OR Spectrum
https://doi.org/10.1007/s00291-024-00786-8

ORIGINAL ARTICLE

Bayesian learning in dynamic portfolio selection under a minimax rule

Xiaoqiang Cai[1] · **Gen Yu**[2]

Received: 19 January 2022 / Accepted: 20 August 2024

Abstract

We are concerned about a multi-period portfolio selection problem where the issue of parameter uncertainty for the distribution of risky asset returns should be addressed properly. For analysis, we first propose a novel dynamic portfolio selection model with an l_∞ risk function, instead of the classic portfolio variance, used as risk measure. The investor in our model is assumed to choose the optimal portfolio by maximizing the expected terminal wealth at a minimum level of cumulative risk, quantified by a weighted sum of the risks in subsequent periods. The proposed multi-period model has a closed-form optimal policy that can be constructed and interpreted intuitively. We introduce Bayesian learning to account for the uncertainty in estimates of unknown parameters and discuss the impact of Bayesian learning on the investor's decision making. Under an i.i.d. normal return-generating process with unknown means and covariance matrix, we show how Bayesian learning promotes diversification and reduces sensitivity of optimal portfolios to changes in model inputs. The numerical results based on real market data further support that the model with Bayesian learning can perform much better than a plug-in model out-of-sample with the extent of performance improvement affected by the investor's level of risk aversion and the amount of data available.

Keywords Investment analysis · Bayesian learning · Parameter uncertainty · l_∞ risk function · Stochastic dynamic programming

✉ Gen Yu
yugen@mail.shufe.edu.cn

Xiaoqiang Cai
xqcai@cuhk.edu.cn

1 The Chinese University of Hong Kong, Shenzhen & Shenzhen Research Institute of Big Data, Shenzhen, China

2 School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

Published online: 17 September 2024

1 Introduction

The seminal mean-variance model proposed by Markowitz (1952) lays down the foundation for the modern portfolio theory. As the name implies, expected returns, variances and covariances are the key input model parameters. In most real-world situations, these parameters are, nevertheless, not known a priori. Much work in this vein has to postulate that those unknown moments can be estimated precisely by sample averages, which is usually known as the plug-in approach. However, point estimates are subject to estimation errors and may result in portfolio allocations far from optimum. In fact, the problem of estimation risk will be even exacerbated in the mean-variance model as more and more researchers find the portfolios produced by the mean-variance model very sensitive to model parameters (see, e.g., Best and Grauer 1991a, b; Chopra et al. 1993; Chopra and Ziemba 1993), that is, a small fluctuation in the estimate might be amplified to a dramatic change in portfolio weights. Thus, the mean-variance model without considering parameter uncertainty is often criticized for entailing extreme positions in the optimal portfolio and delivering poor out-of-sample performances (see Litterman et al. 2004).

There is an extensive literature devoted to applying Bayesian approaches to relieve the adverse impacts of parameter uncertainty in portfolio optimization. Given available data and a portfolio choice problem, an investor might be inclined to discard potential features of uncertainties and make decisions based on nonparametric estimation methods (see Tsybakov 2008). The other extreme is the plug-in approach in which the investor is convinced about an underlying data-generating model that incorporates specific features for predicting the future, through a statistical test or a heuristic argument, assuming that the model parameters are exactly equal to the sample estimates and ignoring the estimation risk of unknown parameters. By contrast to the above two extremes, a Bayesian investor assumes a parametric return-generating model and treats unknown model parameters as random variables with specified priors. A posterior/predictive probability distribution of asset returns, which depends only on the observation of data, can be obtained by integrating out the unknown parameters according to the Bayes' rule and evolves automatically as new data released. The research work on this topic includes Frost and Savarino (1986), Aguilar and West (2000), Polson and Tew (2000), Pástor and Stambaugh (2000), Wang (2005), Black and Litterman (1992), Kolm and Ritter (2017), Zhou (2009), Bauder et al. (2021), Bodnar et al. (2017); Anderson and Cheng (2016) and Marisu and Pun (2023). The majority of the relevant literature is, however, limited in a static framework of the mean-variance model. One exception is Winkler and Barry (1975), who consider Bayesian inference and learning in a multi-period setting, in which the investor is assumed to maximize a utility function of the terminal wealth and the optimal strategy requires a case-by-case discussion. Their work first shows that even for the simple linear and quadratic utilities, the corresponding multi-period model with one risky asset and one risk-free asset involved can only be solved numerically rather than analytically in general.

As intuitively expected, especially for dynamic problems, it is important for the investor to exploit the data observed gradually, recognizing the updated information about the unknown parameters and revising the portfolio accordingly. Nevertheless, as indicated in Winkler and Barry (1975), if the formulated stochastic dynamic program cannot be solved analytically, the development of the optimal portfolio with information updating, which requires computation of conditional expectations and optimization at each time period, will be of great computational challenge even for simple investment cases. As a consequence, there have been few results in the literature on solutions for general dynamic portfolio optimization problems with unknown parameters.[1] On the other hand, recent studies achieve some progress by developing a variety of approximate solution methods. For instance, Barberis (2000) conducts backward induction through discretizing the state space. Brandt et al. (2005) take a Taylor series expansion over the expected utility to obtain an approximate closed-form solution. Soyer and Tanyeri (2006) adopt a surface fitting approach for a two-stage model. Skoulakis (2008) approximates the value functions using feedforward neural network. Jurek and Viceira (2010) obtain their approximate analytical solution by log-linearizing the budget constraint for the log-normal return distribution. Unfortunately, these papers suffer from several deficiencies. First, almost all the previous literature investigating Bayesian learning and dynamic portfolio choice relies on using utilities defined over terminal wealth as objective functions, e.g., the power utility and the exponential utility. Although they are popular in economic studies, investors and portfolio managers may be concerned more about explicit measurement of the investment risk, which cannot be easily seen and analyzed if a utility function of final wealth is used. Second, errors in the portfolios resulted from multi-period approximation and optimization are hard to detect and control, especially when the state variables are of a high dimension. Third, due to the absence of analytical solutions, the traditional models lack a clear interpretation of how Bayesian learning affects investors' decision making. The role of Bayesian learning in forming optimal portfolio allocations is still obscure.

In this paper, we focus on the investment needs of long-term conservative investors who are particularly risk averse. To appropriately model the decision features of those investors, we adopt an l_∞ risk function as the risk measure in our dynamic model formulation. Various risk measures have been proposed in the literature to capture the concerns of investors in different situations. These include, e.g., semivariance (see Markowitz 1959), value at risk (see Duffie and Pan 1997), CVaR (see Rockafellar et al. 2000), and l_1 risk function (see Konno and Yamazaki 1991). Cai et al. (2000) propose a more conservative l_∞ risk function to possibly reflect the risk attitude of those relatively conservative investors. Specifically, this l_∞ risk measure can be defined mathematically as follows:

$$l_\infty(\boldsymbol{x}) = \max_{1\leqslant j\leqslant p} \mathbb{E}\left[|r_j x_j - \mathbb{E}(r_j)x_j|\right],$$

[1] Li and Ng (2000), Yu et al. (2010) and Bodnar et al. (2015) present exact solutions for their dynamic portfolio selection models assuming no unknown parameters. In the continuous setting, Brennan (1998) and Xia (2001) possess analytical solutions in the context of Bayesian learning based on stylized modeling of return process. The results however are hard to generalize to the multi-period setting with general return-generating processes.

where r_j denotes the return rate of asset j, x_j is the amount of allocation of the fund to asset j, p is the total number of assets and $\boldsymbol{x}$ is a vector of the form $\boldsymbol{x} = (x_1, ..., x_p)^\top$. It is clear that the risk of holding a portfolio in one period under $l_\infty(\boldsymbol{x})$ is measured by the maximal expected absolute deviation of individual asset return from its expectation. Therefore, by minimizing the risk proxy $l_\infty(\boldsymbol{x})$ function, an investor can set up a minimax rule to construct the optimal portfolio. In the original work of Cai et al. (2000), the authors introduce the l_∞ risk function in a static portfolio setting and derive an analytical solution for a single-period model. By analyzing the feature of the optimal investment strategy, they show in theory that their model with l_∞ risk function can exhibit some robustness to the errors in the problem inputs. The empirical work in Cai et al. (2004) further supports that the portfolio derived from the l_∞ model is less sensitive to the input data compared with Markowitz's mean-variance model. More studies on l_∞ risk measure have been reported in the literature subsequently; see, e.g., Prigent (2007), Ryals et al. (2007), Park et al. (2019); Vercher and Bermúdez (2015); Sun et al. (2015) and Meng et al. (2022).

The contributions of this paper can be summarized as follows.

(1) We set up a novel dynamic portfolio selection framework in which the estimates of unknown parameters can be updated via Bayesian learning and the l_∞ risk function is used as the risk measure. Specifically, the investor in our model is assumed to choose the optimal portfolio by (i) maximizing the expected terminal wealth; and (ii) minimizing the cumulative investment risk which is defined as a weighted sum of risks the investor will undertake in subsequent periods. We show that the proposed stochastic dynamic program has a closed-form optimal policy, independent of assumptions on the return-generating process. In contrast to previous studies relying on approximate solution methods, our work gives an analytical expression of the optimal portfolio allocation, making it possible to see clearly how the composition of the portfolio is determined and how Bayesian learning affects investor's decision in a dynamic setting. The investment strategy in each period can be intuitively viewed as a three-step decision scheme. First, we rank the individual assets in terms of their expected returns adjusted by available information and anticipation of future decisions. Then, we select assets to be invested by checking a sequence of inequality rules that exploit the information contained in the adjusted expected returns and risks. Finally, the actual amount to be allocated to those selected assets are computed on the basis of the current wealth and their risks, i.e., the mean absolute deviations of individual asset returns. For implementation of the optimal policy, we introduce a least squares Monte Carlo method to approximate complex conditional expectations.

(2) Under an i.i.d. normal return-generating process with unknown means and covariance matrix, we find that, besides providing a formal way to accommodate new information from observed data, Bayesian learning can also help diversify[2] portfolio allocations and reduce sensitivity of optimal portfolios to changes in model inputs. The major insight behind the properties is that incorporating the estimation

[2] Throughout the paper, we will use portfolio size (the number of assets with positive weights) as the main criterion in the discussion of portfolio diversification level. That is, in our analysis, a portfolio is more diversified when it includes more assets with positive investment.

risk of unknown parameters via Bayesian learning makes risky assets more risky, which leads to a more conservative and robust investment policy under our model framework. A typical message from the previous literature is that the use of Bayesian learning induces a negative effect on portfolio weights of risky assets (see Barberis 2000; Brandt et al. 2005; Skoulakis 2008). Barberis (2000) also mentioned that in their case, Bayesian learning affects the sensitivity of the optimal allocation to state variables. However, these findings are mainly based on numerical observations, without clarifying the mechanism or the role of Bayesian learning in forming optimal portfolios. In contrast, our paper aims at interpreting the effects of Bayesian learning from the policy layer in multi-period portfolio selection problems. Our conclusions are consistent with numerical findings in related literature, but we enrich the understanding of Bayesian learning in a new dynamic model scheme.

(3) Our numerical results based on real market data indicate that compared with a plug-in model, using Bayesian learning to account for parameter uncertainty and estimation risk in dynamic portfolio selection problems is able to improve the policy's out-of-sample performance. The performance gap between models with and without Bayesian learning is, however, affected by investor's risk preference and the amount of data available. That is, we observe that the incorporation of Bayesian learning has a significant advantage in out-of-sample performance when the investor's risk tolerance level is high or the amount of available data is small. As the investor becomes more conservative and more data are observed, the gap between the two models narrows. We believe that these findings are valuable in answering the questions on the performance of considering Bayesian learning in practical investments.

The remainder of this paper is organized as follows. In Sect. 2, we formulate the dynamic portfolio selection model with unknown parameters. We solve the proposed stochastic dynamic program in Sect. 3. A plug-in model without learning is discussed in Sect. 4. An empirical study is provided in Sect. 5. We conclude our work in Sect. 6. The proofs of theorems and propositions are included in the "Appendix".

2 A multi-period portfolio selection model under minimax rule

Assume that there is a capital market with p risky assets, $S_1, S_2, ..., S_p$. The decision time is discrete and indexed by $\{0, 1, ..., T-1\}$. An investor joins this market with initial fund V_0 and he can reallocate his fund among these p assets at the beginning of each of the following T consecutive time periods. The returns of risky assets are stochastic and denoted by $\boldsymbol{r}_t = (r_t^1, r_t^2 ..., r_t^p)^\top$, where r_t^j is the return for S_j in period t. Throughout the paper, we use boldface upper and lower case characters to denote vectors and matrices, respectively. $\mathbf{1}$ is the vector of all ones. We use $[t; T]$ to denote an index set $\{t, t+1, ..., T\}$ for short. The monetary values of the risky asset holdings at the beginning of period t are described by a vector $\boldsymbol{x}_t = (x_t^1, ..., x_t^p)^\top$. Let V_t be the wealth at the beginning of period t. Then, the budget constraint is $\sum_{j=1}^p x_t^j = V_t$ for all $t \in [0;T-1]$. Short-selling and transaction cost are not considered.

Uncertainty in risky asset returns is analyzed based on a probability space $(\Omega, \mathcal{F}, P)$ where Ω is the set of possible outcomes with element ω, $\mathcal{F}$ is a σ-algebra and P is a probability measure. The flow of information is modeled by a filtration $\{\mathcal{F}_t\}_{t=0}^T$, where the σ-algebra $\mathcal{F}_t$ describes the information available to the investor at the beginning of period t satisfying $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}_T = \mathcal{F}$ for all $t < T$. Therefore, the expectation operator conditional on available information can be formally defined as $\mathbb{E}(\cdot|\mathcal{F}_t)$, abbreviated as $\mathbb{E}_t(\cdot)$.

In order to evaluate the risk of holding a portfolio $\boldsymbol{x}_t$ in period t, we define a single-period portfolio risk measure based on the l_∞ risk function as

$$l_t(\boldsymbol{x}_t) = \max_{1\leqslant j\leqslant p} \mathbb{E}_t(|r_t^j - \mathbb{E}_t(r_t^j)|)x_t^j = \max_{1\leqslant j\leqslant p} \mathbb{E}_t(|r_t^j - m_t^j|)x_t^j = \max_{1\leqslant j\leqslant p} q_t^j x_t^j, \tag{1}$$

where we use $m_t^j = \mathbb{E}_t(r_t^j)$ and $q_t^j = \mathbb{E}_t(|r_t^j - m_t^j|)$ to denote the (conditional) expected return and its mean absolute deviation (MAD), respectively. The corresponding vector forms are $\boldsymbol{m}_t = \mathbb{E}_t(\boldsymbol{r}_t)$ and $\boldsymbol{q}_t = \mathbb{E}_t(|\boldsymbol{r}_t - \boldsymbol{m}_t|)$. Note that under our notation, r_t^j is $\mathcal{F}_{t+1}$-measurable, and m_t^j, q_t^j and $l_t(\boldsymbol{x}_t)$ are adapted to $\mathcal{F}_t$. Strictly speaking, since m_t^j, q_t^j and $l_t(\boldsymbol{x}_t)$ are measurable functions mapping from the sample space Ω to real line $\mathbb{R}$, their complete forms should be $m_t^j(\omega)$, $q_t^j(\omega)$ and $l_t(\boldsymbol{x}_t, \omega)$. For expositional brevity, we will suppress henceforth the dependency on the element $\omega \in \Omega$ for functions measurable with respect to $\{\mathcal{F}_t\}_{t=0}^T$. Instead, we point out the measurability when necessary. Generally, the dependency could be inferred easily from the context.

By minimizing the risk $l_t(\boldsymbol{x}_t)$ and maximizing the one-period expected return, the so called minimax rule in the single-period model can be formulated as follows:

$$\min_{\boldsymbol{x}_t \in \mathcal{X}_t} \quad \lambda\left(\max_{1\leqslant j\leqslant p} q_t^j x_t^j\right) - (1-\lambda)\sum_{i=1}^{p} m_t^j x_t^j, \tag{2}$$

where $\mathcal{X}_t = \left\{\boldsymbol{x}_t : \sum_{j=1}^p x_t^j = V_t, x_t^j \geqslant 0, j \in [1;p]\right\}$ and $\lambda \in (0,1)$ represents the investor's risk aversion level — the larger the λ, the more conservative the investor is. In multi-objective optimization (to maximize the expected return and minimize the risk), the optimal solution of Problem (2) with respect to a given value of λ corresponds to an efficient solution point of the bi-criteria problem in (2) that considers both risk and return, and the set of optimal solutions with respect to all $\lambda \in (0,1)$ correspond to the efficient frontier. Prior to solving the Problem (2), we first define an ancillary function $G(\boldsymbol{a}_1, \boldsymbol{a}_2, k)$ with input vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and scalar $k \in [0;p-1]$ such that for all $k \in [1;p-1]$,

$$G(\boldsymbol{a}_1, \boldsymbol{a}_2, k) = \sum_{j=0}^{k-1} \frac{a_1^{i_{p-j}(\boldsymbol{a}_1)} - a_1^{i_{p-k}(\boldsymbol{a}_1)}}{a_2^{i_{p-j}(\boldsymbol{a}_1)}}$$

and specially $G(\cdot,\cdot,0) = 0$, where the function $i_j(\boldsymbol{a}_1)$ outputs the index of the jth largest element of a given vector $\boldsymbol{a}_1 = (a_1^1, \ldots, a_1^p)^\top$, i.e., $a_1^{i_1(\boldsymbol{a}_1)} \leqslant a_1^{i_2(\boldsymbol{a}_1)} \leqslant \ldots \leqslant a_1^{i_p(\boldsymbol{a}_1)}$. Then, the optimal portfolio allocation of Problem (2) can be presented in the following lemma.

Lemma 1 *Given $\lambda \in (0, 1)$, the optimal solution of Problem (2) is that*

$$x_t^{j*} = \begin{cases} \dfrac{V_t}{q_t^j}\left(\displaystyle\sum_{j\in\mathcal{A}_t^*}\frac{1}{q_t^j}\right)^{-1}, & j \in \mathcal{A}_t^*, \\ 0, & j \notin \mathcal{A}_t^*, \end{cases} \tag{3}$$

where the set of assets in which to invest $\mathcal{A}_t^*$ can be constructed by the following rule: If there exists an integer $k \in [0;p-2]$ such that

$$G(\boldsymbol{m}_t, \boldsymbol{q}_t, k) < \frac{\lambda}{1-\lambda} \quad \text{and} \quad G(\boldsymbol{m}_t, \boldsymbol{q}_t, k+1) \geqslant \frac{\lambda}{1-\lambda}, \tag{4}$$

then $\mathcal{A}_t^* = \{i_p(\boldsymbol{m}_t), i_{p-1}(\boldsymbol{m}_t), ..., i_{p-k}(\boldsymbol{m}_t)\}$. Otherwise, if the condition above is not satisfied by any integer $k \in [0;p-2]$, then $\mathcal{A}_t^* = [1;p]$.

Lemma 1 can be obtained by solving a set of equations from standard KKT conditions. Readers of interest could refer to the appendix in Cai et al. (2000) for more details. The primary difference between Problem (2) and the original model in Cai et al. (2000) is that Problem (2) considers estimation risk and uses conditional expectations to highlight the value of information flow, while the original model only focuses on point estimates and assumes that the unknown parameters can be estimated precisely.

Despite the wide popularity of single-period models, this static paradigm is known to be difficult to apply to the long-term investors such as pension planners and insurance companies (Mulvey et al. 2003). Moreover, like a wise chess player, it is reasonable to suggest that a successful manager or investor always think ahead and contemplate the inter-temporal effect of multi-period decisions. We next formulate a dynamic model considering both portfolio optimization and Bayesian learning. Let $l_t(\boldsymbol{x}_t)$ in (1) quantify the portfolio risk in period t. We then define the cumulative risk during the investment horizon as follows:

$$L_{t:T} = \mathbb{E}_t\left[\sum_{k=t}^{T-1}\gamma_k l_k(\boldsymbol{x}_k)\right] = \mathbb{E}_t\left[\sum_{k=t}^{T-1}\gamma_k \max_{1\leqslant j\leqslant p} q_k^j x_k^j\right], \quad t = 0, 1, ..., T-1, \tag{5}$$

where $\gamma_k \geqslant 0$ is the weight for each period risk $l_k(\boldsymbol{x}_k)$. By tuning the values of $\{\gamma_k\}_{k=t}^{T-1}$, one can distinguish the importance of risks in different periods.

Analogous to the single-period model (2), we now extend the minimax rule to the multi-period case and propose the dynamic model as follows:

$$\begin{aligned} &\min_{\boldsymbol{x}_0\in\mathcal{X}_0}\mathbb{E}_0\left[\min_{\boldsymbol{x}_1\in\mathcal{X}_1}\mathbb{E}_1\left[\ldots\min_{\boldsymbol{x}_{T-1}\in\mathcal{X}_{T-1}}\mathbb{E}_{T-1}\left[\lambda L_{0:T} - (1-\lambda)V_T\right]\ldots\right]\right] \\ &\text{s.t.} \quad V_{t+1} = V_t + \boldsymbol{r}_t^\top \boldsymbol{x}_t, \quad t \in [0;T-1]. \end{aligned} \tag{6}$$

Several clarifications on Problem (6) are in order. First, the cumulative risk defined in (5) is a natural way to evaluate the total risk the investor undertakes during T periods, and it is widely acceptable in multi-period portfolio optimization literature, see, e.g., Calafiore (2008), Liu and Zhang (2015) and Boyd et al. (2017). Second, the modeling of (6) indicates that the optimal policy should be derived by maximizing the expected final wealth at a minimum level of cumulative risks. Hence, our model formulation is appropriate for the investors who care both final wealth at the expiration and the risks they will undertake during the investment process. The non-negative weights considered $\{\gamma_t\}_{t=0}^{T-1}$ provide extra model flexibility for users in practice. Third, in Problem (6), the investor chooses the optimal portfolio at each decision point with the possibility of rebalancing portfolios in future periods and recognizes that data realizations during the investment horizon contain useful information for updating beliefs on unknown parameters. This decision making process with learning can be described intuitively as follows:

$$\text{Decision}(\boldsymbol{x}_0^*) \rightsquigarrow \text{Observation}(\mathcal{F}_1) \rightsquigarrow \text{Learning} \rightsquigarrow \text{Decision}(\boldsymbol{x}_1^*) \rightsquigarrow \cdots \rightsquigarrow \text{Observation}(\mathcal{F}_{T-1}) \rightsquigarrow \text{Learning} \rightsquigarrow \text{Decision}(\boldsymbol{x}_{T-1}^*).$$

Prior to solving Problem (6), we further investigate the time-consistency property from the perspectives of the multi-period risk measure $L_{t:T}$ and the problem's optimal policy, respectively.

Let $X_k = \gamma_k \max_{1 \leqslant j \leqslant p} q_k^j x_k^j$. The sequence $\bar{X}_t = \{X_k\}_{k=t}^{T-1}$ can be viewed as the loss process with X_k's realization the lower the better, and each X_k is adapted to $\mathcal{F}_k$. Then, $L_{t:T}(\bar{X}_t) = \mathbb{E}_t(\sum_{k=t}^{T-1} X_k)$, that is, a conditional expectation of a sum of future losses. Given $0 \leq t_1 < t_2 \leq T-1$ and loss processes $\bar{X}_{t_1}$ and $\bar{X}'_{t_1}$, if $X_k = X'_k$, $\forall k = t_1, \dots, t_2 - 1$, and $L_{t_2:T}(\bar{X}_{t_2}) \leq L_{t_2:T}(\bar{X}'_{t_2})$, where the equality and inequality between random variables are understood in the almost surely sense, then, given the conditional expectation in $L_{t_1:T}$, it is straightforward to have $L_{t_1:T}(\bar{X}_{t_1}) \leq L_{t_1:T}(\bar{X}'_{t_1})$. Therefore, the multi-period risk measure $L_{t:T}$ is time consistent in the sense that if a position is riskier than another one at some future time (t_2), then the position should also be riskier from the perspective today (t_1). For more discussion on time consistent dynamic risk measures, see Ruszczyński (2010) and references therein.

For the multi-period stochastic programming Problem (6) with further terminal wealth considered, we follow the concept in Shapiro (2009) and time-consistency means that any optimal policy specified today would imply its optimality in future stages. By expanding V_T and $L_{0:T}$, Problem (6) can be reformulated as

$$\min_{\boldsymbol{x}_0 \in \mathcal{X}_0} -(1-\lambda)V_0 + g_0(\boldsymbol{x}_0) + \mathbb{E}_0\left[\min_{\boldsymbol{x}_1 \in \mathcal{X}_1} g_1(\boldsymbol{x}_1) + \mathbb{E}_1\left[\dots + \mathbb{E}_{T-2}\left[\min_{\boldsymbol{x}_{T-1} \in \mathcal{X}_{T-1}} g_{T-1}(\boldsymbol{x}_{T-1})\right]\dots\right]\right],$$

where V_0 is known at the beginning of period 0 and $g_t(\boldsymbol{x}_t) = \lambda\gamma_t \max_{1 \leqslant j \leqslant p} q_t^j x_t^j - (1-\lambda)\mathbb{E}_t[\boldsymbol{r}_t^\top \boldsymbol{x}_t]$ is $\mathcal{F}_t$-measurable. According to Example 2 of Shapiro (2009), we can conclude that Problem (6) is time consistent.

3 Optimal investment policy with learning

In this section, we first solve the Problem (6) and give some structural results on the optimal policy. Then, we set up a Bayesian learning framework under an i.i.d. normal return-generating process with unknown means and covariance matrix. We finally introduce a least squares Monte Carlo method to estimate complex conditional expectations for implementation of the optimal policy.

3.1 An optimal investment policy

It turns out the optimal policy of Problem (6) can be solved analytically. We directly give results in Theorem 1 and defer its proof in "Appendix A".

Theorem 1 *Given non-negative* $\{\gamma_t\}_{t=0}^{T-1}$ *and* $\lambda \in (0,1)$, *the optimal policy of Problem* (6) *is such that for each* $t \in [0;T-1]$,

$$x_t^{j*} = \begin{cases} \dfrac{V_t}{q_t^j}\left(\displaystyle\sum_{j\in\mathcal{A}_t^*}\frac{1}{q_t^j}\right)^{-1}, & j \in \mathcal{A}_t^*, \\ 0, & j \notin \mathcal{A}_t^*, \end{cases} \tag{7}$$

where the set of assets in which to invest $\mathcal{A}_t^*$ *is determined by the following rule: When* $\gamma_t > 0$, *if there exists an integer* $k \in [0;p-2]$ *such that*

$$G(\boldsymbol{v}_t, \gamma_t\boldsymbol{q}_t, k) < \frac{\lambda}{1-\lambda} \quad \text{and} \quad G(\boldsymbol{v}_t, \gamma_t\boldsymbol{q}_t, k+1) \geqslant \frac{\lambda}{1-\lambda}, \tag{8}$$

then $\mathcal{A}_t^* = \{i_p(\boldsymbol{v}_t), i_{p-1}(\boldsymbol{v}_t), \dots, i_{p-k}(\boldsymbol{v}_t)\}$; *otherwise,* $\mathcal{A}_t^* = [1;p]$. *When* $\gamma_t = 0$, $\mathcal{A}_t^* = \{i_p(\boldsymbol{v}_t)\}$. *The vector* $\boldsymbol{v}_t$ *is recursively defined as: For* $t = T-1$, $\boldsymbol{v}_{T-1} = \mathbb{E}_{T-1}(\boldsymbol{r}_{T-1})$ *and* $c_{T-1} = -(1-\lambda)$. *For* $t \in [0;T-2]$, $\boldsymbol{v}_t = -\mathbb{E}_t[(c_{t+1} + \lambda\gamma_{t+1}z_{t+1} - (1-\lambda)y_{t+1})\boldsymbol{r}_t]/(1-\lambda)$ *and* $c_t = \mathbb{E}_t[c_{t+1} + \lambda\gamma_{t+1}z_{t+1} - (1-\lambda)y_{t+1}]$ *where in each stage* $t \in [0;T-1]$,

$$y_t = z_t \sum_{j\in\mathcal{A}_t^*} \frac{v_t^j}{q_t^j} \quad \text{and} \quad z_t = \left(\sum_{j\in\mathcal{A}_t^*}\frac{1}{q_t^j}\right)^{-1}. \tag{9}$$

The optimal policy derived in Theorem 1 is nonanticipative in that V_t, $\boldsymbol{v}_t$ and $\boldsymbol{q}_t$ are $\mathcal{F}_t$-measurable and decisions in $\mathcal{A}_t^*$ and $\boldsymbol{x}_t^*$ depend only on what is known at the beginning of period t. Moreover, the given policy can be intuitively viewed as a three-step decision scheme. First, we rank the p risky assets in terms of their values in the vector $\boldsymbol{v}_t$. Then, we select a set of assets to be included in the portfolio $\mathcal{A}_t^*$ by checking a sequence of inequalities based on their values in $\boldsymbol{v}_t$ and $\boldsymbol{q}_t$. The actual amount allocated from V_t to those selected assets depends on the MADs of their returns in $\boldsymbol{q}_t$ following Eq. (7). Structurally, the policy in Theorem 1 can be decomposed into a selection rule in (8) and an allocation rule in (7), and it has a connection with the single-period solution in Lemma 1. To be specific, regarding the selection rule to determine $\mathcal{A}_t^*$, the inequalities in (4) shows that an asset with a high expected return is always considered prior to an asset with a low expected return and the asset with the highest value in $\boldsymbol{m}_t$ is always included in $\mathcal{A}_t^*$ (but the actual allocation to the selected asset may be close to zero if its risk is high). The policy in Theorem 1 acts in a similar way, but the selection criterion now depends on the comparison of values in $\boldsymbol{v}_t$ instead of the simple expected asset returns and the MADs in function G are weighted by $\{\gamma_t\}_{t=0}^{T-1}$.

As shown in Theorem 1, the vector $\boldsymbol{v}_t$ is the crux to implement the optimal policy. The calculation of $\boldsymbol{v}_t$, however, is far from straightforward. We will introduce a least squares Monte Carlo method in Sect. 3.3 to approximate its value. For now, we deviate and give the following proposition to better understand the meaning of $\boldsymbol{v}_t$ in Theorem 1.

Proposition 1 *For $t \in [0;T-2]$, let*

$$\Delta_{t+1} = \sum_{s=t+1}^{T-1} \frac{\boldsymbol{v}_s^\top \boldsymbol{x}_s^*}{V_s} - \frac{\lambda \gamma_s}{1-\lambda} z_s, \tag{10}$$

where $\boldsymbol{x}_s^*$ denotes the optimal asset holdings at the beginning of period s. Let $\Delta_T = 0$. Then, for all $t \in [0;T-1]$, $\boldsymbol{v}_t$ can be written as

$$\boldsymbol{v}_t = \mathbb{E}_t[(1+\Delta_{t+1})\boldsymbol{r}_t]. \tag{11}$$

An important observation in (11) is that the vector $\boldsymbol{v}_t$ can be regarded as adjusted expected returns (AERs) of risky assets because it is composed of expected returns and adjusted in two perspectives: (i) It is adjusted by the anticipation of future decisions. This type of adjustment is reflected in Δ_{t+1} which is obtained by anticipating the performance of the optimal portfolios $\{\boldsymbol{x}_s^*\}_{s=t+1}^{T-1}$ in future periods. Precisely, in (10), z_s, as defined in (9), generally measures the overall risk of investing the assets in $\mathcal{A}_s^*$, while $\boldsymbol{v}_s^\top \boldsymbol{x}_s^*/V_s$ represents the generalized return rate in period s, evaluated in terms of AERs $\boldsymbol{v}_s$. Thus, this anticipation effect takes account of both risk and return, and it accumulates across all future periods. (ii) It is adjusted by new observed data. This adjustment involves the calculation of conditional expectation $\mathbb{E}_t(\cdot)$. When new data are released, the distribution of $\boldsymbol{r}_t$ as well as future prospects in Δ_{t+1} will be updated accordingly in a Bayesian fashion. Therefore, the impact of future moves from dynamic programming and the role of learning can be well reflected given $\boldsymbol{v}_t$ presented in (11).

Special care should also be paid to the choice of the weights $\{\gamma_t\}_{t=0}^{T-1}$. Investors are allowed to impose personal preference on risks in different periods by setting $\{\gamma_t\}_{t=0}^{T-1}$ properly. For example, if the final period risk really matters, a portfolio manager may raise the value of γ_{T-1}. Increasing γ_{T-1} alone will make $\mathcal{A}_{T-1}^*$ include more assets since G function is decreasing in γ_{T-1} given $\boldsymbol{v}_{T-1}$, $\boldsymbol{q}_{T-1}$ and k. Moreover, for all previous stages $t \in [0;T-2]$, the weight of z_{T-1} will be increased accordingly as shown in (10). As another extreme, if one totally ignores the risk in period $T-1$, he could set $\gamma_{T-1} = 0$. Theorem 1 tells us that in this case, the optimal decision for stage $T-1$ is to invest all the fund V_{T-1} in the asset with the highest value in expected returns $\boldsymbol{m}_{T-1}$, which is not surprising for such a risk-seeking investor.

Similarly, for all previous stages $t \in [0;T-1]$, the effect of z_{T-1} would be removed by setting $\gamma_{T-1} = 0$ in (10).

Note that Theorem 1 solves a general stochastic dynamic programming problem proposed in (6) and thus the derived nonanticipative policy in Theorem 1 is optimal with respect to arbitrary probability measure P that dominates the return process $\{\boldsymbol{r}_t\}_{t=0}^{T-1}$. That is, Theorem 1 provides a general optimal policy without specifying how to compute the conditional expectations and how to update the information. For practical implementation of the policy, we need to further introduce some necessary assumptions on returns and unknown parameters, and develop procedures to learn from data. Next, we will restrict our attention to a parameterized return-generating process and introduce unknown parameters as well as Bayesian learning.

3.2 Bayesian learning framework

In this section, we specify a parameterized return-generating process and further show how the unknown parameters are updated in a Bayesian fashion.

For concentration on parameter uncertainty and estimation risk, we conduct our analysis based on a popular i.i.d. normal return-generating process.[3] Specifically, we assume that the return rates of p risky assets in period t, $t \in [0;T-1]$, follow a linear model given by

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \quad \text{with} \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \tag{12}$$

where $\boldsymbol{\epsilon}_0$, $\boldsymbol{\epsilon}_1$,..., $\boldsymbol{\epsilon}_{T-1}$ are i.i.d. noises. Given true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the returns $\{\boldsymbol{r}_t\}_{t=0}^{T-1}$ are independently and identically distributed as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In most realistic situations, however, the investor cannot know the exact true values of parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In spirit of Bayesian learning, we first suppose that the unknown parameters are random and follow a specified prior distribution, then according to Bayesian rule, the posterior beliefs on probability distributions of unknown parameters could be updated gradually as new data are observed.

To start up, suppose there are h data points before the investment horizon denoted as $\{\boldsymbol{r}_{-h}, ..., \boldsymbol{r}_{-1}\}$. The information set $\mathcal{F}_t$ now can be formally defined as the σ-algebra generated by returns up to time t, that is, $\mathcal{F}_t = \sigma(\{\boldsymbol{r}_{-h}, ..., \boldsymbol{r}_{-1}, \boldsymbol{r}_0, ..., \boldsymbol{r}_{t-1}\})$. Denote the set of return data by $\mathcal{D}_t = \{\boldsymbol{r}_{-h}, ..., \boldsymbol{r}_{-1}, \boldsymbol{r}_0, ..., \boldsymbol{r}_{t-1}\}$. Thus, the conditional expectation $\mathbb{E}(\cdot|\mathcal{F}_t)$ can be written as $\mathbb{E}(\cdot|\mathcal{D}_t)$. This transition facilitates us to deal with conditional expectations defined directly in terms of data process rather that the sequence of σ-algebras. The computation work of Bayesian updating is presented in what follows.

Suppose that the mean vector $\boldsymbol{\mu}$ is unknown while the covariance matrix $\boldsymbol{\Sigma}$ is known. The uncertainty about $\boldsymbol{\mu}$ is further assumed to be described by a multi-variate normal prior, i.e., $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_{-1}, \boldsymbol{\Sigma}_{-1})$. According to Bayesian rule, it can be verified that

[3] Our proposed dynamic model can be easily adapted to more complex considerations, such as return predictability (Barberis 2000), time-varying volatility (Lan 2014; Johannes et al. 2014) or even model uncertainty (Tu and Zhou 2004), with the nice property of analytical solutions maintained.

$\boldsymbol{r}_t|\mathcal{D}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\Sigma}^{-1}\boldsymbol{r}_{t-1} + \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{\mu}_{t-1})$ and $\boldsymbol{\Sigma}_t = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{t-1}^{-1})^{-1}$ for $t > 0$. Specially, $\boldsymbol{\mu}_0 = \boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}^{-1}\sum_{s=-h}^{-1}\boldsymbol{r}_s + \boldsymbol{\Sigma}_{-1}^{-1}\boldsymbol{\mu}_{-1})$ and $\boldsymbol{\Sigma}_0 = (h\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_{-1}^{-1})^{-1}$.

A more general case is that both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. To express some general and objective information on $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ before the realization of data, we assume a conventional "uninformative" prior for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, namely,

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}. \tag{13}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of the matrix $\boldsymbol{\Sigma}$. At the beginning of period 0, one can obtain the posterior distribution after observing h historical data, following the analysis in Zellner (1996). That is, given $\mathcal{D}_0$, we have

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \mathcal{D}_0 \sim \mathcal{N}\big(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/h\big) \quad \text{and} \quad \boldsymbol{\Sigma}|\mathcal{D}_0 \sim \mathcal{IW}\big(h-1, \boldsymbol{\Sigma}_0\big), \tag{14}$$

where $\mathcal{IW}(h-1, \boldsymbol{\Sigma}_0)$ is a inverse Wishart distribution with degree of freedom $(h-1)$ and scale matrix $\boldsymbol{\Sigma}_0$, $\boldsymbol{\mu}_0 = \sum_{s=-h}^{-1}\boldsymbol{r}_s/h$, $\boldsymbol{\Sigma}_0 = (\boldsymbol{R}_0 - \mathbf{1}\boldsymbol{\mu}_0^\top)^\top(\boldsymbol{R}_0 - \mathbf{1}\boldsymbol{\mu}_0^\top)$ and $\boldsymbol{R}_0 = (\boldsymbol{r}_{-h}, ..., \boldsymbol{r}_{-1})^\top$. It is well-known that the normal-inverse-Wishart is a conjugate prior with respect to the normally distributed data. At the beginning of period $t > 0$, we take the posterior in stage $t-1$ as the new prior for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and solve the updated posterior as follows:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \mathcal{D}_t \sim \mathcal{N}\big(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}/(h+t)\big) \quad \text{and} \quad \boldsymbol{\Sigma}|\mathcal{D}_t \sim \mathcal{IW}\big(h+t-1, \boldsymbol{\Sigma}_t\big), \tag{15}$$

where

$$\boldsymbol{\mu}_t = \frac{h+t-1}{h+t}\boldsymbol{\mu}_{t-1} + \frac{1}{h+t}\boldsymbol{r}_{t-1}, \quad \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1} + \frac{h+t-1}{h+t}(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}_{t-1})(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}_{t-1})^\top. \tag{16}$$

For posterior marginal distribution of $\boldsymbol{\mu}$, according to the results in Zellner (1996), one can further obtain that

$$\boldsymbol{\mu}|\mathcal{D}_t \sim \mathcal{T}_{h+t-p}\left(\boldsymbol{\mu}_t, \frac{1}{(h+t)(h+t-p)}\boldsymbol{\Sigma}_t\right), \tag{17}$$

which turns out to be a multivariate t-distribution with degree of freedom $(h+t-p)$, location vector $\boldsymbol{\mu}_t$ and shape matrix $\boldsymbol{\Sigma}_t/[(h+t)(h+t-p)]$. The predictive distribution of $\boldsymbol{r}_t$ is then given by

$$\boldsymbol{r}_t|\mathcal{D}_t \sim \mathcal{T}_{h+t-p}\left(\boldsymbol{\mu}_t, \frac{h+t+1}{(h+t)(h+t-p)}\boldsymbol{\Sigma}_t\right). \tag{18}$$

Based on posterior marginal distributions of unknown parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in (15) and (17), one can compute the first and second moments of $\mu^j|\mathcal{D}_t$ and $(\boldsymbol{\Sigma})_{jj}|\mathcal{D}_t$ for each $j \in [1;p]$ as follows,

$$\mathbb{E}_t(\mu^j) = \mu_t^j, \quad \mathbb{E}_t((\boldsymbol{\Sigma})_{jj}) = \frac{1}{h'}(\boldsymbol{\Sigma}_t)_{jj}, \tag{19}$$

$$\mathrm{Var}(\mu^j|\mathcal{D}_t) = \frac{(\boldsymbol{\Sigma}_t)_{jj}}{(h+t)h'} \quad \text{and} \quad \mathrm{Var}((\boldsymbol{\Sigma})_{jj}|\mathcal{D}_t) = \frac{2(\boldsymbol{\Sigma}_t)_{jj}^2}{(h'-2)h'^2}, \tag{20}$$

where $(\cdot)_{jj}$ denotes the jth diagonal element of the given matrix and $h' = h + t - p - 2$.

To implement the policy with Bayesian learning, we need the following proposition to update $\boldsymbol{m}_t$ and $\boldsymbol{q}_t$ as new data released.

Proposition 2 *Suppose both* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *are unknown and the initial prior follows* (13). *At the beginning of period* $t \in [0;T-1]$, *we have*

$$\boldsymbol{m}_t = \boldsymbol{\mu}_t \quad \text{and} \quad \boldsymbol{q}_t = \frac{2\sqrt{(h+t+1)\mathrm{diag}(\boldsymbol{\Sigma}_t)}}{\sqrt{h+t}(h+t-p-1)B\left(\frac{h+t-p}{2}, \frac{1}{2}\right)}, \tag{21}$$

where $\mathrm{diag}(\cdot)$ *is the operator that takes diagonal elements from the given matrix and* $B(\cdot,\cdot)$ *is a beta function.*

According to the definition of $l_t(\boldsymbol{x}_t)$ in (1) and the result in (21), it is clear that the optimal portfolio of our proposed model dose not depend on the covariances between assets. Interestingly enough, the total portfolio variance, $\boldsymbol{x}_t^\top \boldsymbol{\Sigma} \boldsymbol{x}_t$, can be related to our model, albeit in an implicit way. The formal statement is presented in the following proposition.

Proposition 3 *Suppose* $\boldsymbol{r}_t|\mathcal{D}_t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *for all* $t \in [0;T-1]$. *For arbitrary* $\xi > 0$ *and* $\boldsymbol{x}_t \in \mathcal{X}_t$, *it holds that*

$$2\left[1 - \Phi\left(\frac{\xi}{\sqrt{\boldsymbol{x}_t^\top \boldsymbol{\Sigma} \boldsymbol{x}_t}}\right)\right] \leqslant \frac{p}{\xi} l_t(\boldsymbol{x}_t),$$

where $\Phi(\cdot)$ *denotes the cumulative density function of standard normal distribution and* $l_t(\boldsymbol{x}_t)$ *is defined in* (1).

The inequality in Proposition 3 shows that under the given assumptions, the total portfolio variance, $\boldsymbol{x}_t^\top \boldsymbol{\Sigma} \boldsymbol{x}_t$, that explicitly contains the covariance matrix of asset returns, will be small if $l_t(\boldsymbol{x}_t)$ is kept small (nevertheless, it may not be true the other way around).

As pointed out by Best and Grauer (1991a), the more highly correlated the asset returns, the more sensitive the portfolio holdings from the mean-variance model are to expected returns. Hence, although it is intuitive to consider covariances in making portfolio decisions, an investment policy that removes the explicit dependence on asset covariances in the allocation part, such as the dynamic model proposed in this paper, may benefit from this "counter-intuitive" property in out-of-sample performance. We will provide more numerical evidence later to confirm the potential benefits of this property. It should be noted that the purpose of this paper is not to show whether the

l_∞ risk function is universally better than variance. Instead, we aim to offer alternative investment models different from the traditional mean-variance and utility-based models for heterogeneous users, especially for those relatively long-term conservative investors, and also to provide some insights for handling parameter uncertainty in multi-period problems via incorporating Bayesian learning.

3.3 Least squares Monte Carlo method

According to the analysis in previous subsections, the adjusted expected return vector $\boldsymbol{v}_t$ is one of the keys to obtaining the optimal policy. However, with learning in the dynamic model, its value involves conditional expectations of quantities in future periods and cannot be computed directly. Instead, we introduce a least squares Monte Carlo method to obtain an estimate of $\boldsymbol{v}_t$. This numerical method is first proposed for pricing American options in Longstaff and Schwartz (2001) and Tsitsiklis and Van Roy (2001). Recently, it has been applied to dynamic portfolio selection problems to estimate complex conditional expectations, e.g., Brandt et al. (2005), van Binsbergen and Brandt (2007), Diris et al. (2014), Lan (2014), Denault and Simonato (2017) and Zhang et al. (2019). For convergence analysis of this approach, one can refer to Clément et al. (2002), Stentoft (2004) and Tsitsiklis and Van Roy (2001).

Briefly speaking, the least squares Monte Carlo method is composed of two parts: (i) replace the conditional expectation by projection on a finite set of basis functions of state variables; (ii) use Monte Carlo simulation and least squares regression to compute the estimated values with the replacement in (i) recursively starting from the terminal stage. We now give more details on the implementation.

First, note that $\boldsymbol{v}_t$ is an expectation conditional on data $\mathcal{D}_t$. According to the update rules in (14)–(16), it is clear that $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ are the sufficient state variables to describe the probability distributions. Denote the set of elements in $\boldsymbol{\mu}_t$ and unique elements in $\boldsymbol{\Sigma}_t$ by $\boldsymbol{\theta}_t$. The evolution of $\boldsymbol{\theta}_t$ depends on the newly observed returns, their squares, and cross-products. For $t \in [0;T-2]$ and $j \in [1;p]$, we can write

$$v_t^j = \frac{-1}{1-\lambda}\mathbb{E}\left[\left(c_{t+1} + \lambda\gamma_{t+1}z_{t+1} - (1-\lambda)y_{t+1}\right)r_t^j \middle| \boldsymbol{\theta}_t\right] \tag{22}$$

and

$$c_t = \mathbb{E}\left[c_{t+1} + \lambda\gamma_{t+1}z_{t+1} - (1-\lambda)y_{t+1} \middle| \boldsymbol{\theta}_t\right], \tag{23}$$

which are the two types of conditional expectations we have to estimate in order to obtain $\boldsymbol{x}_t^*$. Essentially, the conditional expectations v_t^j and c_t can be viewed as some functions of $\boldsymbol{\theta}_t$. The theory on Hilbert spaces tells us that any function belonging to this space can be represented as a countable linear combination of basis vectors for the space (see Royden and Fitzpatrick 1988). Therefore, it is reasonable to approximate v_t^j and c_t by a set of basis functions as follows,

$$v_t^j = \sum_{m=1}^{M} a_{mt}^j \phi_m(\boldsymbol{\theta}_t) \quad \text{and} \quad c_t = \sum_{m=1}^{M} a_{mt}^c \phi_m(\boldsymbol{\theta}_t), \tag{24}$$

where $\{\phi_m(\boldsymbol{\theta}_t)\}_{m=1}^M$ are the M basis functions, and $\{a_{mt}^j\}_{m=1}^M$ and $\{a_{mt}^c\}_{m=1}^M$ are the coefficients for v_t^j and c_t, respectively. Specially, $v_{T-1}^j = \mu_{T-1}^j$, $j \in [1;p]$, and $c_{T-1} = -(1-\lambda)$ are obviously simple functions of $\boldsymbol{\theta}_{T-1}$. Note that we do not need to estimate $\boldsymbol{q}_t$ since given $\boldsymbol{\theta}_t$, one can directly compute it following (21).

We next employ Monte Carlo simulation and least squares regression to estimate $\{a_{mt}^j\}_{m=1}^M$ and $\{a_{mt}^c\}_{m=1}^M$. Consider a set of N simulated return paths, denoted as $\{\boldsymbol{r}_t^{(n)}\}_{n=1}^N$, $t \in [0;T-1]$, following the Bayesian update rules in Sect. 3.2 with both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown. Denote the realized values of state variables in path n as $\{\boldsymbol{\theta}_t^{(n)}\}_{t=1}^{T-1}$ and the MADs of returns are $\{\boldsymbol{q}_t^{(n)}\}_{t=1}^{T-1}$, $n \in [1;N]$. The algorithm works backwards from $t = T-1$ to the current decision point $t = 0$. At the beginning of period $T-1$, given $\boldsymbol{\theta}_{T-1}^{(n)}$, one can easily solve a single-period last stage problem following Theorem 1 and obtain the estimated values $\widehat{\boldsymbol{v}}_{T-1}^{(n)} = \sum_{s=-h}^{T-2} \boldsymbol{r}_s^{(n)}/(T-1+h)$, $\hat{c}_{T-1}^{(n)} = -(1-\lambda)$, $\hat{z}_{T-1}^{(n)}$ and $\hat{y}_{T-1}^{(n)}$ in each path. At the beginning of period $t < T-1$, we should already know the estimated values $\widehat{\boldsymbol{v}}_{t+1}^{(n)}$, $\hat{c}_{t+1}^{(n)}$, $\hat{z}_{t+1}^{(n)}$ and $\hat{y}_{t+1}^{(n)}$, $n \in [1;N]$. Then, the realized values of v_t^j and c_t in path n are

$$v_t^{j(n)} = \frac{-1}{1-\lambda}\left(\hat{c}_{t+1}^{(n)} + \lambda\gamma_{t+1}\hat{z}_{t+1}^{(n)} - (1-\lambda)\hat{y}_{t+1}^{(n)}\right) r_t^{j(n)} \tag{25}$$

and

$$c_t^{(n)} = \hat{c}_{t+1}^{(n)} + \lambda\gamma_{t+1}\hat{z}_{t+1}^{(n)} - (1-\lambda)\hat{y}_{t+1}^{(n)}.$$

On the other hand, we have the basis function values $\{\phi_m(\boldsymbol{\theta}_t^{(n)})\}_{m=1}^M$. Therefore, the estimated coefficients $\{\hat{a}_{mt}^j\}_{m=1}^M$, $j \in [1;p]$, and $\{\hat{a}_{mt}^c\}_{m=1}^M$ could be obtained by regressions, that is, they are the solutions of the following minimization problems:

$$\min_{\{a_{mt}^j\}_{m=1}^M} \sum_{n=1}^{N} \left[\sum_{m=1}^{M} a_{mt}^j \phi_m(\boldsymbol{\theta}_t^{(n)}) - v_t^{j(n)} \right]^2,$$

and

$$\min_{\{a_{mt}^c\}_{m=1}^M} \sum_{n=1}^{N} \left[\sum_{m=1}^{M} a_{mt}^c \phi_m(\boldsymbol{\theta}_t^{(n)}) - c_t^{(n)} \right]^2.$$

The fitted values of the regressions, denoted as $\{\hat{v}_t^{j(n)}\}_{j=1}^p$ and $\hat{c}_t^{(n)}$, $n \in [1;N]$, constitute the estimates of conditional expectations in (22) and (23). These estimates of the conditional expectations, in turn, yield estimates of $\hat{z}_t^{(n)}$ and $\hat{y}_t^{(n)}$ for each path n following the results in Theorem 1. At the decision point $t = 0$, since the state variable is fixed on $\boldsymbol{\theta}_0$ for all paths, the fitted value from the regression simply reduces to $\hat{v}_0^j = \sum_{n=1}^N v_0^{j(n)}/N$, $j \in [1;p]$ ($\hat{c}_0$ now is irrelevant to portfolio decision). Based on $\widehat{\boldsymbol{v}}_0$ and $\boldsymbol{q}_0$, the investor can optimally allocate his fund among the assets following Theorem 1.

There are many basis functions we can use for evaluating the conditional expectations, including Hermite, Legendre, Chebyshev, Laguerre polynomials among others. A number of numerical evidence in, e.g., Longstaff and Schwartz (2001), Brandt et al. (2005) and van Binsbergen and Brandt (2007) indicates that the order of the polynomial is not necessary to be very high for obtaining reliable estimates and even the first order (linear) polynomial of state variables is an effective choice in practice.

For path simulation, it should be noted that the sample paths of asset returns are simulated in a Bayesian context to perform learning. In each path n, once new data are revealed, we can compute the updated state variables in $\boldsymbol{\theta}_t^{(n)}$. Given $\boldsymbol{\theta}_t^{(n)}$, we can go on to simulate a new return data point along the path according to the multivariate t-distribution derived in (18). More simulation paths of course will improve the regression fitting, but at the expense of more computational time as the number of portfolio selection problems the algorithm needs to solve increases linearly with the number of simulated paths. The complete implementation process is presented in Algorithm 1.

Algorithm 1 The Optimal Investment Policy with Bayesian Learning

Initiation: Set $t=0$, λ and $\gamma_s \geqslant 0$, $s \in [t; T-1]$.
while *At the beginning of period* $t \in [0; T-2]$, **do**
  Simulate N return paths $\{\boldsymbol{r}_s^{(n)}\}_{n=1}^N$, $s \in [t; T-1]$ with Bayesian updating.
  Compute $\boldsymbol{q}_t$ following (21).
  Set $\widehat{\boldsymbol{v}}_{T-1}^{(n)} = \boldsymbol{\mu}_{T-1}^{(n)}$ and $\hat{c}_{T-1}^{(n)} = -(1-\lambda)$, $n \in [1; N]$.
  Compute $\hat{z}_{T-1}^{(n)}$ and $\hat{y}_{T-1}^{(n)}$ following Theorem 1, $n \in [1; N]$.
  for $s = T-2, T-3, ..., t+1$ **do**
    Compute $\widehat{\boldsymbol{v}}_s^{(n)}$ and $\hat{c}_s^{(n)}$, $n \in [1; N]$, using least squares Monte Carlo approach.
  end
  Compute $\widehat{\boldsymbol{v}}_t = \sum_{n=1}^N \boldsymbol{v}_t^{(n)}/N$ with $\boldsymbol{v}_t^{(n)}$ defined in (25).
  if $\gamma_t > 0$ *and* $\exists k \in [0; p-2]$ *such that* $G(\widehat{\boldsymbol{v}}_t, \gamma_t \boldsymbol{q}_t, k) < \lambda/(1-\lambda)$ *and* $G(\widehat{\boldsymbol{v}}_t, \gamma_t \boldsymbol{q}_t, k+1) \geqslant \lambda/(1-\lambda)$ **then**
    $\mathcal{A}_t^* = \{i_p(\widehat{\boldsymbol{v}}_t), i_{p-1}(\widehat{\boldsymbol{v}}_t), ..., i_{p-k}(\widehat{\boldsymbol{v}}_t)\}$
  else if $\gamma_t > 0$ *and* $G(\widehat{\boldsymbol{v}}_t, \gamma_t \boldsymbol{q}_t, p-1) < \lambda/(1-\lambda)$ **then**
    $\mathcal{A}_t^* = [1; p]$
  else
    $\mathcal{A}_t^* = \{i_p(\widehat{\boldsymbol{v}}_t)\}$
  end
  Compute the optimal portfolio allocation $\boldsymbol{x}_t^*$ following (7).
  Data $\boldsymbol{r}_t$ is revealed and it is included in $\mathcal{D}_{t+1}$. $t := t+1$.
end

4 Plug-in model

4.1 Model formulation

Unlike the Bayesian portfolio selection model, the plug-in model solves the multi-period investment problem believing that the unknown parameters could be estimated precisely by sample estimates given historical data. Let $\tilde{\boldsymbol{m}}_0 = (\tilde{m}_0^1, \ldots, \tilde{m}_0^p)^\top$ and $\tilde{\boldsymbol{q}}_0 = (\tilde{q}_0^1, \ldots, \tilde{q}_0^p)^\top$ denote the point estimates of unconditional expected returns and MADs, respectively, at the beginning of period 0. Given h historical data points in $\mathcal{D}_0$ and the assumption on the normal distribution of asset returns, we have

$$\tilde{\boldsymbol{m}}_0 = \frac{1}{h} \sum_{s=-h}^{-1} \boldsymbol{r}_s \quad \text{and} \quad \tilde{\boldsymbol{q}}_0 = \sqrt{\frac{2}{\pi(h-1)} \mathrm{diag}(\boldsymbol{\Sigma}_0)}, \tag{26}$$

where $\boldsymbol{\Sigma}_0$ has been defined in (14). Consistent with the Bayesian model, the cumulative risk during the investment horizon in this case is defined as

$$\tilde{L}_T = \sum_{t=0}^{T-1} \gamma_t \max_{1 \leqslant j \leqslant p} \tilde{q}_0^j x_t^j, \tag{27}$$

where $\tilde{\boldsymbol{q}}_0$ is used for all future periods $t \in [0;T-1]$, which implies the ignorance of parameter uncertainty and estimation risk in decision making under the plug-in model.

Accordingly, we set up the following multi-period optimization problem for the plug-in model,

$$\begin{aligned} \min_{\boldsymbol{x}_0 \in \mathcal{X}_0, \cdots, \boldsymbol{x}_{T-1} \in \mathcal{X}_{T-1}} \quad & \mathbb{E}\left[\lambda \tilde{L}_T - (1-\lambda) V_T\right] \\ \text{s.t.} \quad & V_{t+1} = V_t + \boldsymbol{r}_t^\top \boldsymbol{x}_t, \quad t \in [0;T-1]. \end{aligned} \tag{28}$$

The policy is self-financing, so we have $\mathbb{E}(V_T) = V_0 + \sum_{t=0}^{T-1} \mathbb{E}(\boldsymbol{r}_t)^\top \boldsymbol{x}_t$. Again, in plug-in model, we replace the uncondtional expected returns $\mathbb{E}(\boldsymbol{r}_t)$ with known sample estimate $\tilde{\boldsymbol{m}}_0$ for all $t \in [0;T-1]$ to emphasize that this model ignores parameter uncertainty.

4.2 The optimal policy

Problem (28) is a standard dynamic program and its solution could be obtained by backwards induction. We present the optimal policy in the following proposition.

Proposition 4 *Given non-negative* $\{\gamma_t\}_{t=0}^{T-1}$ *and* $\lambda \in (0,1)$, *the optimal policy of Problem* (28) *is such that for each stage* $t \in [0;T-1]$, *if* $j \notin \tilde{\mathcal{A}}_t^*$, *then* $\tilde{x}_t^{j*} = 0$; *if* $j \in \tilde{\mathcal{A}}_t^*$, *then* $\tilde{x}_t^{j*} = V_t / (\tilde{q}_0^j \sum_{j \in \tilde{\mathcal{A}}_t^*} 1/\tilde{q}_0^j)$, *where* $\tilde{\mathcal{A}}_t^*$ *can be determined by the rule: When* $\gamma_t > 0$, *if there exists an integer* $k \in [0;p-2]$ *such that* $G(\tilde{\boldsymbol{v}}_t, \gamma_t \tilde{\boldsymbol{q}}_0, k) < \lambda/(1-\lambda)$ *and*

$G(\tilde{\boldsymbol{v}}_t, \gamma_t \tilde{\boldsymbol{q}}_0, k+1) \geqslant \lambda/(1-\lambda)$, then $\tilde{\mathcal{A}}_t^* = \{i_p(\tilde{\boldsymbol{v}}_t), ..., i_{p-k}(\tilde{\boldsymbol{v}}_t)\}$; otherwise, $\tilde{\mathcal{A}}_t^* = [1;p]$. When $\gamma_t = 0$, $\tilde{\mathcal{A}}_t^* = \{i_p(\tilde{\boldsymbol{v}}_t)\}$. The vector $\tilde{\boldsymbol{v}}_t$ is recursively defined as: For $t = T-1$, $\tilde{\boldsymbol{v}}_t = \tilde{\boldsymbol{m}}_0$ and $\tilde{c}_t = -(1-\lambda)$. For $t \in [0;T-2]$, $\tilde{\boldsymbol{v}}_t = -\tilde{c}_t \tilde{\boldsymbol{m}}_0/(1-\lambda)$ and $\tilde{c}_t = \tilde{c}_{t+1} + \lambda\gamma_{t+1}\tilde{z}_{t+1} - (1-\lambda)\tilde{y}_{t+1}$ where for each $t \in [0;T-1]$, $\tilde{z}_t = 1/(\sum_{j\in\tilde{\mathcal{A}}_t^*} 1/\tilde{q}_0^j)$ and $\tilde{y}_t = \tilde{z}_t \sum_{j\in\tilde{\mathcal{A}}_t^*} \tilde{v}_t^j/\tilde{q}_0^j$.

Similar to the case with Bayesian learning, we can further rewrite the AER vector $\tilde{\boldsymbol{v}}_t$ in Proposition 4 as $\tilde{\boldsymbol{v}}_t = (1 + \tilde{\Delta}_{t+1})\tilde{\boldsymbol{m}}_0$ where

$$\tilde{\Delta}_{t+1} = \sum_{s=t+1}^{T-1} \frac{\tilde{\boldsymbol{v}}_s^\top \tilde{\boldsymbol{x}}_s^*}{V_s} - \frac{\lambda\gamma_s}{1-\lambda}\tilde{z}_s. \tag{29}$$

In addition, notice that given sample estimates $\tilde{\boldsymbol{m}}_0$ and $\tilde{\boldsymbol{q}}_0$, the optimal portfolio weights in percentage of wealth $\{\tilde{\boldsymbol{x}}_t^*/V_t\}_{t=0}^{T-1}$ can be exactly known at the beginning of period 0. Therefore, the plug-in investor can choose to follow the deterministic policy $\{\tilde{\boldsymbol{x}}_t^*/V_t\}_{t=0}^{T-1}$ to allocate his fund at each decision point, ignoring the new released data. A "wiser" alternative may be that the plug-in investor only uses $\tilde{\boldsymbol{x}}_0^*$ to decide the portfolio at time point 0 and updates the sample estimates with new observed returns for decisions in future periods. Specifically, according to (26), the update rule is simply that, for $t \in [0;T-1]$,

$$\tilde{\boldsymbol{m}}_t = \frac{1}{h+t}\sum_{s=-h}^{t-1} \boldsymbol{r}_s \quad \text{and} \quad \tilde{\boldsymbol{q}}_t = \sqrt{\frac{2}{\pi(h+t-1)}\mathrm{diag}(\boldsymbol{\Sigma}_t)},$$

where $\boldsymbol{\Sigma}_t$ is defined in (16). Then, at the beginning of period t, the plug-in investor solves Problem (28) with updated point estimates $\tilde{\boldsymbol{m}}_t$ and $\tilde{\boldsymbol{q}}_t$, and only $\tilde{\boldsymbol{x}}_t^*$ is used to construct the portfolio in stage t. We will present the performance of the above two decision approaches for the plug-in investor in our numerical study.

In financial economics literature, the difference in optimal portfolios between a long-term and a short-term investor is often identified as the hedging demand whose theoretical foundation can be dated back to Merton (1969). Here, we compare the multi-period (dynamic) and single-period (myopic) solutions under the plug-in model and the following four cases are considered.

(a): $\tilde{\Delta}_{t+1} > 0$. Because $G(\tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{q}}_0, k) \geqslant G(\tilde{\boldsymbol{m}}_0, \tilde{\boldsymbol{q}}_0, k)$ for all $k \in [0;p-1]$, the dynamic solution is more aggressive than the myopic solution under a positive future prospect. Precisely, the dynamic solution tends to select less assets and focuses on choosing assets with high expected returns to increase portfolio value.

(b): $\tilde{\Delta}_{t+1} = 0$. The dynamic solution degenerates to the myopic solution.

(c): $-1 \leqslant \tilde{\Delta}_{t+1} < 0$. Because $G(\tilde{\boldsymbol{v}}_t, \tilde{\boldsymbol{q}}_0, k) \leqslant G(\tilde{\boldsymbol{m}}_0, \tilde{\boldsymbol{q}}_0, k)$ for all $k \in [0;p-1]$, the dynamic solution performs more conservatively than the myopic solution by selecting more assets to diversify[4] the portfolio under a poor future prospect.

[4] text.

(d): $\tilde{\Delta}_{t+1} < -1$. The plug-investor has an extremely poor future prospect and the ordering of elements in $\tilde{\boldsymbol{v}}_t$ is reversed. Suppose that the MADs in $\tilde{\boldsymbol{q}}_0$ are ordered in the same way as their expected returns in $\tilde{\boldsymbol{m}}_0$. In this case, the dynamic solution prefers investing in assets with low expected returns (e.g., some bonds or treasury bills instead of stocks) to possibly control risk.

4.3 Comparison with Bayesian model

For now, we have shown the optimal policies under Bayesian and plug-in model in Theorem 1 and Proposition 4.1 respectively. Then, we discuss the impact of incorporating Bayesian learning on investor's decision making by comparing these two policies.

Given $\tilde{\boldsymbol{v}}_t$ and $\boldsymbol{v}_t$, we see that $\tilde{\boldsymbol{v}}_t$ is a deterministic function that relies on the sample estimates of historical returns, as shown in $\tilde{\boldsymbol{m}}_0$ and $\tilde{\boldsymbol{q}}_0$, while $\boldsymbol{v}_t$ is a conditional expectation adapted to the available information $\mathcal{F}_t$ in a Bayesian fashion. Particularly, for each $j \in [1;p]$, v_t^j can be understood as an expected value across different realizations of future returns and state variables based on the observed data at the beginning of period t. That is, given $\boldsymbol{\theta}_t$, the evaluation of $\boldsymbol{v}_t$ has an anticipation that the future prospect may deviate from what has been revealed by historical information in $\boldsymbol{\theta}_t$. In contrast, ignoring parameter uncertainty, the plug-in model fully "trusts" the historical information and believes that risky assets in future would perform the same as in history, which may lead to extreme AER values especially under some poor parameter estimates with large errors.

On the other hand, as a consequence of considering extra sources of uncertainty in unknown parameters, accounting for parameter uncertainty via Bayesian learning makes the risky assets more risky in the sense that the MADs estimated in the Bayesian model are larger than those of plug-in model. Specifically, we can rewrite $\boldsymbol{q}_0$ and $\tilde{\boldsymbol{q}}_0$ as $\boldsymbol{q}_0 = \beta_L(h)\sqrt{\text{diag}(\boldsymbol{\Sigma}_0)}$ and $\tilde{\boldsymbol{q}}_0 = \beta_P(h)\sqrt{\text{diag}(\boldsymbol{\Sigma}_0)}$ according to (21) and (26) where

$$\beta_L(h) = \frac{2\sqrt{h+1}}{\sqrt{h}(h-p-1)B\left(\frac{h-p}{2}, \frac{1}{2}\right)} \quad \text{and} \quad \beta_P(h) = \sqrt{\frac{2}{\pi(h-1)}}.$$

Both $\beta_L(h)$ and $\beta_P(h)$ are functions of the amount of historical data h. A more straightforward and quantitative description of these two functions is provided in Fig. 1 where it can be easily observed that $\beta_L(h) > \beta_P(h)$ especially when h is small.

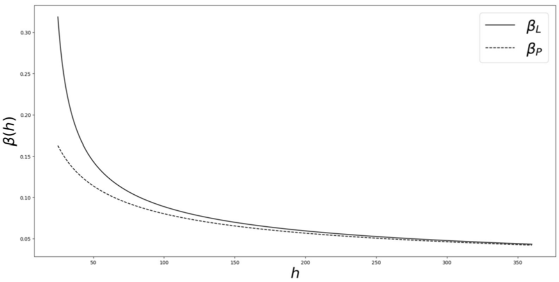

Fig. 1 Comparison of $\beta_L(h)$ and $\beta_P(h)$

Table 1 Data description ($\times 10^{-2}$)

Assests	Name	$\boldsymbol{\mu}$		$(\boldsymbol{\Sigma})_{jj}$	
		EX	SD	EX	SD
S_1	Food	0.718	0.210	0.158	0.012
S_2	Mines	0.666	0.426	0.653	0.050
S_3	Oil	0.616	0.295	0.313	0.024
S_4	Clths	0.877	0.323	0.375	0.029
S_5	Durbl	0.520	0.300	0.324	0.025
S_6	Chems	0.681	0.307	0.339	0.026
S_7	Cnsum	0.800	0.222	0.178	0.014
S_8	Cnstr	0.937	0.308	0.342	0.026
S_9	Steel	0.502	0.443	0.706	0.054
S_{10}	FabPr	0.833	0.290	0.303	0.023
S_{11}	Machn	0.992	0.373	0.501	0.039
S_{12}	Cars	0.598	0.350	0.440	0.034
S_{13}	Trans	0.870	0.268	0.259	0.020
S_{14}	Utils	0.468	0.209	0.157	0.012
S_{15}	Rtail	0.892	0.252	0.229	0.018
S_{16}	Finan	0.790	0.296	0.316	0.024
S_{17}	Other	0.729	0.261	0.245	0.019

This table presents the expectation (EX) and standard deviation (SD) of unknown parameters in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ conditional on $\mathcal{D}_0$

As time passes and more data are revealed, the issue of parameter uncertainty is relieved and thus $\beta_L(h)$ gradually approaches to $\beta_P(h)$ in Fig. 1. Moreover, with larger MADs $\boldsymbol{q}_0$ evaluated, more assets are likely to be selected for investment in the Bayesian model than the case of plug-in model because $G(\boldsymbol{v}_0, \boldsymbol{q}_0, k) \leqslant G(\boldsymbol{v}_0, \tilde{\boldsymbol{q}}_0, k)$ for all $k \in [0;p-1]$ and any given $\boldsymbol{v}_0$.

5 Numerical study

In this numerical study, we first investigate the role of Bayesian learning in the optimal portfolio decision. Then, an out-of-sample performance test is provided for models with and without Bayesian learning based on real market data.

5.1 Data

The market data used in this study consist of monthly return data of 17 industry portfolios from August 1989 to July 2019, that is, we have $p = 17$ and 30-year monthly return data in this experiment. These data are accessible on the website of

Ken French.[5] In Table 1, we list the industry portfolio names and report the expectations and standard deviations of posterior distributions of unknown parameters in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given full data sample. Particularly, in Table 1, we compute $\mathbb{E}_0(\mu^j)$, $\mathbb{E}_0((\boldsymbol{\Sigma})_{jj})$, $\sqrt{\mathrm{Var}(\mu^j|\mathcal{D}_0)}$ and $\sqrt{\mathrm{Var}((\boldsymbol{\Sigma})_{jj}|\mathcal{D}_0)}$, $j \in [1;17]$, following equations in (19) and (20) with $\mathcal{D}_0$ containing the full data sample from August 1989 to July 2019. The results in Table 1 show that the expectations of $\{\mu^j\}_{j=1}^{17}$ conditional on $\mathcal{D}_0$ range from 0.468 to 0.992% with standard deviations $\{\sqrt{\mathrm{Var}(\mu^j|\mathcal{D}_0)}\}_{j=1}^{17}$ varying from 0.209 to 0.443%. Compared to the standard deviations of return variances $\{\sqrt{\mathrm{Var}((\boldsymbol{\Sigma})_{jj}|\mathcal{D}_0)}\}_{j=1}^{17}$ which possess the range from 0.012% to 0.054%, it appears that the uncertainty in return means is a dominated source of parameter uncertainty for investors.

5.2 The role of Bayesian learning

As analyzed in Sect. 4.3, the Bayesian model is likely to produce a more diversified portfolio than the plug-in model. We show this phenomenon by designing an experiment with results presented in Table 2 where we report the conditional expectations $\mathbb{E}_0(\boldsymbol{\mu})$, their standard deviations $\sqrt{\mathrm{Var}(\boldsymbol{\mu}|\mathcal{D}_0)}$, the MADs from two investors and portfolio positions under two scenarios, i.e. "Normal" and "High", with monthly return data from February 2017 to July 2019 (i.e., $h = 30$)[6] contained in $\mathcal{D}_0$. We use the least squares Monte Carlo method to estimate AERs in $\boldsymbol{v}_0$ with $N = 20000$ and basis functions including the state variables and their quadratic values, that is, we have $M = 341$ and

$$(\phi_1(\boldsymbol{\theta}_0), \dots, \phi_{341}(\boldsymbol{\theta}_0)) = \Big(1, \mu_0^1, \dots, \mu_0^{17}, (\boldsymbol{\Sigma}_0)_{1,1}, \dots, \{(\boldsymbol{\Sigma}_0)_{i,j}\}_{i\le j}, \dots, (\boldsymbol{\Sigma}_0)_{17,17}, \\ (\mu_0^1)^2, \dots, (\mu_0^{17})^2, (\boldsymbol{\Sigma}_0)_{1,1}^2, \dots, \{(\boldsymbol{\Sigma}_0)_{i,j}^2\}_{i\le j}, \dots, (\boldsymbol{\Sigma}_0)_{17,17}^2\Big). \tag{30}$$

In Table 2, the "Normal" scenario means that the results are obtained under the real estimates $\boldsymbol{\mu}_0 = \tilde{\boldsymbol{m}}_0 = \sum_{s=-h}^{-1} \boldsymbol{r}_s/h$, while "High" corresponds to the outcomes with $\mu_0^j = \tilde{m}_0^j = \sum_{s=-h}^{-1} \boldsymbol{r}_s/h + \sqrt{\mathrm{Var}(\mu^j|\mathcal{D}_0)}$, $j \in [1;17]$. In other words, we add a perturbation (plus one standard deviation) on the sample average of historical returns in $\mathcal{D}_0$ and want to see the responses of $\boldsymbol{v}_0$, $\tilde{\boldsymbol{v}}_0$ and the optimal portfolio allocations given new mean estimates in the "High" scenario. The investment horizon is set to be $T = 6$ months with $\lambda = 0.4$ and $\gamma_t = 1$, $\forall t \in [0;5]$.

According to Table 2, the results in "High" scenario show that $\boldsymbol{v}_0^{\mathrm{H}}$ is uniformly lower than $\tilde{\boldsymbol{v}}_0^{\mathrm{H}}$, with a smaller range ($1.713 \times 10^{-2}$ in $\boldsymbol{v}_0^{\mathrm{H}}$ versus 1.938×10^{-2} in $\tilde{\boldsymbol{v}}_0^{\mathrm{H}}$) such that $G(\boldsymbol{v}_0^{\mathrm{H}}, \boldsymbol{q}_0, 16) < G(\boldsymbol{v}_0^{\mathrm{H}}, \tilde{\boldsymbol{q}}_0, 16) < G(\tilde{\boldsymbol{v}}_0^{\mathrm{H}}, \tilde{\boldsymbol{q}}_0, 16)$. It is thus reasonable to

[5] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html.

[6] With a relatively short length historical data, the issue of parameter uncertainty could be severe, which is beneficial for observing the differences between the cases with and without learning. More discussions on the length of historical data will be presented in the later subsection.

Table 2 Role of Bayesian learning ($\times 10^{-2}$)

Assets	$\boldsymbol{\mu}$		MAD		Normal				High			
	EX	SD	$\boldsymbol{q}_0$	$\tilde{\boldsymbol{q}}_0$	$\boldsymbol{v}_0$	$\tilde{\boldsymbol{v}}_0$	$\boldsymbol{x}_0^*$	$\tilde{\boldsymbol{x}}_0^*$	$\boldsymbol{v}_0^{\mathrm{H}}$	$\tilde{\boldsymbol{v}}_0^{\mathrm{H}}$	$\boldsymbol{x}_0^{*\mathrm{H}}$	$\tilde{\boldsymbol{x}}_0^{*\mathrm{H}}$
S_1	0.365	0.993	4.314	2.629	0.390	0.383	0	0	1.450	1.533	0	0
S_2	− 0.299	1.728	7.502	4.571	− 0.349	− 0.314	0	0	1.559	1.611	0	0
S_3	− 0.288	1.911	8.299	5.057	− 0.317	− 0.302	0	0	1.715	1.831	0	0
S_4	1.476	1.601	6.952	4.236	1.518	1.547	11.471	11.471	3.163	3.471	11.471	16.283
S_5	0.579	1.584	6.878	4.191	0.568	0.606	0	0	2.236	2.440	0	0
S_6	0.156	1.458	6.331	3.858	0.219	0.163	0	0	1.628	1.820	0	0
S_7	0.539	1.067	4.632	2.822	0.529	0.564	0	0	1.714	1.811	0	0
S_8	1.150	1.572	6.828	4.161	1.167	1.205	11.679	11.679	2.805	3.071	11.679	16.578
S_9	− 0.164	2.264	9.831	5.990	− 0.100	− 0.172	0	0	2.183	2.369	0	0
S_{10}	1.193	1.603	6.962	4.242	1.199	1.249	11.455	11.455	2.881	3.154	11.455	16.261
S_{11}	1.359	1.583	6.873	4.188	1.426	1.423	11.604	11.604	3.051	3.318	11.604	16.471
S_{12}	0.268	1.349	5.860	3.571	0.316	0.281	0	0	1.620	1.825	0	0
S_{13}	1.083	1.614	7.011	4.272	1.145	1.135	11.374	11.374	2.783	3.043	11.374	16.146
S_{14}	0.575	0.823	3.573	2.177	0.626	0.602	0	0	1.476	1.576	0	0
S_{15}	1.593	1.427	6.199	3.777	1.688	1.669	12.865	12.865	3.086	3.407	12.865	18.261
S_{16}	0.976	1.316	5.714	3.482	1.064	1.022	13.957	13.957	2.328	2.585	13.957	0
S_{17}	1.217	1.178	5.114	3.116	1.271	1.275	15.594	15.594	2.483	2.702	15.594	0

The results are obtained with the data sample from February 2017 to July 2019. We set $\lambda = 0.4$, $T = 6$, $V_0 = 1$ and $\gamma_t = 1$, $\forall t \in [0;5]$

expect a more diversified portfolio in Bayesian model than in plug-in model. Consistent with our analysis, we note that in “High” scenario, the plug-in investor selects 6 assets whereas the Bayesian investor selects 8 assets with their portfolio weights denoted by $\tilde{\boldsymbol{x}}_0^{*\mathrm{H}}$ and $\boldsymbol{x}_0^{*\mathrm{H}}$, respectively.

On the other hand, incorporating Bayesian learning to account for parameter uncertainty can also reduce the sensitivity of optimal portfolios to changes in model inputs. To show this, we follow the procedures similar to Best and Grauer (1991a) and focus on the results under the “Normal” scenario in Table 2 where $\boldsymbol{v}_0$ and $\tilde{\boldsymbol{v}}_0$ are close to each other and $\boldsymbol{x}_0^*$ is the same as $\tilde{\boldsymbol{x}}_0^*$. Among the selected 8 assets, for both Bayesian and plug-in investors, asset S_{15} and asset S_{16} feature the maximal and the minimal AER values, respectively. We present the model sensitivity by comparing the sizes of the shifts in the largest AER ($\tilde{v}_0^{15}$ and $\tilde{v}_0^{15}$) required to drive S_{16} from the original optimal portfolio ($\boldsymbol{x}_0^*$ and $\tilde{\boldsymbol{x}}_0^*$). It turns out that, when S_{16} is driven from $\boldsymbol{x}_0^*$, v_0^{15} should increase at least by 132.6% (from 1.688×10^{-2} to 3.926×10^{-2}). On the contrary, the required increase in $\tilde{v}_0^{15}$ is by 16.1% (from 1.669×10^{-2} to 1.938×10^{-2}). Similar results can also be obtained by decreasing q_0^{15} and $\tilde{q}_0^{15}$ to drive asset S_{16} from $\boldsymbol{x}_0^*$ and $\tilde{\boldsymbol{x}}_0^*$ (decrease by 78.2% in q_0^{15} versus 29.4% in $\tilde{q}_0^{15}$). The above results provide the evidence to support that the model with Bayesian learning is more robust to changes in model inputs than the plug-in model that ignores parameter uncertainty and estimation risk. Technically, the robustness gained by the Bayesian model comes from the higher estimates in MADs $\boldsymbol{q}_t$ which could attenuate the impact of changes in $\boldsymbol{v}_t$ and $\boldsymbol{q}_t$ to the variation of function G. Again, consistent with our analysis, we observe that $\boldsymbol{x}_0^*$ remains unchanged in “High” scenario, whereas, the plug-in investor selects less assets in “High” scenario with 8 assets in $\boldsymbol{x}_0^*$ versus 6 assets in $\tilde{\boldsymbol{x}}_0^{*\mathrm{H}}$.

5.3 Out-of-sample performance

In this section, we provide the out-of-sample performance of policies with and without Bayesian learning to further support our findings and analysis.

5.3.1 Models

Six models are considered in this out-of-sample test. We use BL to refer to the model with Bayesian learning with the policy derived by solving Problem (6). Two models for the plug-in investor based on Problem (28) are included, denoted as PI-1 and PI-2. PI-2 follows a deterministic policy solved at the beginning of period 0, ignoring the newly released data. PI-1 keeps updating the point estimates in $\tilde{\boldsymbol{m}}_t$ and $\tilde{\boldsymbol{q}}_t$, and only $\tilde{\boldsymbol{x}}_t^*$ is used for constructing the optimal portfolio in period t. Although both BL and PI-1 can utilize the new released data, the difference is that BL incorporates estimation risk and Bayesian learning while PI-1 does not. We also introduce two single-period models denoted as SP-1 and SP-2. SP-1 is a single-period model with l_∞ risk measure,

$$\min_{x \in \mathcal{X}} \quad \lambda \max_{1 \leqslant j \leqslant p} \tilde{q}_0^j x^j - (1-\lambda)\tilde{\boldsymbol{m}}_0^\top \boldsymbol{x}, \tag{31}$$

where $\mathcal{X} = \left\{ \boldsymbol{x} : \sum_{j=1}^p x^j = V_0, x^j \geqslant 0, j \in [1;p] \right\}$. SP-2 is a mean-variance type model. To make SP-1 and SP-2 comparable, we use portfolio standard deviation, instead of variance, as the risk measure in SP-2, that is, we solve the following problem in SP-2,

$$\min_{x \in \mathcal{X}} \quad \lambda \|\tilde{\boldsymbol{\Sigma}}_0^{1/2} \boldsymbol{x}\|_2 - (1-\lambda)\tilde{\boldsymbol{m}}_0^\top \boldsymbol{x}, \tag{32}$$

where $\tilde{\boldsymbol{\Sigma}}_0$ is the sample covariance matrix given $\mathcal{D}_0$; $\tilde{\boldsymbol{\Sigma}}_0^{1/2}$ is the matrix square root and $\|\cdot\|_2$ is the 2-norm. The naive equally weighted portfolio denoted as $1/p$ is included as well.

For BL, the AER vector $\boldsymbol{v}_t$ is estimated by the least squares Monte Carlo method with settings the same as those used in Table 2. The SP-2 model (32) is solved by CPLEX. Solutions of other models follow Lemma 1, Theorem 1 and Proposition 4.3.

5.3.2 Setup

The out-of-sample test follows a rolling-horizon procedure. We first choose an estimation window with h data points as training data. We set the investment horizon as T, so the following T data points are used as out-of-sample test data. Every month, BL and PI-1 are allowed to use all available data to update model inputs and rebalance their portfolio holdings. The parameter estimates of PI-2, SP-1 and SP-2 depend only on the h training data. During the investment horizon, PI-2 follows its deterministic policy solved at the beginning of period 0. SP-1 and SP-2 repeat using their myopic solutions solved at the beginning of period 0. At the end of the terminal stage, this investment process produces exactly T out-of-sample monthly portfolio return rates for the six models. Then, we repeat the test for the next investment horizon and move the data window by T months, again taking the first h data points as training data and the left T points as test data, until the end of the data set is reached. In this test, we set $T = 6$, $V_0 = 1$, $\lambda \in \{0.2, 0.8\}$, $\gamma_t = 1$ for all $t \in [0;T-1]$ and $h = 30$.

Based on the selected real data sequence from August 1989 to July 2019 (360 months), we further randomly generate nine time permuted return sequences. Note that we focus on the performance of the terminal portfolio return over the T-period investment. Since we set $T = 6$, for each return sequence, we will have 360/6 = 60 terminal portfolio returns, and, in total, we can collect $60 * 10 = 600$ terminal portfolio returns over the 10 monthly return sequences. The T-period investment test will be repeated for 600 times. For ease of reference, we call the kth T-period investment test the kth iteration where $k = 1, 2, ..., 600$.

Table 3 Out-of-sample performance ($p = 17$)

λ	Models	MEAN	STD	SR	VaR95%	VaR99%	PDM	SSQ	PSM	PTO	NSR
0.2	BL	0.046	0.118	0.395	– 0.132	– 0.260	5.036	2.838	2.536	2.155	0.390
	PI-1	0.045	0.123	0.367**	– 0.148	– 0.261	3.812	3.301	3.060	2.197	0.362**
	PI-2	0.045	0.122	0.369**	– 0.141	– 0.280	3.719	3.356	3.130	0.416	0.368*
	SP-1	0.045	0.122	0.370**	– 0.139	– 0.270	3.812	3.303	3.072	0.230	0.370*
	SP-2	0.041	0.121	0.341**	– 0.152	– 0.235	2.817	4.618	4.609	0.230	0.341**
	$1/p$	0.045	0.110	0.413	– 0.137	– 0.225	17.000	1.455	0.000	0.208	0.413
0.8	BL	0.045	0.104	0.433	– 0.126	– 0.213	16.710	1.508	0.228	0.281	0.432
	PI-1	0.045	0.105	0.431	– 0.130	– 0.221	15.639	1.561	0.434	0.450	0.430
	PI-2	0.045	0.105	0.430	– 0.132	– 0.220	15.488	1.570	0.464	0.223	0.429
	SP-1	0.045	0.105	0.431	– 0.132	– 0.220	15.493	1.570	0.463	0.201	0.430
	SP-2	0.038	0.089	0.421	– 0.114	– 0.181	5.348	3.581	2.493	0.173	0.420
	$1/p$	0.045	0.110	0.413***	– 0.137	– 0.225	17.000	1.455	0.000	0.208	0.413***

*p value<0.1, **p value<0.05, ***p value<0.01. The statistical results show the significance of the performance differences between BL and other models. For detailed p values, see Table 4

Table 4 p values of differences in SR and NSR ($p = 17$)

		BL	PI-1	PI-2	SP-1	SP-2	$1/p$
$\lambda = 0.2$	SR	–	0.015	0.045	0.040	0.012	0.312
	NSR	–	0.015	0.092	0.093	0.021	0.214
$\lambda = 0.8$	SR	–	0.517	0.392	0.604	0.620	0.000
	NSR	–	0.455	0.410	0.639	0.624	0.000

BL is the benchmark in the statistical test

5.3.3 Metrics

The performance metrics for the models include mean (MEAN), standard deviation (STD), Sharpe ratio (SR) and value-at-risk at 95% and 99% levels (VaR95%,VaR99%) of the 600 out-of-sample terminal portfolio returns. Besides these, we also report the portfolio turnover which is defined as

$$\mathrm{PTO} = \frac{1}{600}\sum_{k=1}^{600}\sum_{t=0}^{T-1} \|\boldsymbol{x}_{k,t} - \boldsymbol{x}_{k,t^-}\|_1,$$

where $\|\cdot\|_1$ denotes 1-norm, $\boldsymbol{x}_{k,t}$ is the desired portfolio weight vector in iteration k at the beginning of period t and $\boldsymbol{x}_{k,t^-}$ is the portfolio weight before rebalancing but after the realization of the actual asset returns based on $\boldsymbol{x}_{k,t-1}$. We set $\boldsymbol{x}_{k,0^-} = \boldsymbol{0}$. The metric PDM, short for portfolio diversification measure, documents the average number of assets with positive weights in one period, which can be mathematically defined as

$$\mathrm{PDM} = \frac{1}{600T}\sum_{k=1}^{600}\sum_{t=0}^{T-1}\sum_{j=1}^{p}\mathbb{1}\{\boldsymbol{x}_{k,t}^{j} > 0\}.$$

where $\mathbb{1}\{\cdot\}$ is an indicator function and it equals one if $\boldsymbol{x}_{k,t}^{j} > 0$ and zero otherwise. Moreover, we also compute the sum of squares of portfolio weights (SSQ) as another index for observing the fund distribution, with its definition follows that

$$\mathrm{SSQ} = \frac{1}{600}\sum_{k=1}^{600}\sum_{t=0}^{T-1}\left\|\frac{\boldsymbol{x}_{k,t}}{V_{k,t}}\right\|_2,$$

where $V_{k,t}$ is the initial wealth at the beginning of period t in iteration k. Finally, the portfolio sensitivity measure (PSM) (see Palczewski and Palczewski 2014) is given by

$$\mathrm{PSM} = \frac{1}{600}\sum_{k=1}^{600}\sum_{t=0}^{T-1}\left\|\frac{\boldsymbol{x}_{k,t}}{V_{k,t}} - \boldsymbol{w}_t^*\right\|_2,$$

where $\boldsymbol{w}_t^*$ contains the optimal percentage portfolio weights at the beginning of period t with full knowledge of the return distribution. For BL, PI-1 and PI-2, $\boldsymbol{w}_t^*$ is approximated by the optimal policy of a plug-in model with the full data samples. For SP-1 and SP-2, $\boldsymbol{w}_t^*$ is approximated by the solution from (31) and (32), respectively, given the full data samples.

Following DeMiguel et al. (2009), we also consider transaction costs in an *ex post* way. Specifically, the transaction cost arises from the portfolio turnover at the beginning of period t is quantified by $\kappa\|\boldsymbol{x}_{k,t} - \boldsymbol{x}_{k,t^-}\|_1$ where we set $\kappa = 0.002$. After deducting the total transaction costs in T periods, we can obtain a net terminal portfolio return for each iteration and then we can compute the Sharpe ratio net of the transaction costs (NSR) based on the 600 net terminal portfolio returns. To test whether the Sharpe ratios (SR/NSR) of two models are statistically distinguishable, we also compute the p value of the difference following Jobson and Korkie (1981) and Memmel (2003).[7]

5.3.4 Results

Table 3 contains the out-of-sample performance of the six models in terms of ten metrics and Table 4 reports the p values of differences in SR and NSR with benchmark being BL. In Table 3, we see that under $\lambda = 0.2$, BL outperforms PI-1 and PI-2 in SR with the performance gaps both significant at 5% level, and

[7] Specifically, given two portfolios i and n, with $\hat{\mu}_i$, $\hat{\mu}_n$, $\hat{\sigma}_i^2$, $\hat{\sigma}_i^2$, $\hat{\sigma}_{i,n}$ as their estimated means, variances, and covariances over a sample of size S, the test of the hypothesis $H_0 : \hat{\mu}_i/\hat{\sigma}_i - \hat{\mu}_n/\hat{\sigma}_n = 0$ is obtained via the test statistic $\hat{z}_{JK}$, which is asymptotically distributed as a standard normal:

$$\hat{z}_{JK} = \frac{\hat{\sigma}_n\hat{\mu}_i - \hat{\sigma}_i\hat{\mu}_n}{\sqrt{\hat{\vartheta}}},\ \text{with}\ \hat{\vartheta} = \frac{1}{S}\left(2\hat{\sigma}_i^2\hat{\sigma}_n^2 - 2\hat{\sigma}_i\hat{\sigma}_n\hat{\sigma}_{i,n} + \frac{1}{2}\hat{\mu}_i^2\hat{\sigma}_n^2 + \frac{1}{2}\hat{\mu}_n^2\hat{\sigma}_i^2 - \frac{\hat{\mu}_i\hat{\mu}_n}{\hat{\sigma}_i\hat{\sigma}_n}\hat{\sigma}_{i,n}^2\right).$$

the better performance of BL persists even after the consideration of transaction costs (see NSR and p values). Meanwhile, BL has larger MEAN, smaller STD, larger VaR95% and VaR99% than do PI-1 and PI-2, which apparently confirms the superiority of our proposed BL model. BL also has the largest PDM and smallest PSM compared to other models except $1/p$ under both $\lambda = 0.2$ and $\lambda = 0.8$, and these observations are consistent with our analysis in the previous section where we show that Bayesian learning can promote diversification and reduce sensitivity to data changes. Note that, under $\lambda = 0.2$, the metric results of PI-1, PI-2 and SP-1 are close to each other. The similar performance between PI-1 and PI-2 suggests that simply updating point estimates with new data may not improve the quality of the resulted portfolio. While, the similar performance between PI-2 and SP-1 implies that the advantage of dynamic models can be diminished with the existence of parameter uncertainty. The above observations together demonstrate the positive effect and the necessity of incorporating Bayesian learning procedure in dynamic portfolio optimization problems. On the other hand, the positive effect of using l_∞ risk function instead of portfolio variance can be observed from the comparison of models SP-1 and SP-2. Clearly, for both $\lambda = 0.2$ and $\lambda = 0.8$, SP-1 leads to better Sharpe ratios, SR and NSR, and notice that SP-2 has the smallest PDM and the largest SSQ and PSM in Table 3, which is aligned with the criticism on the classic mean-variance model that its solution is sensitive to model parameters and usually concentrates on a few assets (Litterman et al. 2004).

We have seen that BL can outperform PI-1 significantly under $\lambda = 0.2$. When $\lambda = 0.8$, their difference in SR narrows, e.g., 0.433 vs. 0.431 with p value 0.517 in Table 3. We know that the portfolio selection models with l_∞ risk function first select assets to invest according to (adjusted) expected returns and then determine the weight of each selected asset according to risks. So, when the investor is risk-seeking with a small λ, the resulting policy only focuses on several assets with large historical returns and consideration of extra parameter uncertainty can have a noticeable impact on portfolio choice by making the model include more assets than the case without considering parameter uncertainty and reducing sensitivity to input changes as shown in Table 2. On the contrary, when λ is close to 1, the resulting policy will almost invest on all the available assets and in such a case, considering parameter uncertainty cannot affect much the decision on the wealth weights of the selected assets, which causes the similar of BL and PI-1 under $\lambda = 0.8$.

Another interesting observation is that in contrast to the comparison of BL and PI-1, our proposed BL model significantly outperforms $1/p$ under $\lambda = 0.8$ (0.433 vs. 0.413 in SR with p value 0.000), even after the consideration of transaction costs (0.432 vs. 0.413 in NSR with p value 0.000), but their differences in SR and NSR are not significant under $\lambda = 0.2$. The reason is that when λ is large, BL will invest on most of the assets and actively distribute the fund over the selected assets, making use of the risk information, and thus it can have a significant advantage over $1/p$, while, when λ is small, BL that considers only several assets with high historical returns pursues high expected return in the cost of high volatility, which results in a higher out-of-sample return (0.046 vs.

Table 5 Additional out-of-sample test on 12-asset data set (p = 12)

λ	Models	MEAN	STD	SR	VaR95%	VaR99%	PDM	SSQ	PSM	PTO	NSR
0.2	BL	0.045	0.108	0.415	− 0.128	− 0.225	4.153	3.113	2.723	1.968	0.411
	PI-1	0.044	0.112	0.396**	− 0.126	− 0.237	3.465	3.448	3.098	1.897	0.391**
	PI-2	0.044	0.112	0.393**	− 0.133	− 0.234	3.356	3.517	3.196	0.355	0.392*
	SP-1	0.044	0.111	0.390**	− 0.133	− 0.234	3.415	3.480	3.156	0.214	0.390**
	SP-2	0.044	0.116	0.361***	− 0.135	− 0.238	2.702	4.761	4.630	0.218	0.361***
	1/p	0.044	0.100	0.441	− 0.130	− 0.202	12.000	1.732	0.000	0.194	0.440*
0.8	BL	0.044	0.097	0.450	− 0.125	− 0.192	11.960	1.772	0.213	0.235	0.450
	PI-1	0.044	0.098	0.449	− 0.123	− 0.192	11.784	1.785	0.273	0.278	0.448
	PI-2	0.044	0.097	0.450	− 0.118	− 0.191	11.739	1.789	0.295	0.200	0.449
	SP-1	0.044	0.097	0.450	− 0.118	− 0.191	11.728	1.790	0.297	0.188	0.450
	SP-2	0.037	0.087	0.422	− 0.109	− 0.184	5.150	3.617	2.340	0.170	0.421
	1/p	0.044	0.100	0.441***	− 0.130	− 0.202	12.000	1.732	0.000	0.194	0.440***

*p value<0.1, **p value<0.05, ***p value<0.01. The statistical results show the significance of the performance differences between BL and other models. For detailed p values, see Table 6

0.045 in MEAN) but possibly with a statistically insignificant lower Sharpe ratio than 1/p.

5.3.5 Additional tests

To check the robustness of our findings, we reimplement the experiments on another real market data set with monthly returns of 12 risky assets,[8] i.e., $p = 12$. The settings are the same as those in Sects. 5.3.1, 5.3.2 and 5.3.3 and the out-of-sample performance results are presented in Table 5 (ten metrics) and Table 6 (p values). The numerical results in Tables 5 and 6 still support our findings stated in Sect. 5.3.4, that is, (i) our proposed dynamic portfolio selection model with Bayesian learning BL is able to significantly outperform the plug-in models, PI-1 and PI-2, and the equally weighted portfolio 1/p. (ii) The significance of the performance gap is affected by the risk preference level λ. (iii) Compared to the mean-variance model, the use of l_∞ risk function can lead to better out-of-sample performance.

In addition, we also test the effect of Bayesian learning when more data are available with out-of-sample results presented in Table 7 where we set the historical data length $h = 120$ instead of $h = 30$ used in Tables 3 and 5. In Table 7, for both data sets $p = 17$ and $p = 12$, BL, PI-1 and PI-2 have a similar performance under $\lambda = 0.2$, which is in contrast to the observations in Tables 3 and 5. As predicted by

[8] This new data set is also accessible on the website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html. To be consistent with the case of the 17-asset data set, for the new 12-asset data set, we also choose the monthly return sequence from August 1989 to July 2019 (360 months) in the out-of-sample test.

Table 6 p values of differences in SR and NSR ($p = 12$)

		BL	PI-1	PI-2	SP-1	SP-2	$1/p$
$\lambda = 0.2$	SR	–	0.015	0.041	0.018	0.005	0.133
	NSR	–	0.017	0.090	0.047	0.009	0.083
$\lambda = 0.8$	SR	–	0.511	0.879	0.970	0.191	0.002
	NSR	–	0.481	0.912	0.984	0.192	0.002

BL is the benchmark in the statistical test

Table 7 More historical data ($h = 120$)

p	Models	MEAN	STD	SR	VaR95%	VaR99%	PDM	SSQ	PSM	PTO	NSR
17	BL	0.047	0.115	0.405	− 0.141	− 0.225	6.220	2.508	1.922	1.586	0.401
	PI-1	0.047	0.116	0.403	− 0.149	− 0.241	5.921	2.573	2.000	1.032	0.401
	PI-2	0.046	0.115	0.403	− 0.145	− 0.236	5.870	2.584	2.016	0.325	0.402
12	BL	0.046	0.105	0.433	− 0.133	− 0.210	5.457	2.654	1.879	1.354	0.430
	PI-1	0.045	0.105	0.430	− 0.129	− 0.199	5.278	2.701	1.928	0.873	0.428
	PI-2	0.046	0.105	0.437	− 0.129	− 0.198	5.254	2.706	1.948	0.307	0.436

The results are obtained under $h = 120$, $\lambda = 0.2$ for both $p = 17$ and $p = 12$ data sets

Table 8 Out-of-sample results of BL under linear basis

p	λ	MEAN	STD	SR	VaR95%	VaR99%	PDM	SSQ	PSM	PTO	NSR
17	0.2	0.046	0.118	0.395	− 0.133	− 0.260	5.054	2.833	2.529	2.145	0.390
	0.8	0.045	0.104	0.433	− 0.126	− 0.213	16.712	1.508	0.228	0.281	0.432
12	0.2	0.045	0.108	0.415	− 0.128	− 0.225	4.159	3.110	2.720	1.958	0.411
	0.8	0.044	0.097	0.450	− 0.125	− 0.192	11.960	1.772	0.213	0.235	0.450

the trend in Fig. 1, Table 7 numerically shows that the importance of incorporating Bayesian learning decreases as more data are available and the severity of parameter uncertainty reduces.

In Table 8, we report the new out-of-sample results of BL with AERs estimated by a linear basis, that is, we remove the quadratic terms in (30) and again run the regression following Algorithm 1. Compared to the results in Tables 3 and 5 which use the basis vector (30), the performance results in Table 8 with less regressors are almost the same as those listed in Table 3 and Table 5, suggesting that our numerical findings are robust to the setting in least squares Monte Carlo method.

6 Concluding remarks

The issue of parameter uncertainty and estimation risk has been long recognized as a crucial problem in portfolio management. In this paper, we incorporate Bayesian learning to deal with this issue in the framework of a proposed dynamic portfolio selection model where an l_∞ risk function is used as risk measure. The investor in our model is assumed to make decisions by maximizing expected terminal wealth at a minimal level of total risk, quantified by a weighted sum of the risks during the investment horizon. We show that the proposed stochastic dynamic program has a closed-form optimal policy that can be constructed intuitively. For implementation, we introduce a least squares Monte Carlo method to approximate the complex conditional expectations in AERs. We discuss the impact of Bayesian learning on investor's decision making and show how it promotes diversification and reduces sensitivity of optimal portfolios to changes in model inputs under an i.i.d. normal return-generating process with unknown means and covariance matrix. The numerical results based on real market data show that our proposed dynamic portfolio selection model with Bayesian learning can significantly outperform the plug-in models and the equally weighted portfolio with the performance gaps affected by the risk preference level and the amount of data available.

For future research, one can discard the assumption that the return-generating process is known and study the effect of model ambiguity based on the dynamic portfolio selection model developed in this paper. Another direction is to consider transaction costs in an *ex ante* way. As an important element for practical portfolio selection models, it is interesting to investigate how the investment strategy of our dynamic model would change if the transaction costs are included in model formulation. Finally, introducing other advanced Bayesian learning techniques with side information could also be a meaningful way to extend our work.

Appendix A: proofs

A.1. Proof of Theorem 1

Given $\boldsymbol{x}_t$, $\max_{1\leqslant j\leqslant p} q_t^j x_t^j$ is $\mathcal{F}_t$-measurable. Hence, we can split the total risk L_T and reformulate Problem (6) as follows:

$$\min_{\boldsymbol{x}_0\in\mathcal{X}_0} \lambda\gamma_0\left(\max_{1\leqslant j\leqslant p} q_0^j x_0^j\right)+\mathbb{E}_0\left[\min_{\boldsymbol{x}_1\in\mathcal{X}_1} \lambda\gamma_1\left(\max_{1\leqslant j\leqslant p} q_1^j x_1^j\right)+\mathbb{E}_1\left[\cdots \min_{\boldsymbol{x}_{T-1}\in\mathcal{X}_{T-1}} \lambda\gamma_{T-1}\left(\max_{1\leqslant j\leqslant p} q_{T-1}^j x_{T-1}^j\right)-(1-\lambda)\mathbb{E}_{T-1}(V_T)\right]\cdots\right], \tag{33}$$

with budget conditions $V_{t+1} = V_t + \boldsymbol{r}_t^\top \boldsymbol{x}_t$, $t \in [0;T-1]$. At the beginning of period $T-1$, consider the last stage problem:

$$\min_{\boldsymbol{x}_{T-1}\in\mathcal{X}_{T-1}} \lambda\gamma_{T-1}\left(\max_{1\leqslant j\leqslant p} q_{T-1}^j x_{T-1}^j\right)-(1-\lambda)\mathbb{E}_{T-1}(V_T).$$

Since $V_T = V_{T-1} + \boldsymbol{r}_{T-1}^\top \boldsymbol{x}_{T-1}$, we have

$$\min_{\boldsymbol{x}_{T-1}\in\mathcal{X}_{T-1}} \lambda\left(\max_{1\leqslant j\leqslant p} \gamma_{T-1} q_{T-1}^j x_{T-1}^j\right) - (1-\lambda)\boldsymbol{v}_{T-1}^\top \boldsymbol{x}_{T-1} + c_{T-1} V_{T-1}, \tag{34}$$

where $\boldsymbol{v}_{T-1} = \mathbb{E}_{T-1}(\boldsymbol{r}_{T-1})$ and $c_{T-1} = -(1-\lambda)$. Two cases need to be discussed: $\gamma_{T-1} > 0$ and $\gamma_{T-1} = 0$. Assume that $\gamma_{T-1} > 0$. Applying Lemma 1, the optimal allocation $\boldsymbol{x}_{T-1}^*$ is that for $j \in \mathcal{A}_{T-1}^*$, $x_{T-1}^{j*} = \left(\sum_{j\in\mathcal{A}_{T-1}^*} 1/q_{T-1}^j\right)^{-1} V_{T-1}/q_{T-1}^j$ and for $j \notin \mathcal{A}_{T-1}^*$, $x_{T-1}^{j*} = 0$, where the set of assets to be invested, $\mathcal{A}_{T-1}^*$, can be defined by the rule: If there exists an integer $k \in [0;p-2]$ such that $G(\boldsymbol{v}_{T-1}, \lambda_{T-1}\boldsymbol{q}_{T-1}, k) < \lambda(1-\lambda)$ and $G(\boldsymbol{v}_{T-1}, \lambda_{T-1}\boldsymbol{q}_{T-1}, k+1) \geq \lambda(1-\lambda)$, then $\mathcal{A}_{T-1}^* = \{i_p(\boldsymbol{v}_{T-1}), ..., i_{p-k}(\boldsymbol{v}_{T-1})\}$. Otherwise, $\mathcal{A}_{T-1}^* = [1;p]$. When $\gamma_{T-1} = 0$, the optimal solution of (34) is obviously that $x_{T-1}^{i_p(\boldsymbol{v}_{T-1})} = V_{T-1}$ and $x_{T-1}^j = 0$ for all $j \neq i_p(\boldsymbol{v}_{T-1})$. That is, in this case, the optimal strategy is to invest all the fund V_{T-1} on the asset with the largest value in $\boldsymbol{v}_{T-1}$, which is equivalent to choosing $\mathcal{A}_{T-1}^* = \{i_p(\boldsymbol{v}_{T-1})\}$. Therefore, we can combine the two cases and find the value function of period $T-1$ as follows:

$$Q_{T-1}(V_{T-1}) = V_{T-1}[c_{T-1} + \lambda\gamma_{T-1} z_{T-1} - (1-\lambda) y_{T-1}],$$

where $y_{T-1} = z_{T-1}\sum_{j\in\mathcal{A}_{T-1}^*} v_{T-1}^j / q_{T-1}^j$ and $z_{T-1} = 1/(\sum_{j\in\mathcal{A}_{T-1}^*} 1/q_{T-1}^j)$. Here Q_{T-1} is $\mathcal{F}_{T-1}$-measurable and it could be written more explicitly as $Q_{T-1}(V_{T-1}(\omega), \omega)$. The inclusion of V_{T-1} shows that the value function also depends on the portfolio decisions in previous periods.

Next, we use mathematical induction to prove that for all $t \in [0;T-1]$, the value function has the form $Q_t(V_t) = V_t[c_t + \lambda\gamma_t z_t - (1-\lambda)y_t]$. Obviously, $Q_{T-1}(V_{T-1})$ satisfies the condition. Suppose for $t+1 \leq T-1$, $Q_{t+1}(V_{t+1}) = V_{t+1}[c_{t+1} + \lambda\gamma_{t+1} z_{t+1} - (1-\lambda) y_{t+1}]$. At the beginning of period t, we need to solve the optimization problem:

$$\min_{\boldsymbol{x}_t\in\mathcal{X}_t} \lambda\left(\max_{1\leqslant j\leqslant p} \gamma_t q_t^j x_t^j\right) + \mathbb{E}_t\left[Q_{t+1}(V_{t+1})\right].$$

By taking in $V_{t+1} = V_t + \boldsymbol{r}_t^\top \boldsymbol{x}_t$, we have the following equivalent problem

$$\min_{\boldsymbol{x}_t\in\mathcal{X}_t} \lambda\left(\max_{1\leqslant j\leqslant p} \gamma_t q_t^j x_t^j\right) - (1-\lambda)\boldsymbol{v}_t^\top \boldsymbol{x}_t + c_t V_t, \tag{35}$$

where $c_t = \mathbb{E}_t[c_{t+1} + \lambda\gamma_{t+1} z_{t+1} - (1-\lambda) y_{t+1}]$ and $\boldsymbol{v}_t = -\mathbb{E}_t[(c_{t+1} + \lambda\gamma_{t+1} z_{t+1} - (1-\lambda) y_{t+1})\boldsymbol{r}_t]/(1-\lambda)$. The above Problem (35) is essentially the same as the last stage problem, only with parameters indexed by t. Again, one can follow Lemma 1 and obtain that for $j \in \mathcal{A}_t^*$, $x_t^{j*} = \left(\sum_{j\in\mathcal{A}_t^*} 1/q_t^j\right)^{-1} V_t/q_t^j$ and for $j \notin \mathcal{A}_t^*$, $x_t^{j*} = 0$, where the set of assets to be invested $\mathcal{A}_t^*$ is determined by the rule: When $\gamma_t > 0$, if there exists an integer $k \in [0;p-2]$ such that $G(\boldsymbol{v}_t, \gamma_t\boldsymbol{q}_t, k) < \lambda(1-\lambda)$ and $G(\boldsymbol{v}_t, \gamma_t\boldsymbol{q}_t, k+1) \geq \lambda(1-\lambda)$, then $\mathcal{A}_t^* = \{i_p(\boldsymbol{v}_t), ..., i_{p-k}(\boldsymbol{v}_t)\}$; otherwise, $\mathcal{A}_t^* = [1;p]$.

When $\gamma_t = 0$, let $\mathcal{A}_t^* = \{i_p(\boldsymbol{v}_t)\}$. Taking the optimal solution $\boldsymbol{x}_t^*$ into Problem (35), we have the value function such that $Q_t(V_t) = V_t[c_t + \lambda\gamma_t z_t - (1-\lambda)y_t]$ with

$$y_t = z_t \sum_{j\in\mathcal{A}_t^*} \frac{v_t^j}{q_t^j} \quad \text{and} \quad z_t = \left(\sum_{j\in\mathcal{A}_t^*} \frac{1}{q_t^j}\right)^{-1},$$

which completes the proof. Thus, the optimal policy of Problem (6) is that for each $t \in [0;T-1]$, one should choose $\boldsymbol{x}_t^*$, the optimal solution of Problem (35). □

A.2. Proof of Proposition 1

Note that c_{t+1} is $\mathcal{F}_{t+1}$-measurable. According to the law of total expectation, we have

$$\boldsymbol{v}_t = \mathbb{E}_t\Big[\Big(-\frac{c_{t+1}}{1-\lambda} - \frac{\lambda\gamma_{t+1}}{1-\lambda}z_{t+1} + y_{t+1}\Big)\boldsymbol{r}_t\Big] = \mathbb{E}_t\Big(-\frac{c_{t+1}}{1-\lambda}\boldsymbol{r}_t\Big) + \mathbb{E}_t\Big[\Big(y_{t+1} - \frac{\lambda\gamma_{t+1}}{1-\lambda}z_{t+1}\Big)\boldsymbol{r}_t\Big].$$

Replacing c_{t+1} with $\mathbb{E}_{t+1}[c_{t+2} + \lambda\gamma_{t+2}z_{t+2} - (1-\lambda)y_{t+2}]$, we have

$$\begin{aligned}\boldsymbol{v}_t &= \mathbb{E}_t\left[\mathbb{E}_{t+1}\left(-\frac{c_{t+2}}{1-\lambda} - \frac{\lambda\gamma_{t+2}}{1-\lambda}z_{t+2} + y_{t+2}\right)\boldsymbol{r}_t\right] + \mathbb{E}_t\left[\left(y_{t+1} - \frac{\lambda\gamma_{t+1}}{1-\lambda}z_{t+1}\right)\boldsymbol{r}_t\right]\\ &= \mathbb{E}_t\left(-\frac{c_{t+2}}{1-\lambda}\boldsymbol{r}_t\right) + \mathbb{E}_t\left[\left(\sum_{s=t+1}^{t+2} y_s - \frac{\lambda\gamma_s}{1-\lambda}z_s\right)\boldsymbol{r}_t\right]\end{aligned}$$

Given that $c_s = \mathbb{E}_s[c_{s+1} + \lambda\gamma_{s+1}z_{s+1} - (1-\lambda)y_{s+1}]$ for all $s \in [t;T-2]$ and $c_{T-1} = -(1-\lambda)$, we continue to replace $\{c_s\}_{s=t+1}^{T-1}$ under the law of total expectation and finally obtain that:

$$\boldsymbol{v}_t = \mathbb{E}_t(\boldsymbol{r}_t) + \mathbb{E}_t\left[\left(\sum_{s=t+1}^{T-1} y_s - \frac{\lambda\gamma_s}{1-\lambda}z_s\right)\boldsymbol{r}_t\right].$$

y_s can be further represented as

$$y_s = \left(\sum_{j\in\mathcal{A}_s^*} \frac{1}{q_s^j}\right)^{-1} \cdot \sum_{j\in\mathcal{A}_s^*} \frac{v_s^j}{q_s^j} = \frac{\boldsymbol{v}_s^\top \boldsymbol{x}_s^*}{V_s},$$

where $\boldsymbol{x}_s^*$ denotes the optimal allocation in period s. Therefore, we have

$$\boldsymbol{v}_t = \mathbb{E}_t(\boldsymbol{r}_t) + \mathbb{E}_t\Big[\Big(\sum_{s=t+1}^{T-1} \frac{\boldsymbol{v}_s^\top \boldsymbol{x}_s^*}{V_s} - \frac{\lambda\gamma_s}{1-\lambda}z_s\Big)\boldsymbol{r}_t\Big],$$

which completes the proof. □

A.3. Proof of Proposition 2

According to (18), we know that $r_t^j|\mathcal{D}_t$, $j \in [1;p]$, follows a t-distribution. The probability density function is given by

$$f(r_t^j|\mu_t^j,\sigma) = \frac{1}{\sqrt{\nu}B(\nu/2,1/2)}\frac{1}{\sigma}\left[1+\frac{1}{\nu}\left(\frac{r_t^j-\mu_t^j}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}},$$

where $\nu = h+t-p$ $(\nu > 1)$ and $\sigma^2 = \frac{h+t+1}{(h+t)(h+t-p)}(\boldsymbol{\Sigma}_t)_{jj}$. The expectation is simply $\mathbb{E}(r_t^j|\mu_t^j,\sigma) = \mu_t^j$. Suppose X has a standard t-distribution with ν being the degrees of freedom. It is clear that $r_t^j = \mu_t^j + \sigma X$. Thus, the conditional MAD of r_t^j is

$$\begin{aligned}\mathbb{E}(|r_t^j-\mu_t^j||\mu_t^j,\sigma) = \sigma\mathbb{E}(|X||\mu_t^j,\sigma) &= \frac{2\sqrt{\nu}\sigma}{(\nu-1)B(\nu/2,1/2)}\\ &= \frac{2\sqrt{(h+t+1)(\boldsymbol{\Sigma}_t)_{jj}/(h+t)}}{(h+t-p-1)B((h+t-p)/2,1/2)},\end{aligned}$$

which completes the proof. □

A.4. Proof of Proposition 3

let us consider the following probability on the deviation of portfolio return with respect to its expected value: $\mathbb{P}\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right| \geqslant \xi\middle|\mathcal{D}_t\right)$. According to the Markov's inequality, it holds that for arbitrary $\xi > 0$,

$$\mathbb{P}\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right| \geqslant \xi\middle|\mathcal{D}_t\right) \leqslant \frac{1}{\xi}\mathbb{E}_t\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right|\right).$$

Then, according to the triangle inequality, we have

$$\mathbb{E}_t\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right|\right) \leqslant \sum_{j=1}^p \mathbb{E}_t\left(\left|r_t^j x_t^j - m_t^j x_t^j\right|\right).$$

By the definition of $l_t(\boldsymbol{x}_t)$, one can get that $\mathbb{E}_t\left(\left|r_t^j x_t^j - m_t^j x_t^j\right|\right) \leqslant l_t(\boldsymbol{x}_t)$. Therefore, we have proved that

$$\mathbb{P}\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right| \geqslant \xi\middle|\mathcal{D}_t\right) \leq \frac{p}{\xi}l_t(\boldsymbol{x}_t).$$

Under the normal distribution $\boldsymbol{r}_t|\mathcal{D}_t \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, we have

$$\mathbb{P}\left(\left|\boldsymbol{r}_t^\top\boldsymbol{x}_t - \boldsymbol{m}_t^\top\boldsymbol{x}_t\right| \geqslant \xi\middle|\mathcal{D}_t\right) = 2\left(1-\Phi\left(\frac{\xi}{\sqrt{\boldsymbol{x}_t^\top\boldsymbol{\Sigma}\boldsymbol{x}_t}}\right)\right)$$

Table 9 Additional Out-of-Sample Tests under $\lambda = 0.5$

p	Models	MEAN	STD	SR	VaR95%	VaR99%	PDM	SSQ	PSM	PTO	NSR
	BL	0.045	0.108	0.419	– 0.130	– 0.232	11.482	1.830	1.091	1.102	0.416
	PI-1	0.045	0.110	0.409	– 0.126	– 0.239	8.787	2.114	1.517	1.455	0.406
17	PI-2	0.045	0.109	0.416	– 0.123	– 0.236	8.462	2.158	1.578	0.304	0.416
	SP-1	0.045	0.109	0.416	– 0.123	– 0.234	8.558	2.144	1.560	0.209	0.416
	SP-2	0.039	0.098	0.402	– 0.123	– 0.190	4.760	3.718	3.125	0.186	0.401
	1/p	0.045	0.110	0.413	– 0.137	– 0.225	17.000	1.455	2.648	0.208	0.413
	BL	0.045	0.100	0.445	– 0.119	– 0.221	9.061	2.049	1.019	0.855	0.443
	PI-1	0.044	0.101	0.439	– 0.123	– 0.212	7.995	2.188	1.281	1.036	0.436
12	PI-2	0.045	0.101	0.440	– 0.122	– 0.213	7.789	2.221	1.335	0.272	0.440
	SP-1	0.045	0.101	0.442	– 0.122	– 0.206	7.858	2.209	1.318	0.194	0.441
	SP-2	0.037	0.091	0.404*	– 0.115	– 0.195	4.815	3.714	2.872	0.176	0.404*
	1/p	0.044	0.100	0.441	– 0.130	– 0.202	12.000	1.732	2.548	0.194	0.440

*p value < 0.1, **p value < 0.05, ***p value < 0.01. The statistical results show the significance of the performance differences between BL and other models. For detailed p-values, see Table 10

Table 10 p values of differences in SR and NSR ($\lambda = 0.5$)

		BL	PI-1	PI-2	SP-1	SP-2	$1/p$
$p = 17$	SR	–	0.226	0.795	0.766	0.487	0.563
	NSR	–	0.199	0.980	0.971	0.544	0.731
$p = 12$	SR	–	0.163	0.437	0.608	0.052	0.612
	NSR	–	0.137	0.615	0.839	0.062	0.760

BL is the benchmark in the statistical test

where $\Phi(\cdot)$ denotes the cumulative density function of standard normal distribution, which finally completes the proof. □

A.5. Proof of Proposition 4

This proof mimics the proof of Theorem 1. The terminal stage problem is given by

$$\min_{\boldsymbol{x}_{T-1} \in \mathcal{X}_{T-1}} \lambda \left(\max_{1 \leqslant j \leqslant p} \gamma_{T-1} \tilde{q}_0^j x_{T-1}^j \right) - (1 - \lambda) \tilde{\boldsymbol{m}}_0^\top \boldsymbol{x}_{T-1} + \tilde{c}_{T-1} V_{T-1},$$

where $\tilde{\boldsymbol{v}}_{T-1} = \tilde{\boldsymbol{m}}_0$ and $\tilde{c}_{T-1} = -(1 - \lambda)$. We solve it using Lemma 1 and obtain the value function $\tilde{Q}_{T-1}(V_{T-1}) = V_{T-1}[\tilde{c}_{T-1} + \lambda \gamma_{T-1} \tilde{z}_{T-1} - (1 - \lambda) \tilde{y}_{T-1}]$, where $\tilde{y}_{T-1} = \tilde{z}_{T-1} \sum_{j \in \tilde{\mathcal{A}}_{T-1}^*} \tilde{v}_{T-1}^j / \tilde{q}_0^j$ and $\tilde{z}_{T-1} = 1 / (\sum_{j \in \tilde{\mathcal{A}}_{T-1}^*} 1/\tilde{q}_0^j)$. When $\gamma_{T-1} = 0$, set $\tilde{\mathcal{A}}_{T-1}^* = \{ i_p(\tilde{\boldsymbol{v}}_{T-1}) \}$. Next, we use mathematical induction to show that $\tilde{Q}_t(V_t) = V_t[\tilde{c}_t + \lambda \gamma_t \tilde{z}_t - (1 - \lambda) \tilde{y}_t]$ for all $t \in [0;T - 1]$ and give the optimal policy.

Suppose for period $t+1 \leq T-1$, it holds that $\tilde{Q}_{t+1}(V_{t+1}) = V_{t+1}[\tilde{c}_{t+1} + \lambda\gamma_{t+1}\tilde{z}_{t+1} - (1-\lambda)\tilde{y}_{t+1}]$. At the beginning of period t, we need to solve the following problem,

$$\min_{\boldsymbol{x}_t \in \mathcal{X}_t} \lambda \left(\max_{1 \leqslant j \leqslant p} \gamma_t \tilde{q}_0^j x_t^j \right) + \mathbb{E}[\tilde{Q}_{t+1}(V_{t+1})].$$

Given $\mathbb{E}(V_{t+1}) = V_t + \tilde{\boldsymbol{m}}_0^\top \boldsymbol{x}_t$, we have the following equivalent problem

$$\min_{\boldsymbol{x}_t \in \mathcal{X}_t} \lambda \left(\max_{1 \leqslant j \leqslant p} \gamma_t \tilde{q}_0^j x_t^j \right) - (1-\lambda)\tilde{\boldsymbol{v}}_t^\top \boldsymbol{x}_t + \tilde{c}_t V_t, \tag{36}$$

where $\tilde{c}_t = \tilde{c}_{t+1} + \lambda\gamma_{t+1}\tilde{z}_{t+1} - (1-\lambda)\tilde{y}_{t+1}$ and $\tilde{\boldsymbol{v}}_t = -\tilde{c}_t \tilde{\boldsymbol{m}}_0/(1-\lambda)$. According to Lemma 1, one can easily obtain the optimal solution of Problem (B1) as in Proposition 4.1. Then, the value function in stage t is $\tilde{Q}_t(V_t) = V_t[\tilde{c}_t + \lambda\gamma_t\tilde{z}_t - (1-\lambda)\tilde{y}_t]$, where $\tilde{y}_t = \tilde{z}_t \sum_{j \in \tilde{\mathcal{A}}_t^*} \tilde{v}_t^j/\tilde{q}_t^j$ and $\tilde{z}_t = 1/(\sum_{j \in \tilde{\mathcal{A}}_t^*} 1/\tilde{q}_t^j)$, which completes the proof. □

A.6. Supplementary numerical results

We further provide the out-of-sample test results for both cases $p = 17$ and $p = 12$ under the risk parameter $\lambda = 0.5$. The other settings are the same as those in Tables 3 and 5. Specifically, Table 9 reports the models' metric values on both $p = 17$ and $p = 12$ data sets and Table 10 shows the statistical results. We see that although BL still performs better than the benchmark models in various metrics, many performance gaps are not statistically significant for the 0.1 p value threshold. This is because the middle risk preference level $\lambda = 0.5$ is not risk-seeking enough for BL to exhibit the advantage of using Bayesian learning to account for parameter uncertainty over the plug-in models, and $\lambda = 0.5$ is also not risk-averse enough for BL to exhibit the advantage of using l_∞ risk function over the equal weight model.

Acknowledgements This study is supported in part by the National Natural Science Foundation of China (No. 71671146, 72331009), the Leading Talent Program of Guangdong Province (No. 2016LJ06D703), and the Shenzhen Science and Technology Innovation Committee (Grant No. JCYJ20210324115604012).

Funding This study is supported in part by the National Natural Science Foundation of China (No. 71671146, 72331009), the Leading Talent Program of Guangdong Province (No. 2016LJ06D703), and the Shenzhen Science and Technology Innovation Committee (Grant No. JCYJ20210324115604012).

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

References

Aguilar O, West M (2000) Bayesian dynamic factor models and portfolio allocation. J Bus Econ Stat 18(3):338–357

Anderson EW, Cheng AR (2016) Robust Bayesian portfolio choices. Rev Financ Stud 29(5):1330–1375

Barberis N (2000) Investing for the long run when returns are predictable. J Finance 55(1):225–264
Bauder D, Bodnar T, Parolya N, Schmid W (2021) Bayesian mean-variance analysis: optimal portfolio selection under parameter uncertainty. Quant Finance 21(2):221–242
Best MJ, Grauer RR (1991a) On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. Rev Financ Stud 4(2):315–342
Best MJ, Grauer RR (1991b) Sensitivity analysis for mean-variance portfolio problems. Manag Sci 37(8):980–989
Black F, Litterman R (1992) Global portfolio optimization. Financ Anal J 48(5):28–43
Bodnar T, Parolya N, Schmid W (2015) On the exact solution of the multi-period portfolio choice problem for an exponential utility under return predictability. Eur J Oper Res 246(2):528–542
Bodnar T, Mazur S, Okhrin Y (2017) Bayesian estimation of the global minimum variance portfolio. Eur J Oper Res 256(1):292–307
Boyd S, Busseti E, Diamond S, Kahn RN, Koh K, Nystrup P, Speth J (2017) Multi-period trading via convex optimization. arXiv preprint arXiv:1705.00109
Brandt MW, Goyal A, Santa-Clara P, Stroud JR (2005) A simulation approach to dynamic portfolio choice with an application to learning about return predictability. Rev Financ Stud 18(3):831–873
Brennan MJ (1998) The role of learning in dynamic portfolio decisions. Rev Finance 1(3):295–306
Cai X, Teo KL, Yang X, Zhou XY (2000) Portfolio optimization under a minimax rule. Manag Sci 46(7):957–972
Cai X, Teo KL, Yang X, Zhou XY (2004) Minimax portfolio optimization: empirical numerical study. J Oper Res Soc 55(1):65–72
Calafiore GC (2008) Multi-period portfolio optimization with linear control policies. Automatica 44(10):2463–2473
Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. J Portf Manag 19(2):6–11
Chopra VK, Hensel CR, Turner AL (1993) Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. Manag Sci 39(7):845–855
Clément E, Lamberton D, Protter P (2002) An analysis of a least squares regression method for American option pricing. Finance Stoch 6(4):449–471
DeMiguel V, Garlappi L, Uppal R (2009) Optimal versus Naive diversification: how inefficient is the 1/n portfolio strategy? Rev Financ Stud 22(5):1915–1953
Denault M, Simonato JG (2017) Dynamic portfolio choices by simulation-and-regression: revisiting the issue of value function vs portfolio weight recursions. Comput Oper Res 79:174–189
Diris B, Palm F, Schotman P (2014) Long-term strategic asset allocation: an out-of-sample evaluation. Manag Sci 61(9):2185–2202
Duffie D, Pan J (1997) An overview of value at risk. J Deriv 4(3):7–49
Frost PA, Savarino JE (1986) An empirical Bayes approach to efficient portfolio selection. J Financ Quant Anal 21(3):293–305
Jobson JD, Korkie BM (1981) Performance hypothesis testing with the Sharpe and Treynor measures. J Finance 36:889–908
Johannes M, Korteweg A, Polson N (2014) Sequential learning, predictability, and optimal portfolio returns. J Finance 69(2):611–644
Jurek JW, Viceira LM (2010) Optimal value and growth tilts in long-horizon portfolios. Rev Finance 15(1):29–74
Kolm P, Ritter G (2017) On the Bayesian interpretation of Black–Litterman. Eur J Oper Res 258(2):564–572
Konno H, Yamazaki H (1991) Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. Manag Sci 37(5):519–531
Lan C (2014) An out-of-sample evaluation of dynamic portfolio strategies. Rev Finance 19(6):2359–2399
Li D, Ng WL (2000) Optimal dynamic portfolio selection: multiperiod mean-variance formulation. Math Finance 10(3):387–406
Litterman B et al (2004) Modern investment management: an equilibrium approach, vol 246. Wiley, London
Liu YJ, Zhang WG (2015) A multi-period fuzzy portfolio optimization model with minimum transaction lots. Eur J Oper Res 242(3):933–941
Longstaff FA, Schwartz ES (2001) Valuing American options by simulation: a simple least-squares approach. Rev Financ Stud 14(1):113–147

Marisu GP, Pun CS (2023) Bayesian estimation and optimization for learning sequential regularized portfolios. SIAM J Financ Math 14(1):127–157
Markowitz H (1952) Portfolio selection. J Finance 7(1):77–91
Markowitz H (1959) Efficient diversification of investments. Wiley, New York
Memmel C (2003) Performance hypothesis testing with the Sharpe ratio. Finance Lett 21–23
Meng K, Yang H, Yang X, Wai Yu CK (2022) Portfolio optimization under a minimax rule revisited. Optimization 71(4):877–905
Merton RC (1969) Lifetime portfolio selection under uncertainty: the continuous-time case. Rev Econ Stat 51:247–257
Mulvey JM, Pauling WR, Madey RE (2003) Advantages of multiperiod portfolio models. J Portf Manag 29(2):35–45
Palczewski A, Palczewski J (2014) Theoretical and empirical estimates of mean-variance portfolio sensitivity. Eur J Oper Res 234(2):402–410
Park S, Song H, Lee S (2019) Linear programing models for portfolio optimization using a benchmark. Eur J Finance 25(5):435–457
Pástor L, Stambaugh RF (2000) Comparing asset pricing models: an investment perspective. J Financ Econ 56(3):335–381
Polson NG, Tew BV (2000) Bayesian portfolio selection: an empirical analysis of the S &P 500 index 1970–1996. J Bus Econ Stat 18(2):164–173
Prigent JL (2007) Portfolio optimization and performance analysis. Chapman and Hall/CRC, London
Rockafellar RT, Uryasev S et al (2000) Optimization of conditional value-at-risk. J Risk 2:21–42
Royden HL, Fitzpatrick P (1988) Real analysis, vol 32. Macmillan, New York
Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. Math Program 125:235–261
Ryals L, Dias S, Berger M (2007) Optimising marketing spend: return maximisation and risk minimisation in the marketing portfolio. J Mark Manag 23(9–10):991–1011
Shapiro A (2009) On a time consistency concept in risk averse multistage stochastic programming. Oper Res Lett 37(3):143–147
Skoulakis G (2008) Dynamic portfolio choice with Bayesian learning. Available at SSRN 1108816
Soyer R, Tanyeri K (2006) Bayesian portfolio selection with multi-variate random variance models. Eur J Oper Res 171(3):977–990
Stentoft L (2004) Convergence of the least squares Monte Carlo approach to American option valuation. Manag Sci 50(9):1193–1203
Sun Y, Aw G, Teo KL, Zhou G (2015) Portfolio optimization using a new probabilistic risk measure. J Ind Manag Optim 11(4):1275–1283
Tsitsiklis JN, Van Roy B (2001) Regression methods for pricing complex American-style options. IEEE Trans Neural Netw 12(4):694–703
Tsybakov AB (2008) Introduction to nonparametric estimation. Springer, Berlin
Tu J, Zhou G (2004) Data-generating process uncertainty: what difference does it make in portfolio decisions? J Financ Econ 72(2):385–421
van Binsbergen JH, Brandt MW (2007) Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function? Comput Econ 29(3–4):355–367
Vercher E, Bermúdez JD (2015) Portfolio optimization using a credibility mean-absolute semi-deviation model. Expert Syst Appl 42(20):7121–7131
Wang Z (2005) A shrinkage approach to model uncertainty and asset allocation. Rev Financ Stud 18:673–705
Winkler RL, Barry CB (1975) A Bayesian model for portfolio selection and revision. J Finance 30(1):179–192
Xia Y (2001) Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation. J Finance 56(1):205–246
Yu M, Takahashi S, Inoue H, Wang S (2010) Dynamic portfolio optimization with risk control for absolute deviation model. Eur J Oper Res 201(2):349–364
Zellner A (1996) An introduction to Bayesian inference in econometrics. Wiley, New York
Zhang R, Langrené N, Tian Y, Zhu Z, Klebaner F, Hamza K (2019) Dynamic portfolio optimization with liquidity cost and market impact: a simulation-and-regression approach. Quant Finance 19(3):519–532
Zhou G (2009) Beyond Black–Litterman: letting the data speak. J Portf Manag 36(1):36

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.