too restrictive for a general use for all industries and ideally this selection should be made based on prior knowledge of each industry.

The results are in line with earlier studies. Clement (1999) argued that there are differences between experts, and likewise, the regularization of uninformative experts offered higher capital growth compared to the EE model in our application. Cvitanic et al (2006) found average expert judgment not very useful in stock markets, and now we saw a weak performance of the EE model. In addition, reduction of optimizer's curse with more informative prior selection was predicted by Smith and Winkler (2006).

Taleb (2020) compared differences in forecasting accuracy of a binary event and actual returns, claiming that the models that generate accurate forecasts do not necessarily generate high returns. Our simulation study, a similar phenomenon was observed with a continuous target variable (stock returns). For example, in some conditions the EE model's decision performance was better than the Merkle model even though it has worse accuracy in point forecasts. Also with the real data, SBEDE's edge against EE in forecasting accuracy was negligible compared to decision performance. This emphasizes that in addition to point forecasts, estimating the whole return distribution well and the investment objective matters when making portfolio decisions. Making accurate point forecasts is rarely the main objective of the DM.

8 Conclusions

We studied how biased and correlated expert predictions with different levels of expertise can be modeled with a hierarchical Bayesian model and a horseshoe prior for responsiveness to reality. The strength of the proposed SBEDE model is that it automatically debiases and regularizes experts without separate background studies of different experts. The expert model is designed to be combined with a separate asset pricing model and investment strategy. The aim of the model is to make better investment decisions and this study studied parallel forecasting and decision performance, and disappointments caused by the optimizer's curse.

In the real data application, the portfolio decisions based on SBEDE generated higher growth rates compared to the EE model. In addition, SBEDE outperformed all the benchmark models in the simulation study when there was four competent experts. This indicates the usefulness of regularizing uninformative experts with a horseshoe prior distribution.

With real data, SBEDE had a positive premium growth over the index when applied only to industries that were estimated to have multiple competent experts, but we cannot generalize these results to future periods. The simulation study showed outstanding performance for SBEDE but studies with more real data are needed to gather evidence about the sustained premium of the SBEDE model. Also, transaction costs and taxes reduce the benefits of stock selections and should be considered in future studies.

The main weakness of the SBEDE model is the need for a decent amount (we used 16 observations as a threshold) of historical data from the earlier predictions of each expert. As most companies had at most 5–7 years of systematically collected expert data available, we were able to conduct the real data analysis only at one time period.

In portfolio decisions, the SBEDE model has to be accompanied with a reasonable model for unpredictable shocks. That is a challenge with fat-tailed stock returns, and there may be better options than the used symmetric t-distribution. Because of the low number of collected expert target price data, it was not reasonable to model experts' learning process, but this is a potential direction for future research.

For simplicity, we conducted investment decisions at the end of a quarter, which may not be optimal timing. An interesting question for a future study would be: Could SBEDE's performance be improved by timing decisions immediately after informative experts update their target prices?

The sensitivity analysis demonstrated that the DM's prior selections affect the optimizer's curse. Seeking higher performance and reducing the curse with more informative prior distributions based on industry-wise prior information is an interesting direction of future development. With more data and better insights from the DM as informative prior distributions, the SBEDE model has much potential in decision-making for combining and presenting information from historical stock price data, expert predictions and the DM's prior knowledge. The SBEDE model turned out to be usable and could be cautiously applied to real investment decisions keeping the limitations in mind.

Appendix A

See Tables 5, 6.

Table 5 Industries in analyses, and numbers of qualified companies (N) and experts (J)

	Industry	N	J
1	Air Freight & Logistics	4	10
2	Airlines	4	4
3	Apparel Retail	4	7
4	Apparel, Accessories & Luxury Goods	7	10
5	Application Software	7	22
6	Biotechnology	6	16
7	Casinos & Gaming	4	4
8	Communications Equipment	3	6
9	Data Processing & Outsourced Services	6	12
10	Diversified Banks	6	12
11	Electric Utilities	9	5
12	General Merchandise Stores	3	10
13	Health Care Distributors	6	14
14	Health Care Equipment	12	19
15	Hotels, Resorts & Cruise Lines	5	8
16	Industrial Machinery	3	5
17	Integrated Oil & Gas	3	10
18	Interactive Home Entertainment	3	9
19	Interactive Media & Services	3	19
20	Internet & Direct Marketing Retail	4	19
21	Life & Health Insurance	3	4
22	Managed Health Care	4	5
23	Movies & Entertainment	3	20
24	Multi-Utilities	6	6
25	Oil & Gas Equipment & Services	4	7
26	Oil & Gas Exploration & Production	9	19
27	Oil & Gas Refining & Marketing	3	6
28	Oil & Gas Storage & Transportation	3	7
29	Pharmaceuticals	3	9
30	Property & Casualty Insurance	4	7
31	Railroads	3	10
32	Restaurants	5	9
33	Retail REITs	3	3
34	Semiconductors	10	16
35	Specialized REITs	4	9
36	Specialty Stores	5	10
37	Systems Software	3	16

Table 6 Industries and companies with means and standard deviations of returns (x) and expert predictions (M) for the period for training the models

Industry	Company ID	$\bar{x}$	$sd(x)$	$\bar{M}$	$sd(M)$
Air Freight & Logistics	1	0.054	0.154	0.047	0.112
Air Freight & Logistics	2	0.105	0.091	0.048	0.093
Air Freight & Logistics	3	0.035	0.287	0.154	0.103
Air Freight & Logistics	4	0.042	0.132	0.063	0.084
Airlines	1	−0.028	0.351	0.201	0.108
Airlines	2	−0.021	0.320	0.234	0.100
Airlines	3	0.062	0.306	0.097	0.206
Airlines	4	−0.037	0.450	0.185	0.149
Apparel Retail	1	−0.160	0.426	0.319	0.166
Apparel Retail	2	−0.179	0.378	0.062	0.124
Apparel Retail	3	0.157	0.161	0.077	0.090
Apparel Retail	4	0.100	0.128	0.091	0.073
Apparel, Accessories & Luxury Goods	1	−0.095	0.294	0.221	0.187
Apparel, Accessories & Luxury Goods	2	0.162	0.171	0.063	0.124
Apparel, Accessories & Luxury Goods	3	−0.127	0.365	0.155	0.101
Apparel, Accessories & Luxury Goods	4	−0.124	0.290	0.059	0.136
Apparel, Accessories & Luxury Goods	5	−0.148	0.342	0.119	0.190
Apparel, Accessories & Luxury Goods	6	−0.150	0.442	−0.037	0.297
Apparel, Accessories & Luxury Goods	7	0.018	0.233	0.111	0.118
Application Software	1	0.298	0.169	0.098	0.095
Application Software	2	0.291	0.176	0.076	0.083
Application Software	3	0.196	0.189	0.185	0.093
Application Software	4	0.212	0.149	−0.027	0.116
Application Software	5	0.058	0.122	0.102	0.097
Application Software	6	0.254	0.163	0.122	0.085
Application Software	7	0.216	0.196	−0.005	0.124
Biotechnology	1	0.099	0.121	0.145	0.110
Biotechnology	2	−0.004	0.231	0.206	0.130
Biotechnology	3	−0.032	0.184	0.189	0.136
Biotechnology	4	0.083	0.385	0.177	0.172
Biotechnology	5	0.091	0.323	0.136	0.149
Biotechnology	6	0.175	0.316	0.131	0.238
Casinos & Gaming	1	−0.069	0.265	0.170	0.123
Casinos & Gaming	2	−0.040	0.288	0.231	0.108
Casinos & Gaming	3	0.185	0.506	0.214	0.244
Casinos & Gaming	4	−0.152	0.490	0.091	0.237
Communications Equipment	1	0.093	0.164	0.087	0.129
Communications Equipment	2	0.037	0.231	0.097	0.117
Communications Equipment	3	−0.016	0.134	0.056	0.135
Data Processing & Outsourced Services	1	0.144	0.084	0.110	0.085
Data Processing & Outsourced Services	2	0.193	0.112	0.006	0.115

Table 6 (continued)

Industry	Company ID	$\bar{x}$	$sd(x)$	$\bar{M}$	$sd(M)$
Data Processing & Outsourced Services	3	0.256	0.148	0.083	0.119
Data Processing & Outsourced Services	4	0.220	0.153	0.106	0.087
Data Processing & Outsourced Services	5	0.202	0.104	0.121	0.077
Data Processing & Outsourced Services	6	0.032	0.139	−0.005	0.117
Diversified Banks	1	0.066	0.252	0.148	0.120
Diversified Banks	2	−0.002	0.260	0.143	0.172
Diversified Banks	3	−0.038	0.378	0.061	0.108
Diversified Banks	4	0.085	0.178	0.094	0.083
Diversified Banks	5	−0.016	0.184	0.058	0.077
Diversified Banks	6	−0.099	0.264	0.046	0.109
Electric Utilities	1	0.073	0.127	0.020	0.075
Electric Utilities	2	0.017	0.102	0.064	0.080
Electric Utilities	3	0.026	0.099	0.027	0.068
Electric Utilities	4	0.049	0.139	−0.058	0.069
Electric Utilities	5	0.002	0.187	0.046	0.079
Electric Utilities	6	0.045	0.168	0.026	0.079
Electric Utilities	7	0.003	0.200	0.084	0.091
Electric Utilities	8	0.053	0.132	0.049	0.058
Electric Utilities	9	0.037	0.145	−0.007	0.079
General Merchandise Stores	1	0.187	0.167	0.097	0.080
General Merchandise Stores	2	0.080	0.208	0.074	0.104
General Merchandise Stores	3	0.124	0.248	0.063	0.125
Health Care Distributors	1	0.047	0.207	0.116	0.094
Health Care Distributors	2	0.023	0.176	0.064	0.198
Health Care Distributors	3	−0.054	0.176	0.134	0.090
Health Care Distributors	4	0.038	0.144	0.248	0.151
Health Care Distributors	5	−0.026	0.188	0.105	0.121
Health Care Distributors	6	0.106	0.153	0.009	0.079
Health Care Equipment	1	0.131	0.158	0.129	0.135
Health Care Equipment	2	0.355	0.574	0.121	0.187
Health Care Equipment	3	0.140	0.160	0.092	0.085
Health Care Equipment	4	0.115	0.118	0.068	0.072
Health Care Equipment	5	0.163	0.182	0.117	0.091
Health Care Equipment	6	0.353	0.424	0.140	0.154
Health Care Equipment	7	0.269	0.193	0.087	0.092
Health Care Equipment	8	0.136	0.206	0.086	0.098
Health Care Equipment	9	0.075	0.109	0.117	0.062
Health Care Equipment	10	0.138	0.134	0.071	0.075
Health Care Equipment	11	0.193	0.118	0.092	0.075
Health Care Equipment	12	0.035	0.158	0.127	0.111
Hotels, Resorts & Cruise Lines	1	−0.139	0.465	0.156	0.145
Hotels, Resorts & Cruise Lines	2	0.085	0.213	0.273	0.246

Table 6 (continued)

Industry	Company ID	$\bar{x}$	$sd(x)$	$\bar{M}$	$sd(M)$
Hotels, Resorts & Cruise Lines	3	0.074	0.294	0.049	0.131
Hotels, Resorts & Cruise Lines	4	−0.113	0.556	0.186	0.140
Hotels, Resorts & Cruise Lines	5	−0.009	0.441	0.173	0.130
Industrial Machinery	1	0.116	0.156	0.002	0.108
Industrial Machinery	2	0.066	0.225	0.051	0.137
Industrial Machinery	3	0.116	0.192	0.072	0.101
Integrated Oil & Gas	1	−0.071	0.229	0.108	0.100
Integrated Oil & Gas	2	−0.114	0.296	0.115	0.172
Integrated Oil & Gas	3	−0.146	0.247	0.001	0.138
Interactive Home Entertainment	1	0.212	0.318	0.136	0.128
Interactive Home Entertainment	2	0.227	0.284	0.106	0.148
Interactive Home Entertainment	3	0.319	0.246	0.091	0.153
Interactive Media & Services	1	0.202	0.173	0.168	0.151
Interactive Media & Services	2	0.147	0.135	0.156	0.091
Interactive Media & Services	3	−0.051	0.468	−0.053	0.273
Internet & Direct Marketing Retail	1	0.332	0.257	0.158	0.113
Internet & Direct Marketing Retail	2	0.055	0.159	0.138	0.084
Internet & Direct Marketing Retail	3	0.108	0.184	0.324	0.412
Internet & Direct Marketing Retail	3	0.016	0.275	0.128	0.123
Life & Health Insurance	1	0.029	0.186	0.022	0.094
Life & Health Insurance	2	0.066	0.151	−0.025	0.087
Life & Health Insurance	3	−0.109	0.328	0.043	0.191
Managed Health Care	1	0.151	0.209	0.168	0.133
Managed Health Care	2	0.118	0.216	0.172	0.144
Managed Health Care	3	0.187	0.186	0.106	0.128
Managed Health Care	4	0.209	0.137	0.129	0.087
Movies & Entertainment	1	0.066	0.149	0.070	0.111
Movies & Entertainment	2	0.335	0.324	−0.007	0.327
Movies & Entertainment	3	−0.146	0.320	0.175	0.136
Multi-Utilities	1	0.115	0.105	0.025	0.066
Multi-Utilities	2	0.061	0.146	0.200	0.058
Multi-Utilities	3	0.099	0.107	0.050	0.069
Multi-Utilities	4	0.021	0.183	0.108	0.079
Multi-Utilities	5	0.164	0.096	0.063	0.073
Multi-Utilities	6	0.051	0.161	0.000	0.064
Oil & Gas Equipment & Services	1	−0.181	0.292	0.378	0.221
Oil & Gas Equipment & Services	2	−0.261	0.402	0.226	0.191
Oil & Gas Equipment & Services	3	−0.306	0.363	0.062	0.177
Oil & Gas Equipment & Services	4	−0.285	0.331	0.183	0.130
Oil & Gas Exploration & Production	1	−0.367	0.514	0.110	0.195
Oil & Gas Exploration & Production	2	−0.121	0.327	0.137	0.160
Oil & Gas Exploration & Production	3	−0.117	0.202	0.189	0.161

Table 6 (continued)

Industry	Company ID	$\bar{x}$	$sd(x)$	$\bar{M}$	$sd(M)$
Oil & Gas Exploration & Production	4	−0.303	0.468	0.301	0.201
Oil & Gas Exploration & Production	5	−0.138	0.357	0.205	0.157
Oil & Gas Exploration & Production	6	−0.113	0.484	0.225	0.155
Oil & Gas Exploration & Production	7	−0.329	0.550	0.300	0.236
Oil & Gas Exploration & Production	8	−0.311	0.520	0.147	0.135
Oil & Gas Exploration & Production	9	−0.127	0.294	0.282	0.172
Oil & Gas Refining & Marketing	1	−0.045	0.348	0.229	0.151
Oil & Gas Refining & Marketing	2	−0.041	0.245	0.112	0.119
Oil & Gas Refining & Marketing	3	0.002	0.332	0.099	0.132
Oil & Gas Storage & Transportation	1	−0.153	0.370	0.189	0.168
Oil & Gas Storage & Transportation	2	−0.120	0.513	0.067	0.180
Oil & Gas Storage & Transportation	3	−0.154	0.393	0.167	0.222
Pharmaceuticals	1	0.081	0.186	0.105	0.182
Pharmaceuticals	2	0.058	0.100	0.046	0.108
Pharmaceuticals	3	0.152	0.161	0.085	0.145
Property & Casualty Insurance	1	−0.090	0.250	0.078	0.102
Property & Casualty Insurance	2	0.033	0.143	0.082	0.071
Property & Casualty Insurance	3	0.015	0.207	0.090	0.080
Property & Casualty Insurance	4	0.037	0.151	0.027	0.083
Railroads	1	0.139	0.284	0.082	0.109
Railroads	2	0.099	0.226	0.106	0.092
Railroads	3	0.096	0.211	0.091	0.101
Restaurants	1	0.089	0.383	0.033	0.163
Restaurants	2	0.093	0.286	0.087	0.096
Restaurants	3	0.109	0.138	0.074	0.073
Restaurants	4	0.121	0.204	0.074	0.107
Restaurants	5	0.075	0.168	0.242	0.211
Retail REITs	1	−0.057	0.250	0.038	0.069
Retail REITs	2	−0.094	0.303	0.063	0.118
Retail REITs	3	−0.140	0.364	0.098	0.058
Semiconductors	1	0.128	0.144	0.125	0.088
Semiconductors	2	0.097	0.149	0.050	0.132
Semiconductors	3	0.119	0.209	0.136	0.106
Semiconductors	4	0.293	0.164	0.087	0.097
Semiconductors	5	0.099	0.543	0.395	0.221
Semiconductors	6	0.498	0.503	0.034	0.223
Semiconductors	7	0.106	0.265	0.168	0.136
Semiconductors	8	0.045	0.260	0.095	0.122
Semiconductors	9	0.179	0.394	0.137	0.149
Semiconductors	10	0.164	0.138	0.016	0.127
Specialized REITs	1	0.160	0.126	0.129	0.125
Specialized REITs	2	0.216	0.162	0.109	0.107

Table 6 (continued)

Industry	Company ID	$\bar{x}$	$sd(x)$	$\bar{M}$	$sd(M)$
Specialized REITs	3	0.173	0.164	0.133	0.112
Specialized REITs	4	−0.041	0.201	0.136	0.112
Specialty Stores	1	0.118	0.215	0.057	0.127
Specialty Stores	2	0.083	0.202	0.091	0.138
Specialty Stores	3	0.165	0.229	0.085	0.097
Specialty Stores	4	0.096	0.252	0.089	0.146
Specialty Stores	5	0.125	0.327	0.095	0.100
Systems Software	1	0.272	0.291	0.127	0.171
Systems Software	2	0.249	0.130	0.095	0.097
Systems Software	3	0.314	0.242	0.070	0.158

Appendix B: Details of Bayesian modelling

The global shrinkage factor $\tau \sim Gamma(27, 416)$ is based on a prior estimate, where we use inequality $|\phi| > 0.2$ as a proxy for an informative expert. Our heuristic prior estimate was "around 10–20% of the analysts are informative". With the chosen prior distribution, the probability $P(|\phi| > 0.2)$ is within [0.1, 0.2] with a 95% probability. For the correlation matrix between experts, we are using an uninformative distribution $\Omega \sim$ LKJ(1) (Lewadonsky-Kurowicka-Joe distribution, see Lewandowski et al 2009).

For other parameters, we use weakly informative distributions that guide the model towards plausible values and help the model to converge but avoid taking strong subjective opinions. Table 7 shows weakly informative prior distributions for the asset model.

Both degrees of freedom parameters ν_0 and ν have restricted support $(2, \infty)$ as we assume that the related t-distributions have a finite variance. In addition, the prior distribution selections of degrees of freedom parameters are influenced by Juárez and Steel (2010). Table 8 shows weakly informative prior distributions for the SBEDE model.

With the EE model, we use the same prior distributions as with the SBEDE model for parameters κ, ξ and σ^*. However, two prior decisions are specific for the EE model. For ϕ^* we use the weakly informative $N(0, 0.75^2)$ distribution as we do not expect most experts to be very informative but still want to allocate reasonable probability for balanced experts. For a correlation coefficient ρ, we use the uninformative uniform $U(-1, 1)$ distribution.

We fitted the models with the MCMC algorithm "Hamiltonian Monte Carlo with the No-U-turn Sampler" (Hoffman and Gelman 2014) using RStan (Stan Development Team 2023). We run 8 MCMC chains, each having 7500 iterations and 7500 warm-up iterations. We iteratively found out that with this number of iterations, most industries converge well. As a result, we had a total of $S = 60000$ simulated draws from posterior predictive distributions of $x_{i[H,H+4]}$ for all companies in every analyzed industry. These samples present possible future scenarios for next year's

Table 7 Weakly informative prior distributions for the asset model

Parameter	Distribution
Scale parameter for industry shocks, ψ_0	$\mathrm{HN}(0, 0.075^2)$
Average scale for asset shocks within the industry, ψ^*	$\mathrm{HN}(0, 0.075^2)$
Standard deviation of random effect coefficient c_i, ω	$\mathrm{HN}(0, 1.5^2)$
Industry shock distribution's degrees of freedom, ν_0	*Gamma*(2, 0.1)
Asset shock distribution's degrees of freedom, ν	*Gamma*(2, 0.1)

HN refers to a half-normal distribution, and *Gamma* refers to a gamma distribution with shape and rate parameters

Table 8 Weakly informative prior distributions for the SBEDE model

Parameter	Distribution/value
Mean bias in the industry, κ	$\mathrm{N}(0, 0.15^2)$
Standard deviation of biases, ξ	0.3
Average inaccuracy of experts, σ^*	$\mathrm{HN}(0, 0.15^2)$
Standard deviation of random effect coefficients d_j, ι	$\mathrm{HN}(0, 1.5^2)$

N refers to a normal distribution and HN to a half-normal distribution

stock returns. We monitored the convergence of the chains of $x_{i[H,H+4]}$ parameters with statistics $\hat{R}$, bulk effective sample size (bulk-ESS) and tail effective sample size (tail-ESS) (Vehtari et al 2021). We set the criteria for convergence as bulk-ESS > 400 and tail-ESS > 400 as recommended by Vehtari et al (2021). For $\hat{R}$, they suggest the criterion $\hat{R} < 1.01$, but we chose more liberal $\hat{R} < 1.03$. We based the selections for S and the convergence criteria on a satisfying compromise between computing time and the accuracy of the forecasts of $x_{i[H,H+4]}$. One industry did not meet the convergence criteria, but we tried again, increasing S to 600,000, and the criteria was met. R and Stan codes and example data are available in Heikkinen (2024).

Appendix C: Optimized decisions

See Table 9.

Table 9 Selected stocks and their weights based on SBEDE and EE models using S&P500 data with different prior selections

EEi01	SBi0.1	EEw0.1	SBw0.1	EEi0.05	SBi0.05	EEw0.05	SBw0.05	EEi0.2	SBi0.2	EEw0.2	SBw0.2
1	1	1.00	1.00	1	1	1.00	1.00	1	1	1.00	1.00
1	3	1.00	1.00	1	3	1.00	1.00	1	1	1.00	1.00
2	2	1.00	1.00	2	1	1.00	1.00	2	2	1.00	1.00
3	3	1.00	1.00	3	1	1.00	1.00	3	3	1.00	1.00
3	3	1.00	1.00	3	3	0.93	0.83	3	3	1.00	1.00
1	1	1.00	1.00	1	1	1.00	1.00	1	1	1.00	1.00
2	2	1.00	1.00	2	2	1.00	1.00	2	2	1.00	1.00
3	2	1.00	1.00	3	2	1.00	1.00	3	2	1.00	1.00
2	2	1.00	1.00	2	4	1.00	0.80	2	2	1.00	1.00
6	6	0.97	0.98	6	6	0.81	0.88	6	6	1.00	1.00
3	1	1.00	1.00	3	1	1.00	1.00	3	3	1.00	1.00
2	2	1.00	0.92	2	2	0.90	0.85	2	2	1.00	1.00
1	1	1.00	1.00	1	3	0.98	0.59	1	1	1.00	1.00
1	2	1.00	0.95	2	3	0.64	0.66	2	2	1.00	1.00
3	3	1.00	1.00	1	1	0.75	0.69	3	3	1.00	1.00
1	3	1.00	0.53	3	3	0.50	0.51	1	1	1.00	1.00
3	3	1.00	1.00	3	3	1.00	1.00	3	1	0.78	1.00
3	3	1.00	1.00	3	5	1.00	0.90	3	3	1.00	1.00
1	3	1.00	1.00	3	3	0.96	1.00	1	1	1.00	1.00
2	3	1.00	1.00	2	2	1.00	0.96	2	3	1.00	1.00
1	1	1.00	1.00	1	1	1.00	1.00	1	1	1.00	1.00
2	6	1.00	1.00	2	3	1.00	1.00	2	6	1.00	1.00
2	2	1.00	1.00	2	2	1.00	1.00	2	2	1.00	1.00
2	2	1.00	1.00	2	3	1.00	1.00	2	1	1.00	1.00

Table 9 (continued)

EEi01	SBi0.1	EEw0.1	SBw0.1	EEi0.05	SBi0.05	EEw0.05	SBw0.05	EEi0.2	SBi0.2	EEw0.2	SBw0.2
1	1	1.00	1.00	3	1	1.00	0.83	1	1	1.00	1.00
2	1	1.00	1.00	2	4	1.00	0.92	2	1	1.00	1.00
2	7	1.00	1.00	2	7	1.00	1.00	2	2	1.00	1.00
5	6	1.00	1.00	3	3	1.00	1.00	5	5	1.00	1.00
3	3	1.00	1.00	3	3	1.00	1.00	3	2	1.00	1.00
1	1	1.00	1.00	1	1	1.00	1.00	1	1	1.00	1.00
2	4	1.00	1.00	2	4	1.00	0.76	2	2	1.00	1.00
6	6	1.00	1.00	6	6	1.00	1.00	6	6	1.00	1.00
2	2	1.00	1.00	2	2	1.00	1.00	2	2	1.00	1.00
1	2	1.00	1.00	1	2	1.00	1.00	1	1	1.00	1.00
3	3	1.00	1.00	3	3	1.00	1.00	1	1	1.00	1.00
4	4	1.00	1.00	4	5	1.00	1.00	6	4	1.00	1.00
6	6	1.00	1.00	6	6	0.91	1.00	6	6	1.00	1.00

One line represents one industry. *EE* Exchangeable Expert model, *SB* SBEDE model, *i* selected stock, *w* optimized weight in the portfolio. The number refers to standard deviation of the expected return's prior distribution. The main results in Sect. 5 are based on the standard deviation 0.1

Table 10 Sensitivity analysis results (S&P500 data) with prior selection: std of the expected return is 0.05

	Method	GMean G	GMean G Index	Premium	Premium Exp	OC
1	SBEDE	2296	2282	13	304	−290
2	EE	2146	2282	−137	376	−513

Realized average portfolio growth rates (GMean G), return premiums, premium expectations, and realized optimizer's curse (OC), measured as BPS

Table 11 Sensitivity analysis results with prior selection (S&P500 data): std of the expected return is 0.2

	Method	GMean G	GMean G Index	Premium	Premium Exp	OC
1	SBEDE	2290	2282	8	1480	−1473
2	EE	1981	2282	−301	1486	−1787

Realized average portfolio growth rates (GMean G), return premiums, premium expectations, and realized optimizer's curse (OC), measured as BPS

Table 12 Hyperparameters for randomizing stock market returns and expert features

Type	Hyperparameter	Value
Market	ν_0	10
Market	ν	7
Market	ψ_0	0.05
Market	ω	0.1
Market	$\mathbf{E}(\mu)$	0.05
Market	$SD(\mu)$	0.1
Expert	κ	0.1
Expert	σ^*	0.075
Expert	ι	0.5
Expert	$\mathbf{E}(\rho)$	0.45
Expert	hi	0.15

Table 13 Simulation study coverage probabilities (CP) when $n_t = 20$ and $J = 4$

	Method	CP 50	CP 68	CP 90	CP 95
1	SBEDE	0.531	0.708	0.901	0.945
2	EE	0.526	0.704	0.901	0.945
3	Merkle	0.566	0.742	0.924	0.962

Table 14 Simulation study coverage probabilities (CP) when $n_t = 20$ and $J = 8$

	Method	CP 50	CP 68	CP 90	CP 95
1	SBEDE	0.493	0.689	0.904	0.952
2	EE	0.489	0.683	0.898	0.946
3	Merkle	0.570	0.758	0.947	0.975

Table 15 Simulation study coverage probabilities (CP) when $n_t = 40$ and $J = 4$

	Method	CP 50	CP 68	CP 90	CP 95
1	SBEDE	0.492	0.668	0.898	0.948
2	EE	0.496	0.672	0.902	0.946
3	Merkle	0.536	0.711	0.921	0.966

Table 16 Simulation study coverage probabilities (CP) when $n_t = 40$ and $J = 8$

	Method	CP 50	CP 68	CP 90	CP 95
1	SBEDE	0.524	0.700	0.913	0.956
2	EE	0.510	0.694	0.900	0.956
3	Merkle	0.607	0.785	0.957	0.981

Appendix D: Results with alternative prior selections

See Tables 10, 11.

Appendix E: Details of the simulation study

Stock market returns are randomized based on the asset model in Sect. 3.2, and expert features and predictions are randomized based on the SBEDE model in Sect. 3.3 with the following additions. Expected returns

$$\mu_{i[t,t+4]} \sim \mathrm{N}(\mathbf{E}(\mu), SD(\mu)^2) \tag{E1}$$

and correlations between experts

$$\rho_{j,k} \sim \mathrm{Uniform}(\mathbf{E}(\rho) - hi, \mathbf{E}(\rho) + hi), \tag{E2}$$

where *hi* stands for half-interval. Table 12 shows the values for all the needed hyperparameters. They are selected with the guidance of the real data. As these values differ between industries, we encourage readers to repeat these simulations with different selections. The codes are available in Heikkinen (2024). Tables 13, 14, 15 and 16 show the coverage probabilities with different conditions in the simulation study.

Acknowledgements This research is related to the thematic research area DEMO (Decision Analytics utilizing Causal Models and Multiobjective Optimization, jyu.fi/demo) of the University of Jyvaskyla. We want to thank professor Juha-Pekka Junttila warmly for his comments about dividend management in portfolio optimization and CSC (IT Center for Science) for offering computing resources.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Risto Heikkinen. The first draft of the manuscript was written by Risto Heikkinen and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided by University of Jyväskylä (JYU). This work was supported by the Finnish Cultural Foundation, Central Finland Regional Fund under Grant 30202205; Finnish Cultural Foundation under Grant 00210355; and the Foundation for Economic Education under Grant 2211996.

Availability of data and materials A data set for the illustration of SBEDE is available. It includes one example industry with real stock returns and artificial expert predictions. In addition, the SBEDE model, coded with Stan, and R scripts for fitting the models and optimizing a portfolio with an example industry, are available. The Zenodo code repository also includes the codes for conducting a simulation study (https://doi.org/10.5281/zenodo.13805536). The real data is not publicly available as it contains proprietary information that the authors acquired through a license, but the meta-data will be published.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Anderson MJ, Sunder S (1995) Professional traders as intuitive Bayesians. Organ Behav Hum Decis Process 64(2):185–202. https://doi.org/10.1006/obhd.1995.1099

Ardia D, Boudt K, Carl P et al (2011) Differential evolution with DEoptim. R J 3(1):27–34. https://doi.org/10.32614/RJ-2011-005

Barber B, Lehavy R, McNichols M et al (2001) Can investors profit from the prophets? Security analyst recommendations and stock returns. J Financ 56(2):531–563. https://doi.org/10.1111/0022-1082.00336

Batchelder WH, Romney AK (1988) Test theory without an answer key. Psychometrika 53(1):71–92. https://doi.org/10.1007/BF02294195

Black F, Litterman R (1992) Global portfolio optimization. Financ Anal J 48(5):28–43. https://doi.org/10.2469/faj.v48.n5.28

Boni L, Womack KL (2006) Analysts, industries, and price momentum. J Financ Quant Anal 41(1):85–109. https://doi.org/10.1017/S002210900000243X

Bradshaw M, Brown L, Huang K (2013) Do sell-side analysts exhibit differential target price forecasting ability? Rev Acc Stud 18:930–955. https://doi.org/10.1007/s11142-012-9216-5

Brav A, Lehavy R (2003) An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. J Financ 58(2):1933–1967. https://doi.org/10.1111/1540-6261.00593

Breiman L (1961) Optimal gambling systems for favorable games. In: Proceedings of the 4th Berkeley symposium on mathematical statistics and probability, vol 1, pp 63–68

Byrd RH, Lu P, Nocedal J et al (1995) A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput 16(5):1190–1208. https://doi.org/10.1137/0916069

Carvalho CM, Polson NG (2010) The horseshoe estimator for sparse signals. Biometrika 97(2):465–480. https://doi.org/10.1093/biomet/asq017

Chen SD, Lim AEB (2020) A generalized Black-Litterman model. Oper Res 68(2):381–410. https://doi.org/10.1287/opre.2019.1893

Clemen R, Winkler R (1985) Limits for the precision and value of information from dependent sources. Oper Res 33(2):427–442. https://doi.org/10.1287/opre.33.2.427

Clemen RT, Winkler RL (1999) Combining probability distributions from experts in risk analysis. Risk Anal 19(2):127–203. https://doi.org/10.1111/j.1539-6924.1999.tb00399.x
Clement MB (1999) Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? J Account Econ 27:285–303. https://doi.org/10.1016/S0165-4101(99)00013-0
Cooke R (1991) Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, Oxford
Cooke RM, Goossens LLHJ (2008) TU Delft expert judgment data base. Reliab Eng Syst Saf 93(5):657–674. https://doi.org/10.1016/j.ress.2007.03.005
Cvitanic J, Lazrak A, Martellini L et al (2006) Dynamic portfolio choice with parameter uncertainty and the economic value of analysts' recommendations. Rev Financ Stud 19(4):1113–1156. https://doi.org/10.1093/rfs/hhj039
Da Z, Schaumburg E (2011) Relative valuation and analyst target price forecasts. J Financ Mark 14:161–192. https://doi.org/10.1016/j.finmar.2010.09.001
Elton E, Gruber M, Brown S et al (2014) Modern portfolio theory and investment analysis, 9th edn. John Wiley & Sons, Chichester
Fama E (1970) Efficient capital markets: a review of theory and empirical work. J Financ 25(2):383–417. https://doi.org/10.2307/2325486
Gandelman N, Hernandez-Murillo R (2015) Risk aversion at the country level, available at SSRN: https://ssrn.com/abstract=2646134
Gelman A, Carling J, Stern H et al (2013) Bayesian data analysis, 3rd edn. Chapman & Hall / CRC, New York
Grossman S, Stiglitz J (1980) On the impossibility of informationally efficient markets. Am Econ Rev 70:393–408
Hao J, Skinner J (2023) Analyst target price and dividend forecasts and expected stock returns. J Asset Manag 24:108–120. https://doi.org/10.1057/s41260-022-00283-z
Hartley D, French S (2021) A Bayesian method for calibration and aggregation of expert judgement. Int J Approx Reason 130:192–225. https://doi.org/10.1016/j.ijar.2020.12.007
Heikkinen R (2024) RiskyRisto/SBEDE-open: Publication release, *Zenodo*, 0.2,https://doi.org/10.5281/zenodo.13805536
Hoffman M, Gelman A (2014) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15:1593–1623
Hu W, Kercheval A (2010) Portfolio optimization for student t and skewed t returns. Quant Finance 10(1):91–105. https://doi.org/10.1080/14697680902814225
James G, Witten D, Hastie T et al (2021) An introduction to statistical learning: with applications in R, 2nd edn. Springer
Jaspersen J (2022) Convex combinations in judgment aggregation. Eur J Oper Res 299(2):780–794. https://doi.org/10.1016/j.ejor.2021.09.050
Juárez MA, Steel MFJ (2010) Model-based clustering of non-gaussian panel data based on skew-t distributions. J Bus Econ Stat 28(1):52–66. https://doi.org/10.1198/jbes.2009.07145
Kelly J (1956) A new interpretation of information rate. Bell Syst Tech J 35(11):1593–1602. https://doi.org/10.1002/j.1538-7305.1956.tb03809.x
Levy H, Markowitz H (1979) Approximating expected utility by a function of mean and variance. Am Econ Rev 69(3):308–317
Lewandowski D, Kurowicka D, Joe H (2009) Generating random correlation matrices based on vines and extended onion method. J Multivar Anal 100(9):1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008
MacLean L, Thorp E, Ziemba W (2011) Long-term capital growth: the good and bad properties of the Kelly and fractional Kelly capital growth criteria. Quant Finance 10(7):681–687. https://doi.org/10.1080/14697688.2010.506108
Makridakis S, Taleb NN (2009) Decision making and planning under low levels of predictability. Int J Forecast 25(4):716–733. https://doi.org/10.1016/j.ijforecast.2009.05.013
Markowitz H (1952) Portfolio selection. J Finance 7:77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
McNichols M, O'Brien PC (1997) Self-selection and analyst coverage. J Account Res 35:167–199. https://doi.org/10.2307/2491460
Merkle E, Saw G, Davis-Stober C (2020) Beating the average forecast: Regularization based on forecaster attributes. J Math Psychol 98:102419. https://doi.org/10.1016/j.jmp.2020.102419

Nasdaq (2022) Dividend history. Available at: https://www.nasdaq.com/market-activity/quotes/dividend-history
O'Hagan A, Buck C, Daneshkhah A et al (2006) Uncertain judgements: eliciting experts' probabilities. John Wiley & Sons, Chichester
Peters O (2011) Optimal leverage from non-ergodicity. Quant Finance 11(11):917–926. https://doi.org/10.1080/14697688.2010.513338
Piironen J, Vehtari A (2017) Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat 11(2):5018–5051. https://doi.org/10.1214/17-EJS1337SI
Praetz P (1972) The distribution of share price changes. J Bus 45(1):49–55
Pulley L (1981) A general mean-variance approximation to expected utility for short holding periods. J Financ Quant Anal 16(3):361–373. https://doi.org/10.2307/2330243
Satopää V, Salikhov M, Tetlock P et al (2021) Bias, information, noise: The bin model of forecasting. Manage Sci 67(12):599–7618. https://doi.org/10.1287/mnsc.2020.3882
Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Financ 19(3):425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
Sharpe WF (1994) The sharpe ratio. J Portf Manag 21(1):49–58. https://doi.org/10.3905/jpm.1994.409501
Smith J, Winkler R (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Manage Sci 52(3):311–322. https://doi.org/10.1287/mnsc.1050.0451
Stan Development Team (2023) RStan: the R interface to Stan. https://mc-stan.org/, r package version 2.21.8
Taleb NN (2020) On the statistical differences between binary forecasts and real-world payoffs. Int J Forecast 36(4):1228–1240. https://doi.org/10.1016/j.ijforecast.2019.12.004
Thorp E (1971) Portfolio choice and the Kelly criterion. In: Proceedings of the business and economics statistics section proceedings of the American statistical association, pp 215–224
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131. https://doi.org/10.1126/science.185.4157.1124
Vehtari A, Gelman A, Simpson D et al (2021) Rank-normalization, folding, and localization: an improved $\widehat{R}$ for assessing convergence of MCMC. Bayesian Anal 16(2):667–718. https://doi.org/10.1214/20-BA1221
Winkler RL, Clemen RT (2004) Multiple experts vs. multiple methods: Combining correlation assessments. Decis Anal 1(3):167–176. https://doi.org/10.1287/deca.1030.0008
Womack K (1996) Do brokerage analysts' recommendations have investment value? J Financ 51(1):137–167. https://doi.org/10.1111/j.1540-6261.1996.tb05205.x