



HAL
open science

Portfolio diversification and model uncertainty: a robust dynamic mean-variance approach

Huyên Pham, Xiaoli Wei, Chao Zhou

► To cite this version:

Huyên Pham, Xiaoli Wei, Chao Zhou. Portfolio diversification and model uncertainty: a robust dynamic mean-variance approach. 2021. hal-01867133v3

HAL Id: hal-01867133

<https://hal.science/hal-01867133v3>

Preprint submitted on 1 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Portfolio diversification and model uncertainty: a robust dynamic mean-variance approach *

Huyên PHAM[†] Xiaoli WEI[‡] Chao ZHOU[§]

Abstract

This paper focuses on a dynamic multi-asset mean-variance portfolio selection problem under model uncertainty. We develop a continuous time framework for taking into account ambiguity aversion about both expected return rates and correlation matrix of the assets, and for studying the joint effects on portfolio diversification. The dynamic setting allows us to consider time varying ambiguity sets, which include the cases where the drift and correlation are estimated on a rolling window of historical data or when the investor takes into account learning on the ambiguity. In this context, we prove a general separation principle for the associated robust control problem, which allows us to reduce the determination of the optimal dynamic strategy to the parametric computation of the minimal risk premium function. Our results provide a justification for under-diversification, as documented in empirical studies and in the static models Garlappi et al. (2007), Liu & Zeng (2017). Furthermore, we explicitly quantify the degree of under-diversification in terms of correlation bounds and Sharpe ratios proximities, and emphasize the different features induced by drift and correlation ambiguity. In particular, we show that an investor with a poor confidence in the expected return estimation does not hold any risky asset, and on the other hand, trades only one risky asset when the level of ambiguity on correlation matrix is large. We also provide a complete picture of the diversification for the optimal robust portfolio in the three-asset case.

JEL Classification: G11, C61

MSC Classification: 91G10, 91G80, 60H30

Key words: Continuous-time Markowitz problem, model uncertainty, ambiguous drift and correlation, time varying ambiguity sets, separation principle, portfolio diversification.

*The authors benefited from the research grant for the project "Model uncertainty and stochastic control in financial risk management" under the aegis of the Alliance National University of Singapore (NUS)-Sorbonne Paris Cité (USPC). This work started when the first two authors were visiting the National University of Singapore (NUS), whose hospitality is kindly acknowledged. Part of this work was carried out while the third author was invited to the University Paris Diderot, whose hospitality is kindly appreciated.

[†]LPSM, Université de Paris and CREST-ENSAE, [pham at lpsm.paris](mailto:pham@lpsm.paris). The work of this author is supported by the ANR project CAESARS (ANR-15-CE05-0024), and also by FiME and the "Finance and Sustainable Development" EDF - CACIB Chair.

[‡]IEOR, UC Berkeley, [tyswxl at gmail.com](mailto:tyswxl@gmail.com).

[§]Department of Mathematics, National University of Singapore, [matzc at nus.edu.sg](mailto:matzc@nus.edu.sg). The work of this author is supported by Singapore MOE AcRF Grants R-146-000-271-112, R-146-000-255-114, R-146-000-219-112 and NSFC Grant 11871364.

1 Introduction

There are many studies on under-diversification of portfolio in the Finance and Economics literature, where investors hold only a small part of risky assets among a large number of available risky assets. In the extreme case the anti-diversification effect means that investors hold only a single stock (or not even any risky asset) and exclude many others. Empirical studies reported in numerous papers, see French & Poterba (1991), Cooper & Kaplanis (1994), Mitton & Vorkink (2007), Calvet et al. (2007), Guidolin & Liu (2016), have shown the evidence of portfolio under-diversification in practice. For example, in French & Poterba (1991), Cooper & Kaplanis (1994), it is observed that there exists a concentration on (bias towards) domestic assets compared to foreign assets in investors' international equity portfolios. These results are in contrast with the well-diversified portfolio suggested by the classical mean-variance portfolio theory initiated in a single period model in Markowitz (1959), later in Li & Ng (2000) for a multi-period model, and in Li & Zhou (2000) for a continuous-time model. A possible explanation to under-diversification is provided in the Finance and Economics literature by model uncertainty, often also called ambiguity or Knightian uncertainty.

In classical portfolio theory, the model and parameters are assumed to be perfectly known. However, in reality, due to statistical estimation issues, there is always uncertainty (ambiguity) about the model or parameters. In this case, a robust approach, see e.g. Ben-Tal et al. (2009), can be used to compute the optimal portfolio, i.e., the investor makes portfolio decisions under the worst case that corresponds to the least favorable scenario implied by a set of ambiguous parameters or by a set of distributions on the price process, which is usually refereed in operations research literature to distributionally robust optimization.

Abundant research has been conducted to tackle different types of model uncertainty. Robustness to uncertainty over a set of distributions on market factors in portfolio optimization has been analyzed in, e.g., Ghaoui et al. (2003), Natarajan et al. (2008, 2010), Delage & Ye (2010), Goh & Sim (2010), Wiesemann et al. (2014), Hanasusanto & Kuhn (2018), R. Jiang & Guan (2018), primarily in single-period formulations, except Glasserman & Xu (2013) in a multi-period setting. Primary market factor is the price process, and in this case, relevant sets of uncertain distributions correspond to ambiguity on the drift (i.e. the expected rate of return), the volatilities and the correlations when there are multiple assets to be traded. Indeed, drift estimation is known to be notoriously difficult: it is typically computed by maximum likelihood estimator (MLE) from historical time series of assets log-return, and we refer e.g. to Cont (2001) for general issues in statistical estimations of log-return, Chacko & Viceira (2003) for generalized method of moments to estimate drift, and the recent paper T. Bielecki et al. (2017) for the recursive construction of drift confidence region. Moreover, the estimation of correlation between assets may be extremely inaccurate, due to the asynchronous data and lead-lag effect, especially when the number of assets is large, see Cizeau et al. (2001); Liechty et al. (2004), Jagannathan & Ma (2003), Ledoit & Wolf (2004). Related works on robust portfolio optimization include Schied (2007), Jin & Zhou (2015) for uncertainty solely about drift, Denis & Kervarec (2013), Matoussi et al. (2015) for ambiguity about volatility (in a probabilistic setup) with a family of nondominated probability measures, Lin & Riedel (2014), Biagini & Pinar (2017) for combined uncertainty about both drift and volatility, Neufeld & Nutz (2018) for joint ambiguity about drift, volatility and jumps. In this existing literature, the common types of drift uncertainty sets are represented by polyhedral set or ellipsoidal set in Biagini & Pinar (2017), and unified by general ellipsoidal set in Garlappi et al. (2007). This general ellipsoidal representation for the drift ambiguity indicates in particular that drift estimation is affected by correlation estimation. Compared to drift ambiguity and volatility ambiguity, there are rather few results dealing with correlation ambiguity, let us mention however Fouque et al. (2016), J. Jiang & Tian (2016), Huang et al. (2017), Ismail & Pham (2019), and Liu & Zeng (2017).

Our purpose is to explore the joint effects of ambiguity about drift and correlation on portfolio selection and diversification with mean-variance (MV) criterion in continuous time. Notice that in the above cited papers, portfolio selection problems are mainly based on expected utility criterion and the effect on portfolio diversification under continuous-time framework is not really studied. Table 1 summarizes some papers on

Table 1: Some literature on the impact of model uncertainty on portfolio diversification

Objective	static MV	static utility	dynamic MV	dynamic utility
Ambiguity	Goldfarb & Iyengar (2003) (d, N)			Glasserman & Xu (2013) (d, N)
factors	Garlappi et al. (2007) (d, Y)			
drift	Boyle et al. (2012) (d, Y)	Dow & da Costa Werlang (1992) (1, Y)		Uppal & Wang (2003) (d, Y)
covariance (cov)			Ismail & Pham (2019) (d, N)	Matoussi et al. (2015) (d, N)
drift & cov				Biagini & Pinar (2017) (d, N)
drift & cov & jumps				Lin & Riedel (2014) (d, N)
				Neufeld & Nutz (2018) (d, N)
correlation(corr)	Liu & Zeng (2017) (d, Y)	Huang et al. (2017) (2, Y) J. Jiang & Tian (2016) (d, Y)	Ismail & Pham (2019) (2, N)	Fouque et al. (2016) (2, N)
drift & corr				This paper(d, Y)

In the above entries (.,.) the first element 1, 2 or d refers to the number of risky assets considered in the paper [.] , while the second element Y or N indicates whether the portfolio under-diversification is studied or not.

the model uncertainty and its impact on the portfolio diversification, which are related to the study of our paper. The list of related papers is not exhaustive. We distinguish between usual covariance ambiguity and correlation ambiguity in Table 1: Covariance ambiguity refers to the case when the covariance matrix lies between two given bounds $\underline{\Sigma}, \bar{\Sigma}$ in the space of positive definite symmetric matrix, see e.g. Lin & Riedel (2014), Matoussi et al. (2015), Neufeld & Nutz (2018), Yang et al. (2019), or lies in a proper cone, see e.g. Wiesemann et al. (2014). This is not easily interpretable in terms of correlation. Correlation ambiguity means that ambiguity is directly formulated on the correlations between the different assets. It turns out that, in contrast with covariance ambiguity, correlation ambiguity is a relevant indicator for generating under-diversification, as shown in a static model in Huang et al. (2017), J. Jiang & Tian (2016) and Liu & Zeng (2017). As pointed out in Table 1, the existing literature about under-diversification usually focuses on static mean variance criteria or expected utility criteria and on one type of model uncertainty.

Our paper considers both robust mean-variance and utility framework in continuous time to investigate the impact of combined drift and correlation ambiguity on portfolio diversification. Due to the nonlinear dependence on the wealth expectation, the mean-variance criterion is a non standard control problem. To circumvent this issue, the authors in Jin & Zhou (2015) reformulate the mean-variance problem under drift uncertainty into portfolio Sharpe ratio of the terminal wealth. Robust dynamic mean-variance problem under covariance matrix uncertainty, in particular, correlation ambiguity, has been considered in Ismail & Pham (2019) by a McKean-Vlasov dynamic programming approach, but the authors neither tackle the drift uncertainty nor study the portfolio diversification in detail, and mainly focus on the two-asset case $d = 2$. One key assumption in Ismail & Pham (2019) is that one can aggregate a family of processes, a condition, which does not hold true anymore in the case of drift uncertainty. An additional feature of our framework, compared to one-period models, is the consideration of learning on the ambiguity about the drift and correlation of risky assets: for instance, the investor typically gets more and more information about history of asset prices over time, and thus estimation errors about model parameters are reduced. Moreover, compared to models focusing only on one type of parameter ambiguity, our framework provides a unified setting for the joint effects of drift and correlation ambiguity on portfolio diversification. In particular, we are able to consider correlation structure of the assets in the drift uncertainty modelling.

Regarding portfolio diversification, the authors in Boyle et al. (2012), Garlappi et al. (2007), Uppal & Wang (2003) considered ambiguity about the assets' returns. Their frameworks include uncertainty about the joint distribution of returns for all assets and also for different levels of uncertainty of the marginal distribution

of returns for any subsets of these assets. They showed that the different levels of uncertainty on different asset subclasses could result in significant under-diversification. They also applied their theoretical results to real data and found consistent results with the empirical studies in Cooper & Kaplanis (1994), French & Poterba (1991) among others, showing that international equity portfolios are strongly biased towards domestic stocks, and in Huberman (2001) and Schultz (1996), where a similar lack of diversification is revealed on domestic portfolios. The model in Boyle et al. (2012), Garlappi et al. (2007), Uppal & Wang (2003) offers a partial explanation for the observed under-diversification and bias towards familiar securities. More recently, the authors in Liu & Zeng (2017) considered the uncertainty about the correlation of the assets. With a static mean-variance investment, they found that under-diversification of robust optimal portfolio depends on the level of correlation ambiguity. They also provided results with market data and showed, using their uncertain (ambiguous) correlation model, that the investor only holds less than 20 (17 stocks on average) among 100 stocks randomly selected from about 500 stocks in S&P500. In the two-asset case, they also found that the degree of diversification depends on the comparison between the ratio of assets' Sharpe ratios and the correlation ambiguity parameters. A further explanation for under-diversification is that investors can reduce the uncertainty on the model or the parameters through learning. In Van Nieuwerburgh & Veldkamp (2010), the authors built a framework to solve jointly for investment and information choices, with general preferences and information cost functions. They showed that, for some special preferences and information acquisition technologies, investors tend to learn more about the assets with which they are more familiar (typically, the domestic assets rather than the foreign ones), and become even more familiar with those assets after learning. As a consequence of this learning procedure, the investors select those assets they have learnt at the expense of others for which they have less information. Their results are consistent with the empirical studies on portfolios of international investors.

We emphasize that the consideration of joint ambiguity on the drift and correlation is relevant regarding portfolio diversification for several reasons:

- As the portfolio allocation is determined by both the drift and the correlation of the assets, it is consistent that under-diversification should be governed by the ambiguity levels on both these parameters: actually, the ambiguity level on the drift mainly determines absolute position of each asset (invest or not in a asset if the drift ambiguity is small or not), while the ambiguity level on the correlation mainly determines relative position of the assets (not diversify if the correlation is very ambiguous and diversify if not, directional trading if the correlation is small comparing to Sharpe ratio proximity and spread trading if large).
- In Uppal & Wang (2003), the authors consider both global ambiguity for the returns of all risky assets and different levels of ambiguity for any subset of these risky assets. In the case of equal ambiguity for all risky assets, ambiguity about the return would not bias the portfolio towards a particular asset, thus not explaining under-diversification. Instead, in the case of different levels of ambiguity for subsets of these risky assets, the ambiguity difference would bias the portfolio towards assets with smaller ambiguity, thus explaining under-diversification. In our setting, even in the case of equal ambiguity for the returns of all risky assets, we can still explain under-diversification through the ambiguity for the correlations of all assets. Moreover, in the case of different levels of ambiguity for the returns for subsets of these risky assets, we can explain under-diversification through the combination of the ambiguity difference for the returns and the ambiguity for correlations. In particular, we still obtain the optimal strategy in an explicit form, which allows us to understand the different effects of correlation ambiguity and expected return ambiguity.
- In Boyle et al. (2012), the familiarity and unfamiliarity of assets are modelled by different levels of ambiguity on expected rates of return in single period setting. The main feature of their model is that it allows investors to distinguish their ambiguity about one asset class relative to others. Moreover, similarly as in Uppal & Wang (2003), the different levels of ambiguity can explain bias to familiar

assets and under-diversification. The authors showed that the correlation coefficient has an important effect on the portfolio weights, notably, an increase in correlation from 50% to 70% roughly doubles the holding of the familiar asset. Since correlation coefficients are very difficult to estimate with a good accuracy, it is important to take into account correlation ambiguity when building the optimal portfolio which is less sensitive to estimation inaccuracy. This is our main motivation to consider ambiguity on both expected return rates and correlation and study the effects of ambiguity on portfolio strategy, in particular, under-diversification.

- Correlation ambiguity and expected return ambiguity have different features. For example, technically, in a continuous-time setting, a set of absolute continuous probability measures can be used to model the expected return ambiguity. In contrast, a set of mutually singular probability measures is needed for correlation ambiguity. This explains why in the existing literature, such as Ismail & Pham (2019) and Matoussi et al. (2015), only one type of uncertainty is considered. Economically, if the expected return ambiguity level is large, no risky asset is held (see e.g. Boyle et al. (2012)). In contrast, if correlation ambiguity is large, one and only one asset is held (see e.g. Liu & Zeng (2017)). However, by considering only correlation ambiguity as in Liu & Zeng (2017), one could not explain nonparticipation in which the investor does not make any risky investment at all.

To sum up, the contributions of our paper are fourfold:

(1) First, we develop a robust model that takes into account uncertainty on both drift and correlation of multiple risky assets for $d \geq 2$, in a dynamic, continuous time mean-variance portfolio setting. The dynamic setting allows us to consider time varying ambiguity sets, which include the cases where the drift and correlation are estimated on a rolling window of historical data or when the investor takes into account learning on the ambiguity.

(2) Secondly, we state a separation principle for the associated robust control problem formulated as a mean-field type differential game, which allows us to reduce the original min-max problem to the parametric computation of minimal risk premium. We derive the separation principle in a general setting with time varying ambiguity sets and for general preference criteria, including expected utility. In particular, the separation principle also holds true in single period and multi-period models. We can then generalize results in static models as in Boyle et al. (2012) and Garlappi et al. (2007) by incorporating correlation ambiguity, and study the implications of both expected return and correlation ambiguity. The main methodology for the separation principle is based on a weak version of the martingale optimality principle. This extends the classical martingale optimality principle that can not be directly applied in the presence of drift uncertainty, see detailed comments in Remark 3.5.

(3) Furthermore, we illustrate our results in rectangular and ellipsoidal uncertainty set and quantify explicitly the diversification effects on the optimal robust portfolio in terms of the ambiguity level. As in Uppal & Wang (2003) and Garlappi et al. (2007), the uncertainty set is flexible enough to allow for joint uncertainty for all assets or different levels of uncertainty for different subsets of the assets. Both drift uncertainty and correlation uncertainty can lead to under-diversification. In particular, we find that the robust investor does not trade in assets with large expected return ambiguity and trades only one risky asset in presence of high level of ambiguity about correlation. We also obtain closed-form expressions for the robust optimal portfolio, and we provide notably a complete picture of the diversification for the optimal robust portfolio in the case with three risky assets, which is new to the best of our knowledge. In the paper Liu & Zeng (2017) dealing with a rectangular ambiguity set in a static model, the authors proposed an implicit condition for not investing in one asset, i.e., under-diversification, in terms of correlation ambiguity. We provide in our dynamic setting an explicit condition in terms of ambiguity set for under-diversification, and obtain the optimal strategies in explicit form. For our future studies, we may introduce the information acquisition procedure as in Van Nieuwerburgh & Veldkamp (2010) or the multi-agent with heterogeneous beliefs setting as in J. Jiang & Tian (2016) in our framework.

(4) Finally, our results suggest that the diversification effect is mainly determined by the relation between

correlations of risky assets and Sharpe ratio "proximities" (the ratio of the Sharpe ratios). To the best of our knowledge, it is the first time that such an explicit relation is pointed out in a general setting. For simplicity, let us illustrate the idea and explain why classical portfolio theory does not lead to under-diversification in the two-asset case. Indeed, the Sharpe ratio proximity provides a standard to measure if the correlation is large or small. When the correlation is a constant, the investor chooses her portfolio allocation according to the relationship between the correlation and the Sharpe ratio proximity of the two assets. When the correlation is larger than the Sharpe ratio proximity, the optimal strategy is to make a spread trading. When the correlation is smaller than the Sharpe ratio proximity, the optimal strategy is to make a directional trading. Then the only case the investor invests in one asset is when the correlation is equal to the Sharpe ratio proximity. However, in practice, it is almost impossible to have this equality with parameters estimated from market data. However, in the case when correlation uncertainty set is an interval, the investor invests in one asset if the Sharpe ratio proximity lies in the correlation uncertainty interval, which is more likely to occur when the uncertainty interval is large due to lack of market data.

The rest of paper is organized as follows. Part I is devoted to the theoretical developments regarding the robust optimization problem, and contains two sections. Section 2 presents the formulation of the model uncertainty setting and the robust mean-variance problem. In Section 3, we derive the separation principle and explicit robust solution. Part II concerns the financial applications with two sections. Section 4 provides several examples arising from the separation principle, and the implications for the optimal robust portfolio strategy and the portfolio diversification. Section 5 illustrates through two numerical examples the effects of drift and correlation estimation error, thus ambiguity level, on portfolio Sharpe ratio. Finally, Appendix collects proofs of several mathematical results including the separation principle for general expected utility criteria.

Part I: Theoretical developments

2 Problem formulation

2.1 Model uncertainty setting

We consider a financial market with one risk-free asset, assumed to be constant equal to one, and d risky assets on a finite investment horizon $[0, T]$. Model uncertainty is formulated by using a probabilistic setup as in Neufeld & Nutz (2018). We define the canonical state space by $\Omega = \{\omega = (\omega(t))_{t \in [0, T]} \in C([0, T], \mathbb{R}^d) : \omega(0) = 0\}$ representing the continuous paths driving the risky assets. We equip Ω with the uniform norm and the corresponding Borel σ -field \mathcal{F} . We denote by $B = (B_t)_{t \in [0, T]}$ the canonical process, i.e., $B_t(\omega) = \omega(t)$, and by $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ the canonical filtration, i.e. the natural (raw) filtration generated by B .

We assume that the investor knows the marginal volatilities $\sigma_i > 0$ of each asset $i = 1, \dots, d$, typically through a quadratic variation estimation of the assets, and we denote by \mathfrak{S} the known constant¹ diagonal matrix with i -th diagonal term equal to σ_i , $i = 1, \dots, d$. However, there is uncertainty about the drift (expected rate of return) and the correlations of the multi-assets, which are parameters notoriously difficult to estimate in practice.

The ambiguity about drift and correlation matrix is parametrized by a family $\Theta = \{\Theta(t) : t \in [0, T]\}$ of nonempty sets with

$$\Theta(t) \subset \mathbb{R}^d \times \mathbb{C}_{>+}^d, \quad t \in [0, T],$$

where $\mathbb{C}_{>+}^d$ is the subset of all elements $\rho = (\rho_{ij})_{1 \leq i \neq j \leq d} \in [-1, 1]^{d(d-1)}$ with $\rho_{ij} = \rho_{ji}$ s.t. the symmetric

¹we could consider deterministic marginal volatilities, but this does not impact our results, and for simplicity of presentation, we take them constant.

matrix $C(\rho)$ with diagonal terms 1 and off-diagonal terms ρ_{ij}

$$C(\rho) = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1d} \\ \rho_{12} & 1 & \dots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1d} & \cdot & \dots & 1 \end{pmatrix}$$

lies in $\mathbb{S}_{>+}^d$, the set of positive definite symmetric matrices in $\mathbb{R}^{d \times d}$. Notice that $\mathbb{C}_{>+}^d$ is an open convex set of $[-1, 1]^{d(d-1)}$. The first component set of $\Theta(t)$ represents the values taken by the (possibly random) drift of the assets at time t , while the matrices $C(\rho)$, when ρ runs in the second component set of $\Theta(t)$, represent the correlation matrices of the multi-assets at time t . The introduction of a family of sets $\Theta(t)$, $0 \leq t \leq T$, allows us to take into account learning on the ambiguity about the mean return and correlations. The covariance matrices of the assets are given by

$$\Sigma(\rho) := \mathfrak{S}C(\rho)\mathfrak{S} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \dots & \sigma_1\sigma_d\rho_{1d} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \dots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_d\rho_{1d} & \cdot & \dots & \sigma_d^2 \end{pmatrix},$$

and we denote by $\sigma(\rho) = \Sigma^{\frac{1}{2}}(\rho) = (\Sigma(\rho))^{\frac{1}{2}}$ the square-root matrix, called volatility matrix.

An element $\theta = (\theta(t))_{t \in [0, T]} \in \Theta = \{\Theta(t) : t \in [0, T]\}$ can be viewed as a function on $[0, T]$ s.t. $\theta(t) \in \Theta(t)$ for all $t \in [0, T]$, and we write $\theta = (\mathbf{b}, \rho)$ to distinguish the first and second component of this function, with $\mathbf{b} = (b(t))_{t \in [0, T]}$ and $\rho = (\rho(t))_{t \in [0, T]}$.

Let us now introduce the family of (squared) risk premium $\mathbf{R}(\theta) = \{R(\theta(t)) : t \in [0, T]\}$, for $\theta = (\theta(t))_{t \in [0, T]} \in \Theta$, by

$$R(\theta) = \mathbf{b}^\top \Sigma(\rho)^{-1} \mathbf{b} = \|\sigma(\rho)^{-1} \mathbf{b}\|_2^2, \quad \text{for } \theta = (b, \rho) \in \Theta(t), \quad t \in [0, T], \quad (2.1)$$

where $\Sigma(\rho)^{-1} = (\Sigma(\rho))^{-1}$. Hereafter, \top denotes the transpose of matrix and $\|\cdot\|_2$ denotes the Euclidean norm in \mathbb{R}^d .

Remark 2.1 There exist different conditions for characterizing the positive definiteness of the correlation matrix $C(\rho)$. For example, Sylvester's criterion states that $C(\rho)$ is positive definite if and only if all the leading principal minors are positive, e.g., in dimension $d = 2$, $\rho \in (-1, 1)$; in dimension $d = 3$, $\rho_{ij} \in (-1, 1)$ $1 \leq i < j \leq 3$ and $\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 1 - 2\rho_{12}\rho_{13}\rho_{23} < 0$. Alternatively, one can characterize the positive definiteness of $C(\rho)$ using angular coordinates as in Rapisarda et al. (2007). Instead of working directly with $C(\rho) \in \mathbb{S}_{>+}^d$ in the form of matrix, we characterize $C(\rho) \in \mathcal{S}_{>+}^d$ in a more explicit way in terms of parameter $\rho \in \mathbb{C}_{>+}^d$ via Sylvester's criterion. \diamond

The ambiguity sets $\Theta = \{\Theta(t) : t \in [0, T]\}$ for the drift and correlation are assumed to satisfy

$$(\mathbf{H}\Theta) \quad t \mapsto \Theta(t) \text{ is measurable}^2, \text{ and } \Theta(t) \text{ is a bounded convex set of } \mathbb{R}^d \times \mathbb{C}_{>+}^d, \quad t \in [0, T].$$

A relevant class for practical applications of ambiguity sets Θ satisfying $(\mathbf{H}\Theta)$ is the following. Let $\{J_1, \dots, J_l, \dots, J_p\}$, $1 \leq p \leq d$, be a partition of $\{1, \dots, d\}$, and denote by $|J_l|$ the cardinality of J_l , $l = 1, \dots, p$. We consider ambiguity set $\Theta(t)$ in the form

$$\Theta(t) = \{(b, \rho) \in \mathbb{R}^d \times \Gamma(t) : \|\sigma_{J_l}(\rho)^{-1}(b_{J_l} - \hat{b}_{J_l}(t))\|_2 \leq \delta_l(t), \quad l = 1, \dots, p\}, \quad (2.2)$$

²The set-valued map $\Theta : [0, T] \rightarrow 2^{\mathbb{R}^d \times \mathbb{C}_{>+}^d}$ is said to be measurable if and only if for every open set $U \subset \mathbb{R}^d \times \mathbb{C}_{>+}^d$, $\{t \in [0, T] : \Theta(t) \cap U \neq \emptyset\} \in \mathcal{B}([0, T])$. Here $2^{\mathbb{R}^d \times \mathbb{C}_{>+}^d}$ represents the family of all subsets in $\mathbb{R}^d \times \mathbb{C}_{>+}^d$ and $\mathcal{B}([0, T])$ is Borel σ -algebra on $[0, T]$. See e.g., Fryszkowski (2004) for more details on the measurability of set-valued maps.

for some convex set $\Gamma(t)$ of $\mathbb{C}_{>+}^d$, where $\Sigma_{J_l}(\rho)$ is the $|J_l| \times |J_l|$ variance-covariance matrix of assets in subclass J_l and its square root $\sigma_{J_l}(\rho) = (\Sigma_{J_l}(\rho))^{\frac{1}{2}}$. Here $\hat{b}_{J_l}(t)$ is a known vector, representing an estimate of mean return vector b_{J_l} of assets in J_l at time t , and $\delta_l(t) \geq 0$ represents a level of ambiguity around $\hat{b}_{J_l}(t)$ due to estimation error as well as her level of uncertainty aversion.

Remark 2.2 In the particular case when the number of subclasses is equal to the number of risky assets, i.e., $p = d$, $\Theta(t)$ is a rectangular set in the form $\prod_{i=1}^d [\hat{b}_i(t) - \sigma_i \delta_i(t), \hat{b}_i(t) + \sigma_i \delta_i(t)] \times \Gamma(t)$, interpreted as a product set of uncertainty regions with size determined by the level $\delta_i(t)$ for each asset $i = 1, \dots, d$. Instead of setting uncertainty regions for the assets individually, one can do it jointly for all assets by considering the case when $p = 1$, which corresponds to an ellipsoidal set in the form $\{(b, \rho) \in \mathbb{R}^d \times \Gamma(t) : \|\sigma(\rho)^{-1}(b - \hat{b}(t))\|_2 \leq \delta(t)\}$. An extension of the two above sub-cases, allowing for separate estimation and uncertainty regions for different subclasses of assets (due e.g. to different available histories across the assets) is to consider ambiguity sets as in (2.2). This is an extension of expected rate of return uncertainty considered in Ben-Tal et al. (2009), Garlappi et al. (2007) for a single period model by allowing an additional ambiguity on correlation and learning on estimation error through the deterministic level $\delta_l(t)$.

Theoretically, $\Gamma(t)$ can be in the rectangular form $\{\rho \in (-1, 1)^d : \rho_{ij} \in [\hat{\rho}_{ij}(t) - \epsilon_{ij}(t), \hat{\rho}_{ij}(t) + \epsilon_{ij}(t)]\}$, where $\hat{\rho}(t) = (\hat{\rho}_{ij}(t))_{1 \leq i, j \leq d}$ is estimator of correlation, and $\epsilon_{ij}(t)$ represents uncertainty level around $\hat{\rho}_{ij}(t)$. This rectangular correlation set falls in $\mathbb{C}_{>+}^d$ with suitable choice of $\epsilon_{ij}(t)$. When $\Gamma(t) = \mathbb{C}_{>+}^d$, this means that the investor has at time t a full ambiguity about correlation on the d -risky assets. In the opposite case, when $\Gamma(t)$ is a singleton, this means that the investor knows (or is fully confident about) the value of the correlation at time t . Similarly, the case $\delta_l(t) = 0$ means that the mean return vector b_{J_l} for assets in the subclass J_l is known or the investor is fully confident about her estimate at time t . \diamond

Remark 2.3 An interesting extension of our model uncertainty setting would be to consider ambiguity sets that may evolve randomly in time $\Theta(t, \omega)$, e.g., through a factor process or price, and in this case, this would add an additional state variable in the value function and the optimal strategy. For example, in T. R. Bielecki et al. (2019) the threshold δ may depend on the current and past asset prices, which corresponds to an adaptive estimation error from the information flow of the observed asset prices. However, it is not immediate how to extend in this case the weak martingale optimality principle for characterizing the optimal strategy, and whether the separation theorem still holds. \diamond

We denote by \mathcal{V}_Θ the set of \mathbb{F} -progressively measurable processes $\theta. = (\theta_t)_t = (b_t, \rho_t)_t = (b., \rho.)$ valued in Θ , in the sense that θ_t is valued in $\Theta(t)$, $0 \leq t \leq T$, and introduce the set of probability measures \mathcal{P}^Θ :

$$\mathcal{P}^\Theta = \{\mathbb{P}^\theta : \theta. \in \mathcal{V}_\Theta\},$$

where \mathbb{P}^θ is the probability measure on (Ω, \mathcal{F}) s.t. B is a semimartingale on $(\Omega, \mathcal{F}, \mathbb{P}^\theta)$ with absolutely continuous characteristics (w.r.t. the Lebesgue measure dt) $(b., \Sigma(\rho.))$. The probabilities \mathbb{P}^θ are in general non-equivalent, and actually mutually singular, and we say that a property holds \mathcal{P}^Θ -quasi surely (\mathcal{P}^Θ -q.s. in short) if it holds \mathbb{P}^θ -a.s. for all $\theta. \in \mathcal{V}_\Theta$.

The price process $S = (S^1, \dots, S^d)$ of the d risky assets valued in $(0, \infty)^d$ is given by the dynamics

$$\begin{aligned} dS_t &= \text{diag}(S_t)dB_t, \quad 0 \leq t \leq T, \quad \mathcal{P}^\Theta - q.s. \\ &= \text{diag}(S_t)(b_t dt + \sigma(\rho_t)dW_t^\theta), \quad \mathbb{P}^\theta - a.s., \quad \text{for } \theta. = (b., \rho.) \in \mathcal{V}_\Theta, \end{aligned}$$

where W^θ is a d -dimensional Brownian motion under \mathbb{P}^θ . Here $\text{diag}(S_t)$ is the diagonal matrix with i -th element equal to S_t^i . Notice that in this uncertainty modeling, we allow the unknown drift and correlation to be a priori random process, valued in Θ .

2.2 Robust mean-variance problem

An admissible portfolio strategy $\alpha = (\alpha_t)_{0 \leq t \leq T}$ representing the amount invested in the d risky assets, is a \mathbb{R}^d -valued \mathbb{F} -progressively measurable process, satisfying the integrability condition

$$\sup_{\mathbb{P}^\theta \in \mathcal{P}^\Theta} \mathbb{E}_\theta \left[\int_0^T |\alpha_t^\top b_t| dt + \int_0^T \alpha_t^\top \Sigma(\rho_t) \alpha_t dt \right] < \infty, \quad (2.3)$$

and denoted by $\alpha \in \mathcal{A}$. Hereafter, \mathbb{E}_θ denotes the expectation under \mathbb{P}^θ . This integrability condition (2.3) ensures that $\text{diag}(S)^{-1}\alpha$ is S -integrable under any $\mathbb{P} \in \mathcal{P}^\Theta$. For a portfolio strategy $\alpha \in \mathcal{A}$, and an initial capital $x_0 \in \mathbb{R}$, the dynamics of the self-financed wealth process is driven by

$$\begin{aligned} dX_t^\alpha &= \alpha_t^\top \text{diag}(S_t)^{-1} dS_t = \alpha_t^\top dB_t, \quad 0 \leq t \leq T, \quad X_0^\alpha = x_0, \quad \mathcal{P}^\Theta - q.s. \\ &= \alpha_t^\top (b_t dt + \sigma(\rho_t) dW_t^\theta), \quad 0 \leq t \leq T, \quad X_0^\alpha = x_0 \in \mathbb{R}, \quad \mathbb{P}^\theta - a.s. \end{aligned} \quad (2.4)$$

for all $\theta = (b, \rho) \in \mathcal{V}_\Theta$.

Given a risk aversion parameter $\lambda > 0$, the worst-case mean-variance functional under ambiguous drift and correlation is

$$J_{wc}(\alpha) = \inf_{\mathbb{P}^\theta \in \mathcal{P}^\Theta} \left(\mathbb{E}_\theta [X_T^\alpha] - \lambda \text{Var}_\theta (X_T^\alpha) \right) < \infty, \quad \alpha \in \mathcal{A},$$

where $\text{Var}_\theta(\cdot)$ denotes the variance under \mathbb{P}^θ , and the robust mean-variance portfolio selection is formulated as

$$\left\{ V_0 := \sup_{\alpha \in \mathcal{A}} J_{wc}(\alpha) = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} J(\alpha, \theta) \right. \quad (2.5)$$

Notice that problem (2.5) is a non standard stochastic differential game due to the presence of the variance term in the criterion, which prevents the use of classical control method by dynamic programming or maximum principle. We complete this section by recalling the solution to the mean-variance problem when there is no ambiguity on the model parameters, and which will serve later as benchmark for comparison when studying the uncertainty case.

2.3 Case of no model uncertainty

When $\Theta(t) = \{\theta^\circ(t) = (b^\circ(t), \rho^\circ(t))\}$ is a singleton for any $t \in [0, T]$, we are reduced to the Black-Scholes model with time varying deterministic drift $b^\circ(t)$, deterministic covariance matrix $\Sigma^\circ(t) := \Sigma(\rho^\circ(t))$, volatility $\sigma^\circ(t) := \sigma(\rho^\circ(t))$, and deterministic risk premium $R^\circ(t) := R(\theta^\circ(t))$. In this case, two notions of strategy are adopted, see e.g. Li & Zhou (2000); Fischer & Livieri (2016); Pham & Wei (2017) for pre-committed strategy, which we consider here, and e.g. Hu et al. (2012); Björk et al. (2017) for equilibrium. It is known that the optimal mean-variance strategy is given in feedback form by

$$\alpha_t^* = \Lambda^\circ(X_t^*) (\Sigma^\circ(t))^{-1} b^\circ(t), \quad 0 \leq t \leq T,$$

where X^* is the wealth process associated to α^* , and $\Lambda^\circ(X_t^*) > 0$ with

$$\Lambda^\circ(x) := x_0 + \frac{e^{\int_0^T R^\circ(t) dt}}{2\lambda} - x, \quad x \in \mathbb{R},$$

while the optimal performance value is

$$V_0 = x_0 + \frac{1}{4\lambda} \left[e^{\int_0^T R^\circ(t) dt} - 1 \right].$$

The vector $(\Sigma^o(t))^{-1}b^o(t)$, which depends only on the model parameters of the risky assets, determines the allocation in the risky assets. The above expression of α^* shows that, once we know the exact values of the rate of return and covariance matrix, one diversifies her portfolio among all the assets according to the components of the vector $(\Sigma^o(t))^{-1}b^o(t)$, and this is weighted by the scalar term $\Lambda^o(X_t^*)$, which depends on the risk aversion of the investor via the parameter λ , on the current wealth but also on the initial capital x_0 (which is sometimes referred to as the pre-commitment of the mean-variance criterion). Notice that $\Lambda^o(X_t^*)$ is positive. Indeed, observe that

$$\begin{aligned} d\Lambda^o(X_t^*) &= -dX_t^* &= -(\alpha_t^*)^\top (b^o(t)dt + \sigma^o(t)dW_t^o) \\ &= -\Lambda^o(X_t^*)(R^o(t)dt + ((\sigma^o(t))^{-1}b^o(t))^\top dW_t^o), \quad 0 \leq t \leq T, \end{aligned}$$

with $\Lambda^o(X_0^*) = \frac{1}{2\lambda}e^{\int_0^T R^o(t)dt} > 0$, which shows clearly that $\Lambda^o(X_t^*) > 0$, $0 \leq t \leq T$, and decreases with λ .

Let us discuss in particular the allocation in the two-asset case. Notice that the vector $(\Sigma^o(t))^{-1}b^o(t)$ of allocation is then given by

$$(\Sigma^o(t))^{-1}b^o(t) = \frac{1}{1 - |\rho^o(t)|^2} \begin{pmatrix} \frac{\beta_1^o(t) - \rho^o(t)\beta_2^o(t)}{\frac{\sigma_1}{\beta_1^o(t)}} \\ \frac{\beta_2^o(t) - \rho^o(t)\beta_1^o(t)}{\frac{\sigma_2}{\beta_2^o(t)}} \end{pmatrix} =: \begin{pmatrix} \kappa_1^o(t) \\ \kappa_2^o(t) \end{pmatrix},$$

where $\beta_i^o(t) = b_i^o(t)/\sigma_i$ is the Sharpe ratio of the i -th asset, $i = 1, 2$, at time t . To fix the idea, assume that $\beta_1^o(t) > \beta_2^o(t) > 0$. We then see that $\kappa_1^o(t) > 0$, while $\kappa_2^o(t) \geq 0$ if and only if $\frac{\beta_2^o(t)}{\beta_1^o(t)} \geq \rho^o(t)$. The interpretation is the following: the ratio $\frac{\beta_2^o(t)}{\beta_1^o(t)} \in (0, 1)$ measures the ‘‘proximity’’ in terms of Sharpe ratio between the two assets, and has to be compared with the correlation $\rho^o(t)$ between these assets in order to determine whether it is optimal to invest according to a directional trading, i.e., $\kappa_1^o(t)\kappa_2^o(t) > 0$ (thus here long in both assets) or according to a spread trading, i.e., $\kappa_1^o(t)\kappa_2^o(t) < 0$ (long in the first asset and short in the second one) or according to under-diversification, i.e., $\kappa_1^o(t)\kappa_2^o(t) = 0$ (only long in the first asset). Notice that under-diversification only occurs when $\rho^o(t) = \frac{\beta_2^o(t)}{\beta_1^o(t)}$, a condition ‘‘rarely’’ satisfied in practice. For example, when both assets have close Sharpe ratios, and their correlation is not too high, then one optimally invests in both assets with a directional trading. In contrast, when one asset has a much larger Sharpe ratio than the other one, or when the correlation between the assets is high, then one optimally invests in both assets with a spread trading.

In the sequel, we study the quantitative impact of the drift and correlation uncertainty on the optimal robust mean-variance strategy, in particular regarding the portfolio diversification.

3 Separation principle and robust solution

The main result of this section is to state a separation principle for solving the robust dynamic mean-variance problem.

Theorem 3.1 (Separation Principle) *Let us consider a parametric set Θ for model uncertainty as in $(\mathbf{H}\Theta)$. Suppose that there exists a pair $\theta^* = (\theta^*(t))_t = (\mathbf{b}^*, \boldsymbol{\rho}^*) = (b^*(t), \rho^*(t))_t \in \Theta$ solution to $\arg \min_{\theta \in \Theta} \mathbf{R}(\theta)$, i.e., $\theta^*(t) \in \arg \min_{\theta \in \Theta(t)} R(\theta)$, for all $t \in [0, T]$. Then the robust mean-variance problem (2.5) admits an optimal portfolio strategy given in feedback form by*

$$\alpha_t^* = \Lambda^*(X_t^*)\Sigma(\rho^*(t))^{-1}b^*(t), \quad 0 \leq t \leq T, \quad \mathcal{P}^\Theta - q.s., \quad (3.6)$$

where X^* is the state process associated to α_t^* , and $\Lambda^*(X_t^*) > 0$ with

$$\Lambda^*(x) := x_0 + \frac{e^{\int_0^T R(\theta^*(s))ds}}{2\lambda} - x, \quad x \in \mathbb{R}. \quad (3.7)$$

Moreover, the corresponding initial value function is

$$V_0 = x_0 + \frac{1}{4\lambda} [e^{\int_0^T R(\theta^*(s)) ds} - 1].$$

Interpretation. Theorem 3.1 means that the robust mean-variance problem (2.5) can be solved in two steps according to a separation principle. (i) First, at each time $t \in [0, T]$, we search for the infimum of the risk premium function $\theta \in \Theta(t) \mapsto R(\theta)$ as defined in (2.1), which depends only on the inputs of the uncertainty model. Existence and explicit determination of an element $\theta^* = (\mathbf{b}^*, \rho^*) \in \Theta$ attaining this infimum will be discussed and illustrated all along the paper through several examples. (ii) The solution to (2.5) is then given by the solution to the mean-variance problem in the Black-Scholes model with time varying deterministic drift $b^*(t)$ and correlation $\rho^*(t)$, see Section 2.3, and the worst-case scenario of the robust dynamic mean-variance problem is simply given by the family of deterministic parameters $\theta^* = (\mathbf{b}^*, \rho^*)$. Some interesting features show up, especially regarding portfolio diversification, as detailed in the next section. \diamond

Remark 3.1 The existence of the solution θ^* to $\arg \min_{\theta \in \Theta} R(\theta)$ is guaranteed under (2.2) whenever the ambiguity sets $\Gamma(t)$, $t \in [0, T]$, on correlation are compact as the risk premium function R is continuous. Since we also want to consider the case of full ambiguity on correlation, i.e., when $\Gamma(t) = \mathbb{C}_{>+}^d$, which is an open set, we do not impose such compactness condition. \diamond

Remark 3.2 (Relation with static model) Actually, the worst-case scenario of static robust mean-variance problem is also determined by the minimal risk premium. By analogue with model uncertainty described as in Section 2.1, we characterize ambiguity about the rate of return b and correlation ρ as a bounded convex set, i.e. $\theta = (b, \rho) \in \Theta$, and formulate the single-period robust mean-variance problem under ambiguity as

$$\sup_{\alpha \in \mathbb{R}^d} \inf_{\theta \in \Theta} (\alpha^\top b - \lambda \alpha^\top \Sigma(\rho) \alpha) =: \sup_{\alpha \in \mathbb{R}^d} \inf_{\theta \in \Theta} J_{static}(\alpha, \theta).$$

where α is the portfolio held in d risky assets. Assume that there exists $\theta^* = (b^*, \rho^*) \in \arg \min_{\theta \in \Theta} R(\theta)$, we then obtain from Lemma A.1 in Appendix that (α^*, θ^*) with $\alpha^* := \frac{1}{2\lambda} \Sigma(\rho^*)^{-1} b^*$ is a saddle point of $J(\alpha, \theta)$, and

$$\sup_{\alpha \in \mathbb{R}^d} \inf_{\theta \in \Theta} J_{static}(\alpha, \theta) = \inf_{\theta \in \Theta} \sup_{\alpha \in \mathbb{R}^d} J_{static}(\alpha, \theta) = J_{static}(\alpha^*, \theta^*) = \frac{1}{4\lambda} R(\theta^*).$$

Therefore, the key point in all single period, multi-period and continuous-time models is the computation of infimum of risk premium, which will be discussed in the next section. \diamond

The rest of this section is devoted to the proof of Theorem 3.1, and the methodology is based on the following weak version of the martingale optimality principle. The usual martingale optimality principle is introduced in El Karoui (1981) for optimal stopping problems, and later applied in Rouge & El Karoui (2000); Hu et al. (2005) for utility maximization, in Matoussi et al. (2015) for robust utility maximization under the uncertain volatility model, and in Yang et al. (2019) for robust portfolio-consumption strategies with uncertainty on both drift and volatility, to name a few. A comparison of the usual martingale optimality principle and our weak version is given in Remark 3.3 below.

Lemma 3.1 (Weak optimality principle) *Let $\{V_t^{\alpha, \theta}, t \in [0, T], \alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta\}$ be a family of real-valued processes in the form*

$$V_t^{\alpha, \theta} : = v_t(X_t^\alpha, \mathbb{E}_\theta[X_t^\alpha]),$$

for some measurable functions v_t on $\mathbb{R} \times \mathbb{R}$, $t \in [0, T]$, such that :

(i) $v_T(x, \bar{x}) = x - \lambda(x - \bar{x})^2$, for all $x, \bar{x} \in \mathbb{R}$,

(ii) the function $t \in [0, T] \mapsto \mathbb{E}_{\theta^*}[V_t^{\alpha, \theta^*}]$ is nonincreasing for all $\alpha \in \mathcal{A}$ and some $\theta^* \in \mathcal{V}_\Theta$,

(iii) $\mathbb{E}_\theta[V_T^{\alpha^*, \theta} - V_0^{\alpha^*, \theta}] \geq 0$, for some $\alpha^* \in \mathcal{A}$ and all $\theta \in \mathcal{V}_\Theta$.

Then, α^* is an optimal portfolio strategy for the robust mean-variance problem (2.5) with a worst-case scenario θ^* , and

$$\begin{aligned} V_0 &= J_{wc}(\alpha^*) = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} J(\alpha, \theta) = \inf_{\theta \in \mathcal{V}_\Theta} \sup_{\alpha \in \mathcal{A}} J(\alpha, \theta) = v_0(x_0, x_0) \\ &= J(\alpha^*, \theta^*). \end{aligned} \quad (3.8)$$

Proof. First, observe that $V_0^{\alpha, \theta} = v_0(x_0, x_0)$ is a constant that does not depend on α, θ , and from condition (i) that $\mathbb{E}_\theta[V_T^{\alpha, \theta}] = J(\alpha, \theta)$ for all $\alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta$. Then, from condition (ii), we see that

$$v_0(x_0, x_0) = \mathbb{E}_{\theta^*}[V_0^{\alpha, \theta^*}] \geq \mathbb{E}_{\theta^*}[V_T^{\alpha, \theta^*}] = J(\alpha, \theta^*),$$

for all $\alpha \in \mathcal{A}$, and thus $v_0(x_0, x_0) \geq \sup_{\alpha \in \mathcal{A}} J(\alpha, \theta^*) \geq \inf_{\theta \in \mathcal{V}_\Theta} \sup_{\alpha \in \mathcal{A}} J(\alpha, \theta)$. Similarly, from condition (iii), we have

$v_0(x_0, x_0) \leq J(\alpha^*, \theta)$ for all $\theta \in \mathcal{V}_\Theta$, and thus $v_0(x_0, x_0) \leq \inf_{\theta \in \mathcal{V}_\Theta} J(\alpha^*, \theta) = J_{wc}(\alpha^*) \leq \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} J(\alpha, \theta)$.

Recalling that we always have $\sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} J(\alpha, \theta) \leq \inf_{\theta \in \mathcal{V}_\Theta} \sup_{\alpha \in \mathcal{A}} J(\alpha, \theta)$, we obtained the required equality in

(3.8). Then, finally, from (ii) with α^* and (iii) with θ^* , we obtain that $v_0(x_0, x_0) = J(\alpha^*, \theta^*)$. \square

Remark 3.3 The usual martingale optimality principle for stochastic differential games as in robust portfolio selection problem, and with classical expected utility criterion for some nondecreasing and concave utility function U on \mathbb{R} , e.g., $U(x) = -e^{-\eta x}$, $\eta > 0$:

$$\sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} \mathbb{E}_\theta[U(X_T^\alpha)],$$

would consist of finding a family of processes $V_t^{\alpha, \theta}$ in the form $v_t(X_t^\alpha)$ for some measurable functions v_t on \mathbb{R} s.t. (i) $v_T(x) = U(x)$, (ii') the process $(V_t^{\alpha, \theta^*})_t$ is a supermartingale under \mathbb{P}_{θ^*} for all α and some θ^* , and (iii') the process $(V_t^{\alpha^*, \theta})_t$ is a submartingale under \mathbb{P}_θ for some α^* and all θ . Due to the nonlinear dependence on the law of the state wealth process via the variance term in the mean-variance criterion, making the problem *a priori* time inconsistent, we have to adopt a weaker version of the optimality principle. First, the functions v_t depend not only on the state process X_t^α but also on its mean $\mathbb{E}_\theta[X_t^\alpha]$. Second, we replace condition (ii') by the weaker condition (ii) on the mean in Lemma 3.1. Third, condition (iii') is substituted by the weaker condition (iii), which is even weaker than (iii') $t \mapsto \mathbb{E}_\theta[V_t^{\alpha^*, \theta}]$ is nondecreasing for some α^* and all θ . This asymmetry of condition between (ii) and (iii) is explained in more detail in Remark 3.5. \diamond

We shall also use the following property on the infimum of the risk premium function.

Lemma 3.2 Given Θ as in $(\mathbf{H}\Theta)$, and assuming that there exists $\theta^* = (b^*, \rho^*) \in \arg \min_{\theta \in \Theta} \mathbf{R}(\theta)$, let us define the function H_t on $\Theta(t)$, $t \in [0, T]$, by

$$H_t(\theta) := b^\top \Sigma(\rho^*(t))^{-1} \Sigma(\rho) \Sigma(\rho^*(t))^{-1} b^*(t), \quad \text{for } \theta = (b, \rho) \in \Theta(t). \quad (3.9)$$

Then, we have for all $(b, \rho) \in \Theta(t)$:

$$\mathbf{R}(\theta^*(t)) - 2H_t(b, \rho^*(t)) + H_t(b^*(t), \rho) \leq 0. \quad (3.10)$$

Proof. See Section B.1 in Appendix.

In the following, we provide details of the proof for Theorem 3.1 using Lemma 3.1 and Lemma 3.2.

Proof of Theorem 3.1. We aim to construct a family of processes $\{V_t^{\alpha, \theta}, t \in [0, T], \alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta\}$ as in Lemma 3.1, and given the linear-quadratic structure of our optimization problem, we look for measurable functions v_t in the form:

$$v_t(x, \bar{x}) = K_t(x - \bar{x})^2 + Y_t x + \chi_t, \quad t \in [0, T], (x, \bar{x}) \in \mathbb{R}^2, \quad (3.11)$$

for some deterministic processes $(K_t, Y_t, \chi_t)_t$ to be determined. Condition (i) in Lemma 3.1 fixes the terminal condition

$$K_T = -\lambda, \quad Y_T = 1, \quad \chi_T = 0. \quad (3.12)$$

We now consider $\theta^* \in \Theta$ as in Theorem 3.1, hence defining in particular a deterministic process $\theta^* = (\theta^*(t))_t \in \mathcal{V}_\Theta$, and α^* given by (3.6). Let us first check that $\alpha^* \in \mathcal{A}$. The corresponding wealth process X^* satisfies under any \mathbb{P}^θ , $\theta = (b, \rho) \in \mathcal{V}_\Theta$, a linear stochastic differential equation with bounded random coefficients (notice that b and $\sigma(\rho)$ are bounded processes), and thus by standard estimates: $\mathbb{E}_\theta[\sup_{0 \leq t \leq T} |X_t^*|^2] \leq C(1 + |x_0|^2)$ for some constant C independent of $\theta \in \mathcal{V}_\Theta$. It follows immediately that α^* satisfies the integrability condition in (2.3), i.e., $\alpha^* \in \mathcal{A}$.

The main issue now is to show that such a pair (α^*, θ^*) satisfies conditions (ii)-(iii) of Lemma 3.1.

• *Step 1. condition (ii) of Lemma 3.1.*

For any $\alpha \in \mathcal{A}$, with associated wealth process $X := X^\alpha$, let us compute the derivative of the deterministic function $t \mapsto \mathbb{E}_{\theta^*}[V_t^{\alpha, \theta^*}] = \mathbb{E}_{\theta^*}[v_t(X_t, \mathbb{E}_{\theta^*}[X_t])]$ with v_t as in (3.11). From the dynamics of $X = X_t^\alpha$ in (2.4) under \mathbb{P}^{θ^*} and by applying Itô's formula, we obtain

$$\begin{aligned} \frac{d\mathbb{E}_{\theta^*}[X_t]}{dt} &= \mathbb{E}_{\theta^*}[\alpha_t^\top b^*(t)], \\ \frac{d\text{Var}_{\theta^*}(X_t)}{dt} &= 2\text{Cov}_{\theta^*}(X_t, \alpha_t^\top b^*(t)) + \mathbb{E}_{\theta^*}[\alpha_t^\top \Sigma(\rho^*(t)) \alpha_t]. \end{aligned}$$

From the quadratic form of v_t in (3.11), with (K, Y, χ) differentiable in time, we then have

$$\begin{aligned} \frac{d\mathbb{E}_{\theta^*}[V_t^{\alpha, \theta^*}]}{dt} &= \frac{d\mathbb{E}_{\theta^*}[v_t(X_t, \mathbb{E}_{\theta^*}[X_t])]}{dt} \\ &= \dot{K}_t \text{Var}_{\theta^*}(X_t) + K_t \frac{d\text{Var}_{\theta^*}(X_t)}{dt} + \dot{Y}_t \mathbb{E}_{\theta^*}[X_t] + Y_t \frac{d\mathbb{E}_{\theta^*}[X_t]}{dt} + \dot{\chi}_t \\ &= \dot{K}_t \text{Var}_{\theta^*}(X_t) + \dot{Y}_t \mathbb{E}_{\theta^*}[X_t] + \dot{\chi}_t + \mathbb{E}_{\theta^*}[G_t(\alpha)], \end{aligned} \quad (3.13)$$

where \dot{K}_t , \dot{Y}_t and $\dot{\chi}_t$ represent the time derivatives of K_t , Y_t and χ_t respectively, and

$$G_t(\alpha) := \alpha_t^\top Q_t \alpha_t + \alpha_t^\top [2U_t(X_t - \mathbb{E}_{\theta^*}[X_t]) + O_t],$$

with the deterministic coefficients

$$Q_t = K_t \Sigma(\rho^*(t)), \quad U_t = K_t b^*(t), \quad O_t = Y_t b^*(t).$$

By square completion, we rewrite $G_t(\alpha)$ as

$$G_t(\alpha) = (\alpha_t - \hat{\alpha}_t(X_t, \mathbb{E}_{\theta^*}[X_t]))^\top Q_t (\alpha_t - \hat{\alpha}_t(X_t, \mathbb{E}_{\theta^*}[X_t])) - \zeta_t,$$

where for $t \in [0, T]$, $x, \bar{x} \in \mathbb{R}^2$,

$$\hat{\alpha}_t(x, \bar{x}) := -Q_t^{-1} U_t (x - \bar{x}) - \frac{1}{2} Q_t^{-1} O_t,$$

and

$$\zeta_t := U_t^\top Q_t^{-1} U_t \text{Var}_{\theta^*}(X_t) + \frac{1}{4} O_t^\top Q_t^{-1} O_t = K_t R(\theta^*(t)) \text{Var}_{\theta^*}(X_t) + \frac{Y_t^2}{4K_t} R(\theta^*(t)).$$

The expression in (3.13) is then rewritten as

$$\begin{aligned} \frac{d\mathbb{E}_{\theta^*}[V_t^{\alpha, \theta^*}]}{dt} &= (\dot{K}_t - K_t R(\theta^*(t))) \text{Var}_{\theta^*}(X_t) + \dot{Y}_t \mathbb{E}_{\theta^*}[X_t] + \dot{\chi}_t - \frac{Y_t^2}{4K_t} R(\theta^*(t)) \\ &\quad + K_t \mathbb{E}_{\theta^*}[(\alpha_t - \hat{a}_t(X_t, \mathbb{E}_{\theta^*}[X_t]))^\top \Sigma(\rho^*(t)) (\alpha_t - \hat{a}_t(X_t, \mathbb{E}_{\theta^*}[X_t]))]. \end{aligned} \quad (3.14)$$

Therefore, whenever

$$\begin{cases} \dot{K}_t - K_t R(\theta^*(t)) = 0, \\ \dot{Y}_t = 0, \\ \dot{\chi}_t - \frac{Y_t^2}{4K_t} R(\theta^*(t)) = 0, \end{cases} \quad (3.15)$$

holds for all $t \in [0, T]$, which yields, together with the terminal condition (3.12), the explicit forms:

$$K_t = -\lambda e^{\int_t^T R(\theta^*(s)) ds} < 0, \quad Y_t = 1, \quad \chi_t = \frac{1}{4\lambda} [e^{\int_t^T R(\theta^*(s)) ds} - 1], \quad (3.16)$$

we have

$$\frac{d\mathbb{E}_{\theta^*}[V_t^{\alpha, \theta^*}]}{dt} = K_t \mathbb{E}_{\theta^*}[(\alpha_t - \hat{a}_t(X_t, \mathbb{E}_{\theta^*}[X_t]))^\top \Sigma(\rho^*) (\alpha_t - \hat{a}_t(X_t, \mathbb{E}_{\theta^*}[X_t]))],$$

which is nonpositive for all $\alpha \in \mathcal{A}$, i.e., the process V_t^{α, θ^*} satisfies the condition (ii) of Lemma 3.1. Moreover, notice that in this case,

$$V_0^{\alpha, \theta^*} = v_0(x_0, x_0) = x_0 + \frac{1}{4\lambda} [e^{\int_0^T R(\theta^*(t)) dt} - 1], \quad (3.17)$$

and

$$\hat{a}_t(x, \bar{x}) = -\Sigma(\rho^*(t))^{-1} b^*(t) (x - \bar{x} - \frac{1}{2\lambda} e^{\int_t^T R(\theta^*(s)) ds}). \quad (3.18)$$

Notice that in this step, we have not yet used the property that θ^* attains the infimum of the risk premium function. This will be used in the next step.

• *Step 2. condition (iii) of Lemma 3.1.*

Let us now prove that $V_0^{\alpha^*, \theta^*} \leq \mathbb{E}_{\theta^*}[V_T^{\alpha^*, \theta^*}]$, for all $\theta^* \in \mathcal{V}_{\Theta}$. A sufficient condition is the nondecreasing monotonicity of the function $t \mapsto \mathbb{E}_{\theta^*}[V_t^{\alpha^*, \theta^*}]$, by proving that $\frac{d\mathbb{E}_{\theta^*}[V_t^{\alpha^*, \theta^*}]}{dt}$ is nonnegative, for all $\theta^* \in \mathcal{V}_{\Theta}$. However, while this nondecreasing property is valid when there is no uncertainty on the drift, this does not hold true in the general uncertainty case as shown in Remark 3.5. We then proceed by computing directly the difference: $\mathbb{E}_{\theta^*}[V_T^{\alpha^*, \theta^*}] - V_0^{\alpha^*, \theta^*}$. Notice from (3.6), (2.4), that the dynamics of $\Lambda^*(X^*)$, with $\Lambda^*(x)$ defined in (3.7), under \mathbb{P}^{θ^*} , $\theta^* \in \mathcal{V}_{\Theta}$, is given by

$$d\Lambda^*(X_t^*) = -\Lambda^*(X_t^*) (b^*(t))^\top \Sigma(\rho^*(t))^{-1} [b_t dt + \sigma(\rho_t) dW_t^\theta],$$

with $\Lambda^*(x_0) = \frac{e^{\int_0^T R(\theta^*(t)) dt}}{2\lambda}$. By setting $N_t^* := \frac{2\lambda}{e^{\int_0^T R(\theta^*(t)) dt}} \Lambda^*(X_t^*)$, we deduce that

$$\begin{aligned} N_t^* &= \exp\left(-\int_0^t (b_s^\top \Sigma(\rho^*(t))^{-1} b^*(t) + \frac{1}{2} (b^*(t))^\top \Sigma(\rho^*(t))^{-1} \Sigma(\rho) \Sigma(\rho^*(t))^{-1} b^*(t)) ds \right. \\ &\quad \left. - \int_0^t (b^*(t))^\top \Sigma(\rho^*(t))^{-1} \sigma(\rho_s) dW_s^\theta\right), \quad 0 \leq t \leq T, \quad \mathbb{P}^{\theta^*} - a.s. \end{aligned}$$

$$X_t^* = x_0 + \frac{e^{\int_0^T R(\theta^*(t)) dt}}{2\lambda} (1 - N_t^*), \quad 0 \leq t \leq T, \quad \mathcal{P}^{\Theta} - q.s.,$$

and thus

$$\begin{cases} \mathbb{E}_\theta.[X_t^*] &= x_0 + \frac{e^{\int_0^T R(\theta^*(t))dt}}{2\lambda}(1 - \mathbb{E}_\theta.[N_t^*]), \\ \text{Var}_\theta.(X_t^*) &= \frac{e^{2\int_0^T R(\theta^*(t))dt}}{4\lambda^2}\text{Var}_\theta.(N_t^*). \end{cases} \quad (3.19)$$

By using the quadratic form (3.11) of v_t , together with the terminal condition (3.12), (3.17), and (3.19), we then obtain for all $\theta. \in \mathcal{V}_\Theta$:

$$\begin{aligned} \mathbb{E}_\theta.[V_T^{\alpha^*, \theta.}] - V_0^{\alpha^*, \theta.} &= \mathbb{E}_\theta.[v_T(X_T^*, \mathbb{E}_\theta.[X_T^*])] - v_0(x_0, x_0) \\ &= -\lambda \text{Var}_\theta.(X_T^*) + \mathbb{E}_\theta.[X_T^*] - x_0 - \frac{1}{4\lambda}(e^{\int_0^T R(\theta^*(t))dt} - 1) \\ &= -\frac{e^{2\int_0^T R(\theta^*(t))dt}}{4\lambda}\text{Var}_\theta.(N_T^*) + \frac{e^{\int_0^T R(\theta^*(t))dt}}{2\lambda}(1 - \mathbb{E}_\theta.[N_T^*]) - \frac{1}{4\lambda}(e^{\int_0^T R(\theta^*(t))dt} - 1) \\ &= \frac{e^{\int_0^T R(\theta^*(t))dt}}{4\lambda}\left(1 - e^{\int_0^T R(\theta^*(t))dt}\mathbb{E}_\theta.[|N_T^*|^2]\right) + \frac{1}{4\lambda}\left(e^{\int_0^T R(\theta^*(t))dt}\mathbb{E}_\theta.[N_T^*] - 1\right)^2 \\ &\geq \frac{e^{\int_0^T R(\theta^*(t))dt}}{4\lambda}\left(1 - e^{\int_0^T R(\theta^*(t))dt}\mathbb{E}_\theta.[|N_T^*|^2]\right) =: \frac{e^{\int_0^T R(\theta^*(t))dt}}{4\lambda}\Delta_T^*(\theta.). \end{aligned} \quad (3.20)$$

Noting that N^* is rewritten in terms of H introduced in Lemma 3.2 as

$$N_t^* = \exp\left(-\int_0^t (H_s(b_s, \rho^*(s)) + \frac{1}{2}H_s(b^*(s), \rho_s))ds - \int_0^t (b^*(s))^\top \Sigma(\rho^*(s))^{-1} \sigma(\rho_s) dW_s^\theta\right),$$

for $t \in [0, T]$, $\mathbb{P}^{\theta.}$ - a.s., and observing that $|(b^*(s))^\top \Sigma(\rho^*(s))^{-1} \sigma(\rho_s)|^2 = H_s(b^*(s), \rho_s)$, we see that

$$|N_t^*|^2 = \exp\left(-\int_0^t (2H_s(b_s, \rho^*(s)) - H(b^*(s), \rho_s))ds\right)M_t^*,$$

where

$$M_t^* := \exp\left(-2\int_0^t |(b^*(s))^\top \Sigma(\rho^*(s))^{-1} \sigma(\rho_s)|^2 ds - 2\int_0^t (b^*(s))^\top \Sigma(\rho^*(s))^{-1} \sigma(\rho_s) dW_s^\theta\right),$$

is an exponential Doléans-Dade local martingale under any $\mathbb{P}^{\theta.}$, $\theta. \in \mathcal{V}_\Theta$. Actually, the Novikov criterion is satisfied. Indeed,

$$\begin{aligned} &\mathbb{E}_\theta.\left[\exp\left(\frac{1}{2}\int_0^T |2(b^*(t))^\top \Sigma(\rho^*(t))^{-1} \sigma(\rho_t)|^2 dt\right)\right] = \mathbb{E}_\theta.\left[\exp\left(2\int_0^T H_t(b^*(t), \rho_t) dt\right)\right] \\ &= \mathbb{E}_\theta.\left[\exp\left(2\int_0^T \kappa(b^*(t), \rho^*(t))^\top \Sigma(\rho_t) \kappa(b^*(t), \rho^*(t)) dt\right)\right] \\ &= \mathbb{E}_\theta.\left[\exp\left(2\int_0^T \sum_{i=1}^d \kappa^i(b^*(t), \rho^*(t))^2 + 2\sum_{1 \leq i < j \leq d} \rho_{ij,t} \kappa^i(b^*(t), \rho^*(t)) \kappa^j(b^*(t), \rho^*(t)) dt\right)\right] \\ &\leq \mathbb{E}_\theta.\left[\exp\left(\int_0^T \sum_{i=1}^d \kappa^i(b^*(t), \rho^*(t))^2 + 2\sum_{1 \leq i < j \leq d} |\kappa^i(b^*(t), \rho^*(t))| |\kappa^j(b^*(t), \rho^*(t))|\right)\right] \\ &= \mathbb{E}_\theta.\left[\exp\left(2\int_0^T \left(\sum_{i=1}^d |\kappa^i(b^*(t), \rho^*(t))|\right)^2 dt\right)\right] < \infty, \end{aligned}$$

where the first inequality comes from the fact that the process $\rho_{ij,t}$, $1 \leq i < j \leq d$, is valued in $(-1, 1)$. Therefore, $(M_t^*)_{0 \leq t \leq T}$ is a martingale under any $\mathbb{P}^{\theta.}$, $\theta. \in \mathcal{V}_\Theta$. Consequently, we have

$$\begin{aligned} \Delta_T^*(\theta.) &= 1 - \mathbb{E}_\theta.\left[\exp\left(\int_0^T (R(\theta^*(s)) - 2H_s(b_s, \rho^*(s)) + H_s(b^*(s), \rho_s))ds\right)M_T^*\right] \\ &\geq 1 - \mathbb{E}_\theta.[M_T^*] = 1 - M_0^* = 0, \end{aligned}$$

where we used (3.10) in the above inequality. From (3.20), this proves condition (iii) of Lemma (3.1), and finally concludes the proof of Theorem 3.1. \square

Remark 3.4 The optimal strategy α^* given in (3.6) can be expressed in feedback form as

$$\alpha_t^* = \hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta^*}[X_t^*]), \quad 0 \leq t \leq T, \mathcal{P}^{\Theta} - q.s., \quad (3.21)$$

where $\hat{\alpha}_t$ is defined in (3.18). Indeed, denoting by $\hat{\alpha} \in \mathcal{A}$ the process defined by $\hat{\alpha}_t = \hat{\alpha}_t(\hat{X}_t, \mathbb{E}_{\theta^*}[\hat{X}_t])$, $0 \leq t \leq T$, $\mathcal{P}^{\Theta} - q.s.$, where \hat{X} is the wealth process associated to $\hat{\alpha}$, we see from (2.4) that \hat{X} satisfies the dynamics under \mathbb{P}^{θ^*} :

$$d\hat{X}_t = - \left[\hat{X}_t - \mathbb{E}_{\theta^*}[\hat{X}_t] - \frac{1}{2\lambda} e^{\int_t^T R(\theta^*(s))ds} \right] (b^*(t))^\top \Sigma(\rho^*(t))^{-1} [b^*(t)dt + \sigma(\rho^*(t))dW_t^{\theta^*}].$$

By taking expectation under \mathbb{P}^{θ^*} , we get: $d\mathbb{E}_{\theta^*}[\hat{X}_t] = \frac{1}{2\lambda} e^{\int_t^T R(\theta^*(s))ds} R(\theta^*)dt$, and thus

$$\begin{aligned} \mathbb{E}_{\theta^*}[\hat{X}_t] &= x_0 + \frac{e^{\int_0^T R(\theta^*(t))dt}}{2\lambda} [1 - e^{-\int_0^t R(\theta^*(s))ds}], \\ \hat{\alpha}_t &= \Lambda^*(\hat{X}_t) \Sigma(\rho^*(t))^{-1} b^*(t), \quad 0 \leq t \leq T, \mathcal{P}^{\Theta} - q.s. \end{aligned}$$

This implies that \hat{X} and X^* satisfy the same linear SDE under \mathbb{P}^{θ} , for any $\theta \in \mathcal{V}_{\Theta}$, and so $\hat{X}_t = X_t^*$, $0 \leq t \leq T$, \mathcal{P}^{Θ} -q.s. This proves that $\alpha^* = \hat{\alpha}$, equal to (3.21). \diamond

Remark 3.5 By similar derivation as in (3.14), and using (3.15), (3.21), we have that for all $\theta = (\theta_t)_t = (b_t, \rho_t)_t \in \mathcal{V}_{\Theta}$, $t \in [0, T]$,

$$\begin{aligned} \frac{d\mathbb{E}_{\theta}[V_t^{\alpha^*, \theta}]}{dt} &= K_t (R(\theta^*(t)) - R(\theta_t)) \text{Var}_{\theta}(X_t^*) + \frac{1}{4K_t} (R(\theta^*(t)) - R(\theta_t)) \\ &\quad + K_t \mathbb{E}_{\theta} \left[\left(\hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta^*}[X_t^*]) - \hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta}[X_t^*]) \right)^\top \Sigma(\rho_t) \right. \\ &\quad \left. \left(\hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta^*}[X_t^*]) - \hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta}[X_t^*]) \right) \right] \\ &\geq K_t \mathbb{E}_{\theta} \left[\left(\hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta^*}[X_t^*]) - \hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta}[X_t^*]) \right)^\top \Sigma(\rho_t) \right. \\ &\quad \left. \left(\hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta^*}[X_t^*]) - \hat{\alpha}_t(X_t^*, \mathbb{E}_{\theta}[X_t^*]) \right) \right] \end{aligned} \quad (3.22)$$

by definition of $\theta^* \in \arg \min_{\theta \in \Theta} R(\theta)$, and as $K_t < 0$. In the case when there is no uncertainty on the drift, i.e., for any $\theta = (b, \rho) \in \mathcal{V}_{\Theta}$, b is a deterministic function equal to $b^o(t)$, $t \in [0, T]$, the dynamics of X^* under any \mathbb{P}^{θ} , $\theta \in \mathcal{V}_{\Theta}$, is given by

$$dX_t^* = \left[x_0 + \frac{e^{\int_0^T R(\theta^*(t))dt}}{2\lambda} - X_t^* \right] (b^o(t))^\top \Sigma(\rho^*(t))^{-1} [b^o(t)dt + \sigma(\rho_t)dW_t^{\theta}],$$

from which, we deduce by taking expectation under \mathbb{P}^{θ} :

$$\mathbb{E}_{\theta}[X_t^*] = x_0 + \frac{e^{\int_0^T R(\theta^*(t))dt}}{2\lambda} [1 - e^{-\int_0^t R(\theta^*(s))ds}].$$

This means that the expectation under \mathbb{P}^{θ} of the optimal wealth process X^* does not depend on $\theta \in \mathcal{V}_{\Theta}$, and the r.h.s. of (3.23) is then equal to zero. Therefore, the function $t \mapsto \mathbb{E}_{\theta}[V_t^{\alpha^*, \theta}]$ is nondecreasing for all $\theta \in \mathcal{V}_{\Theta}$, which implies in particular condition (iii) of Lemma 3.1.

However, in the case of drift uncertainty, we cannot conclude as above, and actually this nondecreasing property does not always hold true. Indeed, consider for example the case where there is only drift uncertainty

in a single asset model $d = 1$, with $\Theta(t) = \Theta = \{\theta \in [\underline{b}, \bar{b}]\}$, $0 \leq \underline{b} < \bar{b}$, and known variance Σ^o normalized to one. Notice that $R(\theta) = \theta^2$, and $\theta^* = \arg \min_{\theta \in \Theta} R(\theta) = \underline{b}$. For any constant process equal to $\theta \in \Theta$, we can compute explicitly from (3.19) the expectation and variance of X^* under \mathbb{P}^θ :

$$\begin{aligned}\mathbb{E}_\theta[X_t^*] &= \frac{1}{2\lambda} e^{R(\theta^*)T} [1 - e^{-\theta\theta^*t}], \\ \text{Var}_\theta(X_t^*) &= \frac{1}{4\lambda^2} e^{2R(\theta^*)T} [e^{(R(\theta^*)-2\theta\theta^*)t} - e^{-2\theta\theta^*t}].\end{aligned}$$

Plugging into (3.22), and using also the expression of K , \hat{a} in (3.16), (3.18), we have for all $\theta \in \Theta$, $t \in [0, T]$, after some straightforward rearrangement:

$$\begin{aligned}\frac{d\mathbb{E}_\theta[V_t^{\alpha^*, \theta}]}{dt} &= \frac{1}{2\lambda} e^{R(\theta^*)T} \left[ce^{-2ct} - e^{-R(\theta^*)t} (1 - e^{-ct}) \left(\frac{R(\theta^*)}{2} - \left(\frac{R(\theta^*)}{2} + c \right) e^{-ct} \right) \right] \\ &=: f(t, c),\end{aligned}$$

where we set $c = (\theta - \theta^*)\theta^* \geq 0$. Now, we easily see that for all $t \in [0, T]$, $f(t, c)$ converges to $-\frac{R(\theta^*)}{4\lambda} e^{R(\theta^*)(T-t)} < 0$, as c goes to infinity. Then, by continuity of f with respect to c , we deduce that for θ large enough (hence for c large enough), $\frac{d\mathbb{E}_\theta[V_t^{\alpha^*, \theta}]}{dt}$ is negative, which means that the function $t \mapsto \mathbb{E}_\theta[V_t^{\alpha^*, \theta}]$ is not nondecreasing for all $\theta \in \Theta$. Actually, we have proved in Theorem 3.1 the weaker condition (iii) of Lemma 3.1, that is, $V_0^{\alpha^*, \theta} \leq \mathbb{E}_\theta[V_T^{\alpha^*, \theta}]$, for all $\theta \in \mathcal{V}_\Theta$. \diamond

Remark 3.6 Assume that there is only correlation ambiguity, hence the known drift is a deterministic function denoted by $b^o(t)$, and $\hat{\rho}(t)$ is a point estimation of correlation lying in $\Gamma(t)$. Then the robust portfolio strategy α_t^* in (3.6) is less risky than standard mean-variance portfolio strategy denoted by $\alpha_t^{MV,*}$ (to distinguish with robust portfolio strategy) in Section 2.3 in the sense that, for each $t \in [0, T]$

$$\text{Var}_{\hat{\theta}}(\alpha_t^*) := \mathbb{E}_{\hat{\theta}}[(\alpha_t^*)^\top \Sigma(\hat{\rho}(t)) \alpha_t^*] \leq \mathbb{E}_{\hat{\theta}}[(\alpha_t^{MV,*})^\top \Sigma(\hat{\rho}(t)) \alpha_t^{MV,*}] =: \text{Var}_{\hat{\theta}}(\alpha_t^{MV,*}), \quad (3.24)$$

where $\mathbb{E}_{\hat{\theta}}$ is expectation under probability measure $\mathbb{P}^{\hat{\theta}}$ with $\hat{\theta} := (b^o(t), \hat{\rho}(t))$. Indeed, from the expression of α_t^* and $\alpha_t^{MV,*}$, we then have

$$\begin{aligned}\text{Var}_{\hat{\theta}}(\alpha_t^*) &= \frac{1}{4\lambda^2} (b^o(t))^\top \Sigma(\rho^*(t))^{-1} \Sigma(\hat{\rho}(t)) \Sigma(\rho^*(t))^{-1} b^o(t) e^{\int_t^T R(b^o(s), \rho^*(s)) ds} \\ &\quad e^{-\int_0^t (R(b^o(s), \rho^*(s)) ds - (b^o(s))^\top \Sigma(\rho^*(s))^{-1} \Sigma(\hat{\rho}(s))^{-1} \Sigma(\rho^*(s))^{-1} b^o(s)) ds}, \\ \text{Var}_{\hat{\theta}}(\alpha_t^{MV,*}) &= \frac{1}{4\lambda^2} R(b^o(t), \hat{\rho}(t)) e^{\int_t^T R(b^o(s), \hat{\rho}(s)) ds}.\end{aligned}$$

From Lemma 3.2 with $b^o(t) = b^*(t)$, together with $\rho^*(t) \in \arg \min_{\rho \in \Gamma(t)} R(b^o(t), \rho)$, we obtain for $t \in [0, T]$

$$(b^o(t))^\top \Sigma(\rho^*(t))^{-1} \Sigma(\hat{\rho}(t)) \Sigma(\rho^*(t))^{-1} b^o(t) \leq R(b^o(t), \rho^*(t)) \leq R(b^o(t), \hat{\rho}(t)),$$

which implies (3.24). This point was observed in Liu & Zeng (2017) under a single period setting, and extended here in a continuous-time setting. However, notice that this result does not always hold in the case of both drift and correlation ambiguity. \diamond

Conclusion of Part I We complete this part by highlighting the key mathematical result in this paper about the separation principle and (weak) martingale optimality principle for solving robust portfolio selection problem. A related methodology has been used in Ismail & Pham (2019), however only in the case of ambiguity set for the covariance matrix: it is pointed out in their Remark 4.3 that the employed method relying on a

verification theorem for McKean-Vlasov control problem (which itself is derived from an associated optimality principle in the Wasserstein space of probability measures) cannot tackle the ambiguity about mean return rate. In the paper Fouque et al. (2016), which considers a special setting with two-asset model and uncertain correlation, a similar separation principle is obtained, but not explicitly written in terms of a risk premium function, and it is not clear how their conditions (see Theorem 2.2 in Fouque et al. (2016)) can be expressed in a multi-asset case. An important contribution of our paper is to state this separation principle in a more general framework including uncertainty both on the mean return rates and on the correlations of multi-assets, and with ambiguity sets that may decrease over time, taking into account, for example, learning about the true parameter to reduce the estimation error. On the other hand, such a result holds not only for mean-variance problems, but for other popular classes of performance measures like utility criteria, and also in discrete-time setting. This is detailed and discussed in Appendix A, where we used martingale optimality principle as explained in Remark 3.3. It could be also applied to other time-inconsistent robust optimization, like V@R, or CV@R, in the future work. Finally, we point out that we are not able to tackle uncertainty on both marginal volatilities and correlation, as we would lose in this case the convexity of covariance matrix on parameter, which is required in the proof of the separation principle.

Part II: Applications

4 Applications and examples

We provide in this section several examples for the determination of the minimal risk premium arising from the separation principle in Theorem 3.1, and the implications for the optimal robust portfolio strategy and the portfolio diversification. We shall focus in this section on ambiguity sets $\Theta = \{\Theta(t), t \in [0, T]\}$ as in (2.2), i.e., in the ellipsoidal form

$$\Theta(t) = \{(b, \rho) \in \mathbb{R}^d \times \Gamma(t) : \|\sigma_{J_l}(\rho)^{-1}(b_{J_l} - \hat{b}_{J_l}(t))\|_2 \leq \delta_l(t), \quad l = 1, \dots, p\}. \quad (4.25)$$

Given $\Theta(t)$ as in (4.25), we denote by $\hat{\beta}_i(t) := \frac{\hat{b}_i(t)}{\sigma_i}$ the instantaneous Sharpe ratio of the i -th asset associated with estimated mean return $\hat{b}_i(t)$, and marginal volatility $\sigma_i > 0$, $i = 1, \dots, d$. In what follows, we assume that $\max_{1 \leq j \leq d} |\hat{\beta}_j(t)| \neq 0$ (otherwise $\hat{\beta}_j(t) = 0$ for each $1 \leq j \leq d$, i.e., $\hat{b}(t) = 0$, meaning that the optimal portfolio strategy is to never trade, i.e., $\alpha_t^* = 0$). We define the Sharpe ratio "proximity" between i -th asset and j -th asset, $1 \leq i \neq j \leq d$, by

$$\hat{\varrho}_{ij}(t) = \hat{\varrho}_{ji}(t) := \frac{\hat{\beta}_j(t)}{\hat{\beta}_i(t)} \mathbf{1}_{\{|\hat{\beta}_i(t)| > |\hat{\beta}_j(t)|\}} + \frac{\hat{\beta}_i(t)}{\hat{\beta}_j(t)} \mathbf{1}_{\{|\hat{\beta}_i(t)| \leq |\hat{\beta}_j(t)|\}} \in [-1, 1], \quad (4.26)$$

with the convention that $\hat{\varrho}_{ij}(t) = \hat{\varrho}_{ji}(t) = 0$ when $\hat{\beta}_i(t) = \hat{\beta}_j(t) = 0$.

We first provide the general explicit expression of the robust optimal strategy in the case of ellipsoidal ambiguity set.

Proposition 4.1 *Let $\Theta(t)$ be an ellipsoidal set as in (4.25) with $p = 1$, and assume that there exists $\rho^*(t) \in \arg \min_{\rho \in \Gamma(t)} \|\sigma(\rho)^{-1} \hat{b}(t)\|_2$. Then, an optimal portfolio strategy for (2.5) is given by, for $t \in [0, T]$,*

$$\begin{aligned} \alpha_t^* &= \left[x_0 + \frac{1}{2\lambda} e^{\int_0^t (\|\sigma(\rho^*(s))^{-1} \hat{b}(s)\|_2 - \delta(s))^2 \mathbf{1}_{\{\|\sigma(\rho^*(s))^{-1} \hat{b}(s)\|_2 > \delta(s)\}} ds} - X_t^* \right] \\ &\quad \left(1 - \frac{\delta(t)}{\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2} \right) \mathbf{1}_{\{\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2 > \delta(t)\}} \Sigma(\rho^*(t))^{-1} \hat{b}(t). \end{aligned} \quad (4.27)$$

Proof. See Section B.2 in Appendix.

Remark 4.1 (Financial interpretation: no risky investment) We have seen in the previous section that $\rho^*(t)$ exists when $\Gamma(t)$ is compact (in particular when it is a singleton, i.e., there is no ambiguity on correlation) or when $\Gamma(t) = \mathbb{C}_{>+}^d$, i.e., there is full ambiguity on correlation. From (4.27), we observe notably that whenever $\delta(t) \geq \|\sigma(\rho^*(t))^{-1}\hat{b}(t)\|_2$, $\alpha_t^* = 0$. In other words, when, at time t , the investor is poorly confident about her estimation on the expected rate of return $\hat{b}(t)$, or when the level of uncertainty aversion about the expected rate of return is high, then she does not make risky investment at all. \diamond

4.1 Full ambiguity correlation and anti-diversification

In this paragraph, we consider the case of full ambiguity on correlation, i.e., $\Gamma(t) = \mathbb{C}_{>+}^d$, and investigate the impact on optimal robust portfolio strategy.

Theorem 4.1 (Full ambiguity correlation) *Let $\Theta(t)$ be an ellipsoidal set as in (4.25), with $\Gamma(t) = \mathbb{C}_{>+}^d$ for all $t \in [0, T]$, and assume that for each $1 \leq l \leq p$, there exists a function $m_l(t) \in J_l$ s.t. $|\hat{\beta}_{m_l(t)}(t)| > \max_{j \in J_l, j \neq m_l(t)} |\hat{\beta}_j(t)|$. Assume further that there exists a function $k(t) \in \{1, \dots, p\}$ s.t. $|\hat{\beta}_{m_{k(t)}}(t)| - \delta_{k(t)}(t) > \max_{1 \leq l \leq p, l \neq k(t)} (|\hat{\beta}_{m_l(t)}(t)| - \delta_l(t))$. Then an optimal portfolio strategy for the robust mean-variance problem (2.5) is explicitly given by, for $t \in [0, T]$,*

$$\alpha_t^* = \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T (|\hat{\beta}_{m_{k(s)}}(s)| - \delta_{k(s)}(s))^2 1_{\{|\hat{\beta}_{m_{k(s)}}(s)| > \delta_{k(s)}(s)\}} ds} - X_t^* \right] \left(1 - \frac{\delta_{k(t)}(t)}{|\hat{\beta}_{m_{k(t)}}(t)|} \right) 1_{\{|\hat{\beta}_{m_{k(t)}}(t)| > \delta_{k(t)}(t)\}} (0, \dots, 0, \frac{\hat{b}_{m_{k(t)}}(t)}{\sigma_{m_{k(t)}}^2}, 0, \dots, 0)^\top.$$

Proof. See Section B.3 in Appendix.

Remark 4.2 Notice that in the particular case when $p = 1$, the conditions of Theorem 4.1 simply assume that there exists $m(t) \in \{1, \dots, d\}$ s.t. $|\hat{\beta}_{m(t)}(t)| > \max_{1 \leq j \leq d, j \neq m(t)} |\hat{\beta}_j(t)|$. Then an optimal portfolio strategy for the robust mean-variance problem (2.5) is explicitly given by, for $t \in [0, T]$,

$$\alpha_t^* = \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T (|\hat{\beta}_{m(s)}(s)| - \delta(s))^2 1_{\{|\hat{\beta}_{m(s)}(s)| > \delta(s)\}} ds} - X_t^* \right] \left(1 - \frac{\delta(t)}{|\hat{\beta}_{m(t)}(t)|} \right) 1_{\{|\hat{\beta}_{m(t)}(t)| > \delta(t)\}} (0, \dots, 0, \frac{\hat{b}_{m(t)}(t)}{\sigma_{m(t)}^2}, 0, \dots, 0)^\top.$$

\diamond

Remark 4.3 (Financial interpretation: anti-diversification) Observe that both reference Sharpe ratio and drift ambiguity level play an important role in the portfolio choice of an investor. If this investor is poorly confident on the drift estimate, i.e., whenever all $\delta_l(t)$, $1 \leq l \leq p$ are large enough, then she does not make risky investments at all, i.e., $\alpha_t^* = 0$. When the reference Sharpe ratio of an asset is large enough to offset the effect of its own drift ambiguity but not others, she would only invest in this asset, namely the one with the highest lower Sharpe ratio (the reference Sharpe ratio (absolute value) of i -th asset minus its drift estimation error), which might change from one subclass to another subclass with time.

This anti-diversification result under full ambiguity about correlation has been also observed in Liu & Zeng (2017) for a single-period mean-variance problem without drift uncertainty, and is extended here in a continuous-time framework. Moreover, under this general setting, the single risky asset that the investor trades may change over time because of the time varying ambiguity set.

Compared to Remark 4.1, we see that expected return rates ambiguity and correlation ambiguity have different effects on portfolio selection. When expected return rates ambiguity is large, no risky investment is made, while when correlation ambiguity is large, one and only one risky asset is traded. \diamond

4.2 Partial diversification

• Two-asset model: $d = 2$

We provide a complete picture of the optimal robust portfolio strategy in a two-asset model with ambiguous drift and correlation.

Theorem 4.2 (Ambiguous drift and correlation in the two-asset case) *Let $\Theta(t)$*

= $\{(b, \rho) \in \mathbb{R}^2 \times [\underline{\rho}(t), \bar{\rho}(t)] : \|\sigma(\rho)^{-1}(b - \hat{b}(t))\|_2 \leq \delta(t)\}$, with $-1 < \underline{\rho}(t) \leq \bar{\rho}(t) < 1$, $t \in [0, T]$, and assume at each time $t \in [0, T]$, $\max(|\hat{\beta}_1(t)|, |\hat{\beta}_2(t)|) \neq 0$.

Then, an optimal portfolio strategy is given by

$$\alpha_t^* = \begin{cases} \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T R(\theta^*(s)) ds} - X_t^* \right] \left(1 - \frac{\delta(t)}{\max(|\hat{\beta}_1(t)|, |\hat{\beta}_2(t)|)} \right) \\ \quad \mathbf{1}_{\{\max(|\hat{\beta}_1(t)|, |\hat{\beta}_2(t)|) > \delta(t)\}} \begin{pmatrix} \frac{\hat{b}_1(t)}{|\sigma_1|^2} \mathbf{1}_{\{|\hat{\beta}_1(t)| > |\hat{\beta}_2(t)|\}} \\ \frac{\hat{b}_2(t)}{|\sigma_2|^2} \mathbf{1}_{\{|\hat{\beta}_2(t)| > |\hat{\beta}_1(t)|\}} \end{pmatrix}, & \text{if } \hat{\varrho}_{12}(t) \in [\underline{\rho}(t), \bar{\rho}(t)] \\ \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T R(\theta^*(s)) ds} - X_t^* \right] \left(1 - \frac{\delta(t)}{\|\sigma(\bar{\rho}(t))^{-1} \hat{b}(t)\|_2} \right) \\ \quad \mathbf{1}_{\{\|\sigma(\bar{\rho}(t))^{-1} \hat{b}(t)\|_2 > \delta(t)\}} \Sigma(\bar{\rho}(t))^{-1} \hat{b}(t), & \text{if } \bar{\rho}(t) < \varrho_{12}(t) \\ \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T R(\theta^*(s)) ds} - X_t^* \right] \left(1 - \frac{\delta(t)}{\|\sigma(\underline{\rho}(t))^{-1} \hat{b}(t)\|_2} \right) \\ \quad \mathbf{1}_{\{\|\sigma(\underline{\rho}(t))^{-1} \hat{b}(t)\|_2 > \delta(t)\}} \Sigma(\underline{\rho}(t))^{-1} \hat{b}(t), & \text{if } \underline{\rho}(t) > \varrho_{12}(t) \end{cases}$$

where for each $s \in [0, T]$

$$\begin{aligned} & R(\theta^*(s)) \\ = & (\max(|\hat{\beta}_1(s)|, |\hat{\beta}_2(s)|) - \delta(s))^2 \mathbf{1}_{\{\max(|\hat{\beta}_1(s)|, |\hat{\beta}_2(s)|) > \delta(s)\}} \mathbf{1}_{\{\hat{\varrho}_{12}(s) \in [\underline{\rho}(s), \bar{\rho}(s)]\}} \\ & + (\|\sigma(\bar{\rho}(s))^{-1} \hat{b}(s)\|_2 - \delta(s))^2 \mathbf{1}_{\{\|\sigma(\bar{\rho}(s))^{-1} \hat{b}(s)\|_2 > \delta(s)\}} \mathbf{1}_{\{\bar{\rho}(s) < \hat{\varrho}_{12}(s)\}} \\ & + (\|\sigma(\underline{\rho}(s))^{-1} \hat{b}(s)\|_2 - \delta(s))^2 \mathbf{1}_{\{\|\sigma(\underline{\rho}(s))^{-1} \hat{b}(s)\|_2 > \delta(s)\}} \mathbf{1}_{\{\underline{\rho}(s) > \hat{\varrho}_{12}(s)\}}. \end{aligned} \tag{4.28}$$

Proof See Section B.4 in Appendix.

Remark 4.4 When there is only ambiguity on correlations and correlation ambiguity set does not vary with time, we retrieve the results obtained in Ismail & Pham (2019) for the correlation ambiguity between two assets (see their Theorem 4.2). Our Theorem includes in addition the case when there is uncertainty on the expected rate of return and the ambiguity sets vary over time. Also, with time varying ambiguity sets, we see that the investor may switch between well-diversification and under-diversification over time, moreover, when under-diversification occurs, she may only invest in the first asset or the second asset; when well-diversification occurs, she may switch between directional trading and spread trading. This feature reflects dynamic changes in real markets. \diamond

Remark 4.5 (Financial interpretation) At a given time $t \in [0, T]$, we have three possible cases depending on the relation between the bounds of correlations, more precisely, the value of variance risk ratio on the bounds of correlations, and Sharpe ratio proximity. Notice that $R(\hat{b}(t), \rho)$ is convex, the first order derivative $\frac{\partial R(\hat{b}(t), \rho)}{\partial \rho} = -\sigma_1 \sigma_2 \hat{\kappa}_t^1(\rho) \hat{\kappa}_t^2(\rho)$ is increasing, hence $\rho \in [\underline{\rho}(t), \bar{\rho}(t)] \mapsto \hat{\kappa}_t^1(\rho) \hat{\kappa}_t^2(\rho)$ decreasing. Moreover, we have $\lim_{\rho \rightarrow 1} \hat{\kappa}_t^1(\rho) \hat{\kappa}_t^2(\rho) < 0$, $\lim_{\rho \rightarrow -1} \hat{\kappa}_t^1(\rho) \hat{\kappa}_t^2(\rho) > 0$ and $\hat{\kappa}_t^1(\hat{\varrho}_{12}(t)) \hat{\kappa}_t^2(\hat{\varrho}_{12}(t)) = 0$.

In the first case when correlation ambiguity interval includes $\hat{\varrho}_{12}(t)$, anti-diversification phenomenon occurs, i.e., one and only one asset is invested. Indeed, note that when the range of correlation ambiguity

interval is larger than $\max\{1 - \hat{\varrho}_{12}(t), 1 + \hat{\varrho}_{12}(t)\}$, Sharpe ratio proximity $\hat{\varrho}_{12}(t) \in [\underline{\rho}(t), \bar{\rho}(t)]$, or in other words, the correlation can be either larger or smaller than the Sharpe ratio proximity. In this case, the optimal strategy under the worst-case scenario is to invest in the asset with highest Sharpe ratio.

In the second case when $\bar{\rho}(t) < \hat{\varrho}_{12}(t)$, meaning that the correlation taking value in $[\underline{\rho}(t), \bar{\rho}(t)]$ is small compared to the Sharpe ratio proximity, then it is optimal to invest in both assets with a directional trading, that is, buying or selling simultaneously. And the worst-case correlation refers to the highest correlation $\bar{\rho}(t)$ where the diversification effect is minimal. This case corresponds to $\hat{\kappa}_t^1(\bar{\rho}(t))\hat{\kappa}_t^2(\bar{\rho}(t)) > 0$. Together with the monotonicity of $\hat{\kappa}_t^1(\rho)\hat{\kappa}_t^2(\rho)$, we have $\hat{\kappa}_t^1(\rho)\hat{\kappa}_t^2(\rho) > 0$ for any $\rho \in [\underline{\rho}(t), \bar{\rho}(t)]$, meaning that it is optimal to take directional trading, and the worst-scenario correlation is upper bound of correlation under which the diversification effect is minimal.

In the third case when $\underline{\rho}(t) > \hat{\varrho}_{12}(t)$, meaning that the correlation taking value in $[\underline{\rho}(t), \bar{\rho}(t)]$ is large compared to the Sharpe ratio proximity, then it is optimal to invest in both assets with a spread trading, that is, buying one and selling another. And the worst-case correlation corresponds to the lowest correlation $\underline{\rho}(t)$ where the profit from the spread trading is minimal. This case corresponds to $\hat{\kappa}_t^1(\underline{\rho}(t))\hat{\kappa}_t^2(\underline{\rho}(t)) < 0$. By analogy with the second case, we have $\hat{\kappa}_t^1(\rho)\hat{\kappa}_t^2(\rho) < 0$ for any $\rho \in [\underline{\rho}(t), \bar{\rho}(t)]$, meaning that it is optimal to invest with spread trading with lower bound of correlation.

This diversification result with only correlation uncertainty has been also observed in the literature Fouque et al. (2016) for a continuous-time expected utility problem, Liu & Zeng (2017) for a single-period mean-variance problem and Ismail & Pham (2019) for a continuous-time mean-variance problem, and is extended here in a continuous time framework with time varying ambiguity set for both drift and correlation uncertainty. One interesting additional finding in continuous time is that at some time we can be in the first case where under-diversification occurs, and next at a future time, in case 2 or 3 with directional or spread trading due to a change in the relation between the bounds of correlations and Sharpe ratio proximity. \diamond

• **Three-asset model:** $d = 3$

We finally provide an explicit description of the optimal robust strategy in a three-asset model under drift uncertainty and ambiguous correlation where ambiguity set does not vary with time for simplicity. We introduce the so-called variance risk ratio $\hat{\kappa}(\rho)$,

$$\Sigma(\rho)^{-1}\hat{b} =: \hat{\kappa}(\rho) = (\hat{\kappa}^1(\rho), \hat{\kappa}^2(\rho), \hat{\kappa}^3(\rho))^\top,$$

which represents (up to a scalar term) the vector of allocation in the assets when the drift is \hat{b} and the correlation is ρ .

Theorem 4.3 *Let $\Theta = \{(b, \rho) \in \mathbb{R}^3 \times \Gamma : \|\sigma(\rho)^{-1}(b - \hat{b})\|_2 \leq \delta\}$, with $\Gamma = [\underline{\rho}_{12}, \bar{\rho}_{12}] \times [\underline{\rho}_{13}, \bar{\rho}_{13}] \times [\underline{\rho}_{23}, \bar{\rho}_{23}] \subset \mathbb{C}_{>+}^3$, and assume w.l.o.g. that $|\hat{\beta}_1| \geq |\hat{\beta}_2| \geq |\hat{\beta}_3|$ and $\hat{\beta}_1 \neq 0$. Then, we have the following possible exclusive cases:*

1. (Anti-diversification) *If $\hat{\varrho}_{12} \in [\underline{\rho}_{12}, \bar{\rho}_{12}]$, and $\hat{\varrho}_{13} \in [\underline{\rho}_{13}, \bar{\rho}_{13}]$, then an optimal portfolio strategy is explicitly given by*

$$\alpha_t^* = \left[x_0 + \frac{1}{2\lambda} e^{(|\hat{\beta}_1| - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{|\hat{\beta}_1|} \right) \mathbf{1}_{\{|\hat{\beta}_1| > \delta\}} \begin{pmatrix} \frac{\hat{b}_1}{\sigma_1^2} \\ 0 \\ 0 \end{pmatrix}, \quad 0 \leq t \leq T, \quad \mathcal{P}^\Theta - q.s..$$

2. (Under-diversification: no investment in the third asset)

(i) If $\bar{\rho}_{12} < \hat{\varrho}_{12}$, and $\hat{\kappa}^3(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})\hat{\kappa}^3(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is

$$\begin{aligned} \begin{pmatrix} \alpha_t^{1,*} \\ \alpha_t^{2,*} \\ \alpha_t^3 \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 > \delta\}} \Sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3} \\ \alpha_t^{3,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 > \delta$, then $\alpha_t^{1,*} \alpha_t^{2,*} > 0$.

(ii) If $\underline{\rho}_{12} > \hat{\varrho}_{12}$, and $\hat{\kappa}^3(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})\hat{\kappa}^3(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is

$$\begin{aligned} \begin{pmatrix} \alpha_t^{1,*} \\ \alpha_t^{2,*} \\ \alpha_t^3 \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-3}(\underline{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-3}(\underline{\rho}_{12})^{-1}\hat{b}_{-3}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-3}(\underline{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 > \delta\}} \Sigma_{-3}(\bar{\rho}_{12})^{-1}\hat{b}_{-3} \\ \alpha_t^{3,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-3}(\underline{\rho}_{12})^{-1}\hat{b}_{-3}\|_2 > \delta$, then $\alpha_t^{1,*} \alpha_t^{2,*} < 0$.

3. (Under-diversification: no investment in the second asset)

(i) If $\bar{\rho}_{13} < \hat{\varrho}_{13}$, and $\hat{\kappa}^2(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})\hat{\kappa}^2(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is

$$\begin{aligned} \begin{pmatrix} \alpha_t^{1,*} \\ \alpha_t^{2,*} \\ \alpha_t^{3,*} \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-2}(\bar{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-2}(\bar{\rho}_{13})^{-1}\hat{b}_{-2}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-2}(\bar{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 > \delta\}} \Sigma_{-2}(\bar{\rho}_{13})^{-1}\hat{b}_{-2} \\ \alpha_t^{2,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-2}(\bar{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 > \delta$, then $\alpha_t^{1,*} \alpha_t^{3,*} > 0$.

(ii) If $\underline{\rho}_{13} > \hat{\varrho}_{13}$, and $\hat{\kappa}^2(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})\hat{\kappa}^2(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is given by

$$\begin{aligned} \begin{pmatrix} \alpha_t^{1,*} \\ \alpha_t^{2,*} \\ \alpha_t^{3,*} \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-2}(\underline{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-2}(\underline{\rho}_{13})^{-1}\hat{b}_{-2}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-2}(\underline{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 > \delta\}} \Sigma_{-2}(\underline{\rho}_{13})^{-1}\hat{b}_{-2} \\ \alpha_t^{2,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-2}(\underline{\rho}_{13})^{-1}\hat{b}_{-2}\|_2 > \delta$, then $\alpha_t^{1,*} \alpha_t^{3,*} < 0$.

4. (Under-diversification: no investment in the first asset)

(i) If $\bar{\rho}_{23} < \hat{\varrho}_{23}$, and $\hat{\kappa}^1(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})\hat{\kappa}^1(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is

$$\begin{aligned} \begin{pmatrix} \alpha_t^{1,*} \\ \alpha_t^{2,*} \\ \alpha_t^{3,*} \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-1}(\bar{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-1}(\bar{\rho}_{23})^{-1}\hat{b}_{-1}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-1}(\bar{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 > \delta\}} \Sigma_{-1}(\bar{\rho}_{23})^{-1}\hat{b}_{-1} \\ \alpha_t^{1,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-1}(\bar{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 > \delta$, then $\alpha_t^{2,*} \alpha_t^{3,*} > 0$.

(ii) If $\underline{\rho}_{23} > \hat{\varrho}_{23}$, and $\hat{\kappa}^1(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23})\hat{\kappa}^1(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) \leq 0$, then an optimal portfolio strategy is

$$\begin{aligned} \begin{pmatrix} \alpha_t^{2,*} \\ \alpha_t^{3,*} \end{pmatrix} &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma_{-1}(\underline{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma_{-1}(\underline{\rho}_{23})^{-1}\hat{b}_{-1}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma_{-1}(\underline{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 > \delta\}} \Sigma_{-1}(\underline{\rho}_{23})^{-1}\hat{b}_{-1} \\ \alpha_t^{1,*} &\equiv 0, \end{aligned}$$

and if $\|\sigma_{-1}(\underline{\rho}_{23})^{-1}\hat{b}_{-1}\|_2 > \delta$, then $\alpha_t^{2,*} \alpha_t^{3,*} < 0$.

5. (Well-diversification)

(i) If $\hat{\kappa}^1\hat{\kappa}^2(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23}) > 0$, and $\hat{\kappa}^1\hat{\kappa}^3(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23}) > 0$, then an optimal portfolio strategy is given by

$$\begin{aligned} \alpha_t^* &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2 > \delta\}} \Sigma(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}. \end{aligned}$$

(ii) If $\hat{\kappa}^1\hat{\kappa}^2(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23}) < 0$, and $\hat{\kappa}^1\hat{\kappa}^3(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23}) < 0$, then an optimal portfolio strategy is given by

$$\begin{aligned} \alpha_t^* &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}\|_2 > \delta\}} \Sigma(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23})^{-1}\hat{b}. \end{aligned}$$

(iii) If $\hat{\kappa}^1\hat{\kappa}^2(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) > 0$, and $\hat{\kappa}^1\hat{\kappa}^3(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) < 0$, then an optimal portfolio strategy is given by

$$\begin{aligned} \alpha_t^* &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2 > \delta\}} \Sigma(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}. \end{aligned}$$

(iv) If $\hat{\kappa}^1\hat{\kappa}^2(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) < 0$, and $\hat{\kappa}^1\hat{\kappa}^3(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) > 0$, then an optimal portfolio strategy is given by

$$\begin{aligned} \alpha_t^* &= \left[x_0 + \frac{1}{2\lambda} e^{(\|\sigma(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2 - \delta)^2 T} - X_t^* \right] \left(1 - \frac{\delta}{\|\sigma(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2} \right) \\ &\quad \mathbf{1}_{\{\|\sigma(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}\|_2 > \delta\}} \Sigma(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23})^{-1}\hat{b}. \end{aligned}$$

Proof. See Section B.5 in Appendix.

Remark 4.6 (Financial interpretation) In case **1** when $\hat{\varrho}_{12} \in [\underline{\rho}_{12}, \bar{\rho}_{12}]$, $\hat{\varrho}_{13} \in [\underline{\rho}_{13}, \bar{\rho}_{13}]$, it is optimal to invest only in the first asset, namely the one with the highest estimated Sharpe ratio, which is consistent with the anti-diversification result obtained in Theorem 4.1 (see also Remark 4.3). Case **1** happens when the range of $[\underline{\rho}_{1j}, \bar{\rho}_{1j}]$, $j = 2, 3$, is larger than $\max\{1 + \hat{\varrho}_{1j}, 1 - \hat{\varrho}_{1j}\}$.

In case **2**, corresponding to a large correlation ambiguity for the third asset, the investor does not invest in the third asset, but only in the first and second assets. Large correlation ambiguity for the third asset is quantified by the fact that the function $(\rho_{13}, \rho_{23}) \mapsto \hat{\kappa}(\bar{\rho}_{12}, \rho_{13}, \rho_{23})$ evaluated at the lower bounds $(\underline{\rho}_{13}, \underline{\rho}_{23})$ and the upper bounds $(\bar{\rho}_{13}, \bar{\rho}_{23})$ have opposite signs. Moreover, depending on whether the correlation of the first and second assets is small or large compared to the Sharpe ratio proximity ($\bar{\rho}_{12} < \hat{\varrho}_{12}$ or $\underline{\rho}_{12} > \hat{\varrho}_{12}$), the investment in the first and second assets follows a directional trading or a spread trading.

We have a similar under-diversification effect in cases **3** and **4**, and notice that it may happen that one does not invest in the first asset even though it has the highest reference Sharpe ratio. The result in case **4** is quite interesting and is *a priori* unexpected. Intuitively, an investor should always invest in the asset with the greatest absolute Sharpe ratio. For example, this is the case when anti-diversification occurs and also in cases **1**, **2**, **3**, **5**. However, the case **4** means that the asset with the greatest absolute Sharpe ratio (the first asset) may not be traded in the optimal portfolio while the one with the smallest absolute Sharpe ratio (the third asset) may be traded. The idea is that depending on the drift and correlation ambiguity levels, investing in the two other assets may achieve higher risk premium than investing in the first asset. Take case **4**(i) for example, in this case, the risk premium is $R(\hat{b}_{-1}, \bar{\rho}_{23}) = \hat{\beta}_{-1}^\top C(\bar{\rho}_{23})^{-1} \hat{\beta}_{-1}$ where $\hat{\beta}_{-1} := (\hat{\beta}_2, \hat{\beta}_3)^\top$, especially, $R(\hat{b}_{-1}, \bar{\rho}_{23}) = |\hat{\beta}_2|^2 + |\hat{\beta}_3|^2$ when $\bar{\rho}_{23} = 0$. It follows from (B.73) in appendix that $R(\hat{b}_{-1}, \bar{\rho}_{23}) > |\hat{\beta}_1|^2$. In Liu & Zeng (2017), the authors constructed a simple example where such scenario occurs in a single period model in the case where the second and third assets are independent, hence with no correlation ambiguity.

Finally, in case **5**, corresponding to a small correlation ambiguity, the investor has incentive to well-diversify her portfolio among the three assets. More precisely, Case **5** involves explicitly the signs of $\hat{\kappa}^1 \hat{\kappa}^2$ and $\hat{\kappa}^1 \hat{\kappa}^3$ at the correlation bounds. Assuming that these functions $\hat{\kappa}^1 \hat{\kappa}^2$ and $\hat{\kappa}^1 \hat{\kappa}^3$ do not vanish at some point $\rho \in [\underline{\rho}_{12}, \bar{\rho}_{12}] \times [\underline{\rho}_{13}, \bar{\rho}_{13}] \times [\underline{\rho}_{23}, \bar{\rho}_{23}]$, then by continuity, and provided that the range of these correlation bounds are small enough, we see that one should fall into one of the 4 subcases **5**.(i), (ii), (iii), (iv), and for which the worst-case correlation is obtained on the upper or lower correlation bounds. \diamond

5 Portfolio Sharpe ratio

In this section, we illustrate through two examples how drift estimation error $\delta(t)$ and correlation estimation interval denoted by $\epsilon(t)$ affect the portfolio Sharpe ratio of a strategy.

5.1 The impact of drift estimation error $\delta(t)$ on portfolio Sharpe ratio

We consider a market with one risky asset, and assume that the true dynamics of the stock price is given by the Black-Scholes model

$$dS_t = S_t(b^\circ dt + \sigma^\circ dW_t),$$

where the true drift $b^\circ > 0$ and the true volatility $\sigma^\circ > 0$ are constants, and W is Brownian motion under some probability measure \mathbb{P} . The portfolio Sharpe ratio of a strategy $\alpha \in \mathcal{A}$ over the finite horizon T is defined by

$$SR_T(\alpha) = \frac{\mathbb{E}[X_T^\alpha] - x_0}{\sqrt{\text{Var}(X_T^\alpha)}},$$

that is the excess of the expected return per unit of the standard deviation under the true probability measure \mathbb{P} .

• Let us first consider an investor who knows the true drift b° and true volatility σ° . In other words, she knows that the stock price is governed by a Black-Scholes model of parameter (b°, σ°) . Therefore, from our Section 2.3, the optimal mean-variance portfolio strategy of this investor with risk-aversion parameter $\lambda > 0$, and initial capital x_0 is given by

$$\alpha_t^* = \left[x_0 + \frac{e^{|\beta^\circ|^2 T}}{2\lambda} - X_t^* \right] \frac{b^\circ}{|\sigma^\circ|^2},$$

where $\beta^\circ = b^\circ/\sigma^\circ$, and X_t^* is the wealth process with feedback strategy α^* . By noting that the evolution of her wealth process X_t^* under \mathbb{P} is governed by

$$dX_t^* = \alpha_t^* b^\circ dt + \alpha_t^* \sigma^\circ dW_t,$$

we get, after straightforward calculation, that her terminal wealth is given by

$$X_T^* - x_0 = \frac{1}{2\lambda} \left[e^{|\beta^\circ|^2 T} - e^{-\frac{1}{2}|\beta^\circ|^2 T - \beta^\circ W_T} \right], \quad \mathbb{P} - a.s.$$

Therefore, its expectation and variance under \mathbb{P} are explicitly given by

$$\mathbb{E}[X_T^*] - x_0 = \frac{1}{2\lambda} [e^{|\beta^\circ|^2 T} - 1], \quad \text{Var}(X_T^*) = \frac{1}{4\lambda^2} [e^{|\beta^\circ|^2 T} - 1],$$

and thus the portfolio Sharpe ratio of the first investor following a portfolio strategy α^* is

$$SR_T^{(1)} := SR_T(\alpha^*) = \sqrt{e^{|\beta^\circ|^2 T} - 1}.$$

• Let us next consider a second investor with risk-aversion parameter λ , initial capital x_0 , who knows the true volatility but is uncertain about the drift: she believes that the drift lies in an ellipsoidal set around b° with constant δ . From Proposition 4.1, her robust optimal portfolio strategy denoted by $\tilde{\alpha}$ is given by

$$\tilde{\alpha}_t = \left[x_0 + \frac{1}{2\lambda} e^{(\beta^\circ - \delta)^2 T} - \tilde{X}_t \right] \left(1 - \frac{\delta}{\beta^\circ} \right) 1_{\{\beta^\circ > \delta\}} \frac{b^\circ}{|\sigma^\circ|^2},$$

where \tilde{X}_t is the wealth process associated to $\tilde{\alpha}$. By noting that the evolution of \tilde{X} under the true probability measure \mathbb{P} is

$$d\tilde{X}_t = \tilde{\alpha}_t b^\circ dt + \tilde{\alpha}_t \sigma^\circ dW_t,$$

we get its explicit expression under true probability measure \mathbb{P}

$$\tilde{X}_T - x_0 = \frac{1}{2\lambda} \left[e^{(\beta^\circ - \delta)^2 T} - e^{\frac{1}{2}(\delta^2 - |\beta^\circ|^2)T + (\delta - \beta^\circ)W_T} \right] 1_{\{\beta^\circ > \delta\}}, \quad \mathbb{P} - a.s.$$

It follows that the excess expected return and variance under \mathbb{P} are explicitly given by

$$\begin{aligned} \mathbb{E}[\tilde{X}_T] - x_0 &= \frac{1}{2\lambda} \left[e^{(\beta^\circ - \delta)^2 T} - e^{\delta(\delta - \beta^\circ)T} \right] 1_{\{\beta^\circ > \delta\}}, \\ \text{Var}(\tilde{X}_T) &= \frac{1}{4\lambda^2} \left[e^{(\beta^\circ - \delta)(\beta^\circ - 3\delta)T} - e^{2\delta(\delta - \beta^\circ)T} \right] 1_{\{\beta^\circ > \delta\}}. \end{aligned}$$

Therefore, the portfolio Sharpe ratio of the second investor following a portfolio strategy $\tilde{\alpha}$ is

$$SR_T^{(2)} := SR_T(\tilde{\alpha}) = \frac{e^{\beta^\circ(\beta^\circ - \delta)T} - 1}{\sqrt{e^{(\beta^\circ - \delta)^2 T} - 1}} 1_{\{\beta^\circ > \delta\}},$$

with the convention that $SR_T^{(2)} = 0$ when $\tilde{X}_T = x_0$.

• Finally, let us consider a third investor with risk-aversion parameter λ , initial capital x_0 , who knows the true volatility and has ambiguity about the drift which lies in ellipsoidal set around b° , but compared to the second investor, she learns information about drift by performing MLE or recursive point estimator with new coming observation, so that her estimation error is $\delta(t) = \frac{\delta}{\sqrt{1+t}}$, hence decreasing with time. It is consistent with T. Bielecki et al. (2017) in which the recursive point estimator of drift converges to the true value with convergence rate $\frac{1}{\sqrt{1+t}}$. It follows from Proposition 4.1 that an optimal robust portfolio strategy $\hat{\alpha}$ is

$$\hat{\alpha}_t = \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T (\beta^\circ - \frac{\delta}{\sqrt{1+s}})^2 1_{\{\beta^\circ > \frac{\delta}{\sqrt{1+s}}\}} ds} - \hat{X}_t \right] \left(1 - \frac{\delta}{\beta^\circ \sqrt{1+t}} \right) 1_{\{\beta^\circ > \frac{\delta}{\sqrt{1+t}}\}} \frac{b^\circ}{|\sigma^\circ|^2}.$$

Similarly as the previous two investors, we compute the terminal wealth associated to $\hat{\alpha}_t$ under the true probability measure \mathbb{P} as

$$\begin{aligned}\hat{X}_T - x_0 &= \frac{1}{2\lambda} \left[e^{\int_0^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}})^2 dt} \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+T}}\}} \right. \\ &\quad \left. - e^{\int_0^T \frac{1}{2} (\frac{\delta^2}{1+t} - |\beta^\circ|^2) dt - \int_0^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}}) \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+t}}\}} dW_t} \right] \\ &= \frac{1}{2\lambda} \left[e^{\int_{t^\circ}^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}})^2 dt} - e^{\int_{t^\circ}^T \frac{1}{2} (\frac{\delta^2}{1+t} - |\beta^\circ|^2) dt - (\beta^\circ - \frac{\delta}{\sqrt{1+t}}) dW_t} \right] \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+T}}\}},\end{aligned}$$

where we set $t^\circ = \max(\frac{\delta^2}{|\beta^\circ|^2} - 1, 0)$. Therefore, the expectation and variance of \hat{X}_T under \mathbb{P} are given by

$$\begin{aligned}\mathbb{E}[\hat{X}_T] - x_0 &= \frac{1}{2\lambda} \left[e^{\int_{t^\circ}^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}})^2 dt} - e^{\int_{t^\circ}^T (\frac{\delta^2}{1+t} - \frac{\delta\beta^\circ}{\sqrt{1+t}}) dt} \right] \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+T}}\}}, \\ \text{Var}(\hat{X}_T) &= \frac{1}{4\lambda^2} \left[e^{\int_{t^\circ}^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}})^2 dt} - e^{\int_{t^\circ}^T (\frac{2\delta^2}{1+t} - \frac{2\delta\beta^\circ}{\sqrt{1+t}}) dt} \right] \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+T}}\}}.\end{aligned}$$

It follows that the portfolio Sharpe ratio of the third investor following the strategy $\hat{\alpha}$ is

$$SR_T^{(3)} := SR_T(\hat{\alpha}) = \frac{e^{\int_{t^\circ}^T \beta^\circ (\beta^\circ - \frac{\delta}{\sqrt{1+t}}) dt} - 1}{\sqrt{e^{\int_{t^\circ}^T (\beta^\circ - \frac{\delta}{\sqrt{1+t}})^2 dt} - 1}} \mathbf{1}_{\{\beta^\circ > \frac{\delta}{\sqrt{1+T}}\}}$$

with the convention that $SR_T^{(3)} = 0$ when $\hat{X}_T = x_0$.

Let us illustrate numerically the effect of the estimation error on the robust optimal portfolio strategy. We take $b^\circ = 20\%$, $\sigma^\circ = 30\%$. Figure 1 shows the portfolio Sharpe ratio of investors when varying the estimation error δ at investment horizon $T = 2$, and varying terminal time T with estimation error $\delta = 0.5$. We see that the Sharpe ratio of the first investor who knows true parameters is always better than the two other investors who have ambiguity on drift. Moreover, we notice that the Sharpe ratio decreases as the estimation error increases, and that when investment horizon T is large, the Sharpe ratio of the third investor who learns information with time performs noticeably better than the one of the second investor.

5.2 The impact of correlation uncertainty level $\epsilon(t)$ on portfolio Sharpe ratio

We consider a market with two risky assets, and assume that the true dynamics of the stock prices $S = (S^1, S^2)$ is governed by

$$\begin{aligned}dS_t &= \text{diag}(S_t)(b^\circ dt + \sigma(\rho^\circ)dW_t) \\ &= \text{diag}(S_t) \left(\begin{pmatrix} b_1^\circ \\ b_2^\circ \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \sqrt{1 - |\rho^\circ|^2} & \sigma_1 \rho^\circ \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} \right),\end{aligned}$$

where the drift $b^\circ = (b_1^\circ, b_2^\circ)^\top$, $\sigma_1 > 0$, $\sigma_2 > 0$, and the true correlation $\rho^\circ \in (-1, 1)$ are known constants under some probability measure \mathbb{P} . Recall $\beta_i^\circ = \frac{b_i^\circ}{\sigma_i}$, $i = 1, 2$, we assume w.l.o.g. $\beta_1^\circ \geq \beta_2^\circ > 0$, and denote by $\varrho_{12}^\circ = \frac{\beta_2^\circ}{\beta_1^\circ}$ the Sharpe ratio proximity. Note that the true correlation ρ° is not necessarily equal to ϱ_{12}° .

• Let us consider the first investor who knows the true correlation ρ° . Therefore, from our Section 2.3, the optimal mean-variance portfolio strategy of this investor with risk-aversion parameter $\lambda > 0$, and initial capital x_0 is given by

$$\alpha_t^* = \left[x_0 + \frac{1}{2\lambda} e^{R^\circ T} - X_t^* \right] \Sigma(\rho^\circ)^{-1} b^\circ,$$

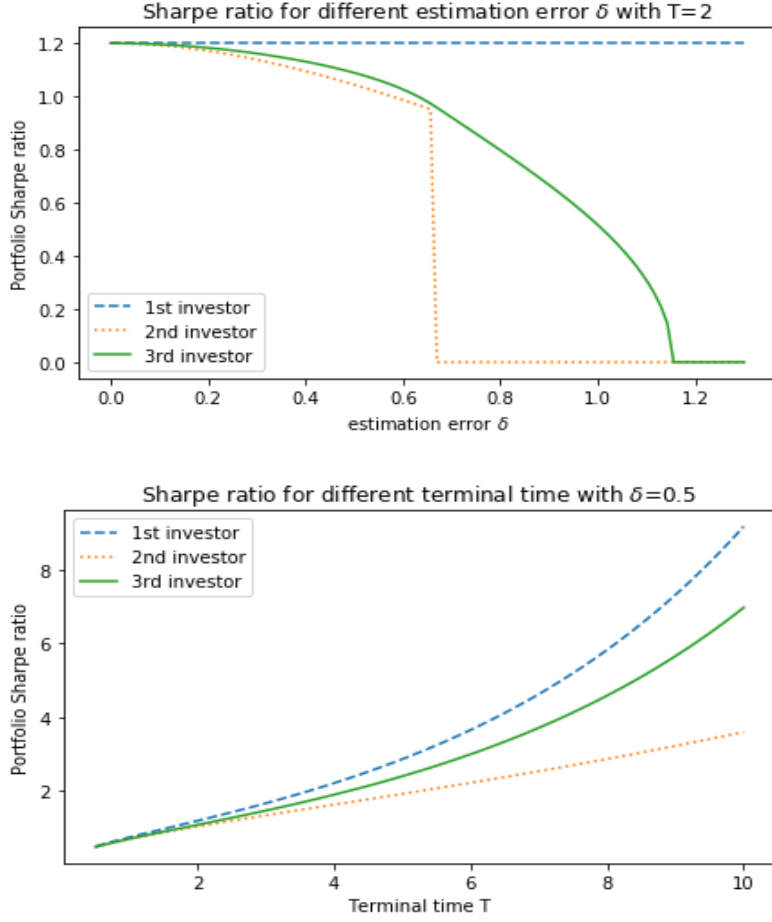


Figure 1: Portfolio Sharpe ratios for different estimation errors (above) and terminal horizons

where $R^o = (b^o)^\top \Sigma(\rho^o)^{-1} b^o$ and X_t^* is the wealth process with feedback strategy α^* . Since her wealth process X_t^* under \mathbb{P} is given by

$$dX_t^* = (\alpha_t^*)^\top (b^o dt + \sigma(\rho^o) dW_t),$$

we get, after straightforward calculations, that her terminal wealth is given by

$$X_T^* - x_0 = \frac{1}{2\lambda} [e^{R^o T} - e^{-\frac{1}{2} R^o T - (b^o)^\top \sigma(\rho^o)^{-1} W_T}], \quad \mathbb{P} - a.s.$$

Therefore, its expectation and variance under \mathbb{P} are explicitly given by

$$\mathbb{E}[X_T^*] - x_0 = \frac{1}{2\lambda} [e^{R^o T} - 1], \quad \text{Var}(X_T^*) = \frac{1}{4\lambda^2} [e^{R^o T} - 1],$$

hence the portfolio Sharpe ratio of this investor associated to α^* is

$$SR_T^{(1)} := SR_T(\alpha^*) = \sqrt{e^{R^o T} - 1}.$$

- Let us consider the second investor who knows the true expected rate of return b^o but is uncertain about correlation. She believes that the correlation lies in an interval $[\underline{\rho}, \bar{\rho}] = [\rho^o - \epsilon, \rho^o + \epsilon] \in (-1, 1)$, with ϵ a

positive constant. From Theorem 4.2, the robust optimal portfolio strategy denoted by $\tilde{\alpha}^{(2)}$ of the second investor is given by

$$\tilde{\alpha}_t^{(2)} = \left[x_0 + \frac{1}{2\lambda} e^{R(\rho^{(2),*})T} - \tilde{X}_t^{(2)} \right] \Sigma(\rho^{(2),*})^{-1} b^o,$$

where $R(\rho) = (b^o)^\top \Sigma(\rho)^{-1} b^o$, $\tilde{X}_t^{(2)}$ is wealth process associated to $\tilde{\alpha}^{(2)}$, and

$$\begin{aligned} \rho^{(2),*} &= \varrho_{12}^o \mathbf{1}_{\{\varrho_{12}^o \in [\underline{\rho}, \bar{\rho}]\}} + \bar{\rho} \mathbf{1}_{\{\bar{\rho} < \varrho_{12}^o\}} + \underline{\rho} \mathbf{1}_{\{\rho > \varrho_{12}^o\}}, \\ R(\rho^{(2),*}) &= |\beta_1^o|^2 \mathbf{1}_{\{\varrho_{12}^o \in [\underline{\rho}, \bar{\rho}]\}} + (b^o)^\top \Sigma(\bar{\rho})^{-1} b^o \mathbf{1}_{\{\bar{\rho} < \varrho_{12}^o\}} + (b^o)^\top \Sigma(\underline{\rho})^{-1} b^o \mathbf{1}_{\{\rho > \varrho_{12}^o\}}. \end{aligned} \quad (5.29)$$

By noting that the evolution of $\tilde{X}^{(2)}$ under probability measure \mathbb{P} is

$$d\tilde{X}_t^{(2)} = (\tilde{\alpha}_t^{(2)})^\top (b^o dt + \sigma(\rho^o) dW_t),$$

we get its explicit expression under \mathbb{P}

$$\tilde{X}_T^{(2)} - x_0 = \frac{1}{2\lambda} \left[e^{R(\rho^{(2),*})T} - e^{-(b^o)^\top \Sigma(\rho^{(2),*})^{-1} \sigma(\rho^o) W_T - \frac{1}{2} (b^o)^\top \Sigma(\rho^{(2),*})^{-1} \Sigma(\rho^o) \Sigma(\rho^{(2),*})^{-1} b^o T} \right],$$

and thus the expectation and variance of $\tilde{X}_T^{(2)}$ under \mathbb{P} are

$$\begin{aligned} \mathbb{E}[\tilde{X}_T^{(2)}] - x_0 &= \frac{1}{2\lambda} \left[e^{R(\rho^{(2),*})T} - 1 \right], \\ \text{Var}(\tilde{X}_T^{(2)}) &= \frac{1}{4\lambda^2} \left(e^{(b^o)^\top \Sigma(\rho^{(2),*})^{-1} \Sigma(\rho^o) \Sigma(\rho^{(2),*})^{-1} b^o T} - 1 \right), \end{aligned}$$

therefore the portfolio Sharpe ratio of the investor is given by

$$SR_T^{(2)} := SR_T(\tilde{\alpha}^{(2)}) = \frac{e^{R(\rho^{(2),*})T} - 1}{\sqrt{e^{(b^o)^\top \Sigma(\rho^{(2),*})^{-1} \Sigma(\rho^o) \Sigma(\rho^{(2),*})^{-1} b^o T} - 1}}.$$

Substituting $\underline{\rho} = \rho^o - \epsilon$ and $\bar{\rho} = \rho^o + \epsilon$ into (5.29), we get the explicit form of $\rho^{(2),*}$.

(i) If $\rho^o = \varrho_{12}^o$, then $\rho^{(2),*} = \varrho_{12}^o$.

(ii) If $\rho^o < \varrho_{12}^o$, then

$$\rho^{(2),*} = \begin{cases} \rho^o + \epsilon, & \text{if } \epsilon < \varrho_{12}^o - \rho^o, \\ \varrho_{12}^o, & \text{if } \epsilon \geq \varrho_{12}^o - \rho^o, \end{cases}$$

(iii) If $\rho^o > \varrho_{12}^o$, then

$$\rho^{(2),*} = \begin{cases} \rho^o - \epsilon, & \text{if } \epsilon < \rho^o - \varrho_{12}^o, \\ \varrho_{12}^o, & \text{if } \epsilon > \rho^o - \varrho_{12}^o, \end{cases}$$

• The third investor is more informed than the second investor, and believes that the correlation lies in an interval $[\underline{\rho}(t), \bar{\rho}(t)] = [\rho^o - \frac{\epsilon}{\sqrt{1+t}}, \rho^o + \frac{\epsilon}{\sqrt{1+t}}] \subset (-1, 1)$. From Theorem 4.2, the robust optimal portfolio strategy denoted by $\tilde{\alpha}^{(3)}$ of the third investor is given by

$$\tilde{\alpha}_t^{(3)} = \left[x_0 + \frac{1}{2\lambda} e^{\int_0^T R(\rho^{(3),*}(s)) ds} - \tilde{X}_t^{(3)} \right] \Sigma(\rho^{(3),*}(t))^{-1} b^o,$$

where $\tilde{X}_t^{(3)}$ is wealth process associated to $\tilde{\alpha}^{(3)}$, and

$$\begin{aligned}\rho^{(3),*}(t) &= \varrho_{12}^o 1_{\{\varrho_{12}^o \in [\underline{\rho}(t), \bar{\rho}(t)]\}} + \bar{\rho}(t) 1_{\{\bar{\rho}(t) < \varrho_{12}^o\}} + \underline{\rho}(t) 1_{\{\underline{\rho}(t) > \varrho_{12}^o\}}, \\ R(\rho^{(3),*}(t)) &= |\beta_1^o|^2 1_{\{\varrho_{12}^o \in [\underline{\rho}(t), \bar{\rho}(t)]\}} + (b^o)^\top \Sigma(\bar{\rho}(t))^{-1} b^o 1_{\{\bar{\rho}(t) < \varrho_{12}^o\}} + (b^o)^\top \Sigma(\underline{\rho}(t))^{-1} b^o 1_{\{\underline{\rho}(t) > \varrho_{12}^o\}}.\end{aligned}\tag{5.30}$$

The portfolio Sharpe ratio of the third investor is computed in the same way as the second investor, and is given by:

$$SR_T^{(3)} := SR_T(\tilde{\alpha}^3) = \frac{e^{\int_0^T R(\rho^{(3),*}(s)) ds} - 1}{\sqrt{e^{\int_0^T (b^o)^\top \Sigma(\rho^{(3),*}(s))^{-1} \Sigma(\rho^o) \Sigma(\rho^{(3),*}(s))^{-1} b^o ds} - 1}}.$$

Plugging $\underline{\rho}(t) = \rho^o - \frac{\epsilon}{\sqrt{1+t}}$, and $\bar{\rho}(t) = \rho^o + \frac{\epsilon}{\sqrt{1+t}}$ into (5.30), we get the explicit expression of $\rho^{(3),*}(t)$.

(i) If $\rho^o = \varrho_{12}^o$, then $\rho^{(3),*}(t) = \varrho_{12}^o$.

(ii) If $\rho^o < \varrho_{12}^o$, then

$$\rho^{(3),*}(t) = \begin{cases} \rho^o + \frac{\epsilon}{\sqrt{1+t}}, & \text{if } \epsilon < \varrho_{12}^o - \rho^o, \\ \varrho_{12}^o 1_{\{t \in [0, \frac{\epsilon^2}{|\varrho_{12}^o - \rho^o|^2} - 1]\}} + (\rho^o + \frac{\epsilon}{\sqrt{1+t}}) 1_{\{t \in [\frac{\epsilon^2}{|\varrho_{12}^o - \rho^o|^2} - 1, T]\}}, & \text{if } \frac{\epsilon}{\sqrt{1+T}} < \varrho_{12}^o - \rho^o < \epsilon, \\ \varrho_{12}^o, & \text{if } \frac{\epsilon}{\sqrt{1+T}} \geq \varrho_{12}^o - \rho^o. \end{cases}$$

(iii) If $\rho^o > \varrho_{12}^o$, then

$$\rho^{(3),*}(t) = \begin{cases} \rho^o - \frac{\epsilon}{\sqrt{1+t}}, & \text{if } \epsilon < \rho^o - \varrho_{12}^o, \\ \varrho_{12}^o 1_{\{t \in [0, \frac{\epsilon^2}{|\varrho_{12}^o - \rho^o|^2} - 1]\}} + (\rho^o - \frac{\epsilon}{\sqrt{1+t}}) 1_{\{t \in [\frac{\epsilon^2}{|\varrho_{12}^o - \rho^o|^2} - 1, T]\}}, & \text{if } \frac{\epsilon}{\sqrt{1+T}} < \rho^o - \varrho_{12}^o < \epsilon, \\ \varrho_{12}^o, & \text{if } \frac{\epsilon}{\sqrt{1+T}} > \rho^o - \varrho_{12}^o. \end{cases}$$

Let us illustrate numerically the effect of the correlation ambiguity on the robust optimal portfolio strategy. We fix investment horizon $T = 2$ and take $\beta_1^o = 1.5$, $\beta_2^o = 0.5$, hence giving Sharpe ratio proximity $\varrho_{12}^o = \frac{1}{3}$. We select $\rho^o = \varrho_{12}^o = \frac{1}{3}$, or $\rho^o = 0 < \varrho_{12}^o$, or $\rho^o = \frac{1}{2} > \varrho_{12}^o$. Figure 2 shows the effect of the correlation ambiguity level on the portfolio Sharpe ratio. We see that portfolio Sharpe ratio decreases with correlation ambiguity level except when true correlation equals Sharpe ratio proximity. In this case, when true correlation equals Sharpe ratio proximity under which under-diversification occurs, whatever the correlation level ϵ is, portfolio Sharpe ratio is a constant.

Conclusion of Part II. In this part, we have provided a complete picture for two-asset and three-asset setting. The same patterns can hold true for $d - 1$ -asset and d -asset settings with $d > 3$. More precisely, we can apply the following induction argument. Suppose that we have all the scenarios for $d - 1$ assets: anti-diversification (investment in one asset), under-diversification (investment in $2, \dots, d - 2$ assets), and well-diversification (investment in $d - 1$ assets), then we obtain a complete picture of d assets depending on whether d -th asset is invested or not, which corresponds to the evaluation of variance risk ratios $\hat{\kappa}^j(\rho)$, $\hat{\kappa}^d(\rho)$ on $\hat{\varrho}_{jd}$, $\underline{\rho}_{jd}$, $\bar{\rho}_{jd}$, $1 \leq j \leq d - 1$. Mathematically, we have to compute the minima of risk premium function (convex) on $\prod_{1 \leq i \neq j \leq d} [\underline{\rho}_{ij}, \bar{\rho}_{ij}] \subset \mathbb{C}_{>+}^d$. Since there are too many sub-cases, we do not provide all the details for $d > 3$.

Acknowledgement: We would like to thank the Associate Editor and the two referees for their careful reading of our paper and their valuable comments. We feel that, thanks to all the comments we received, the clarity and the scientific value of the paper have been greatly improved.

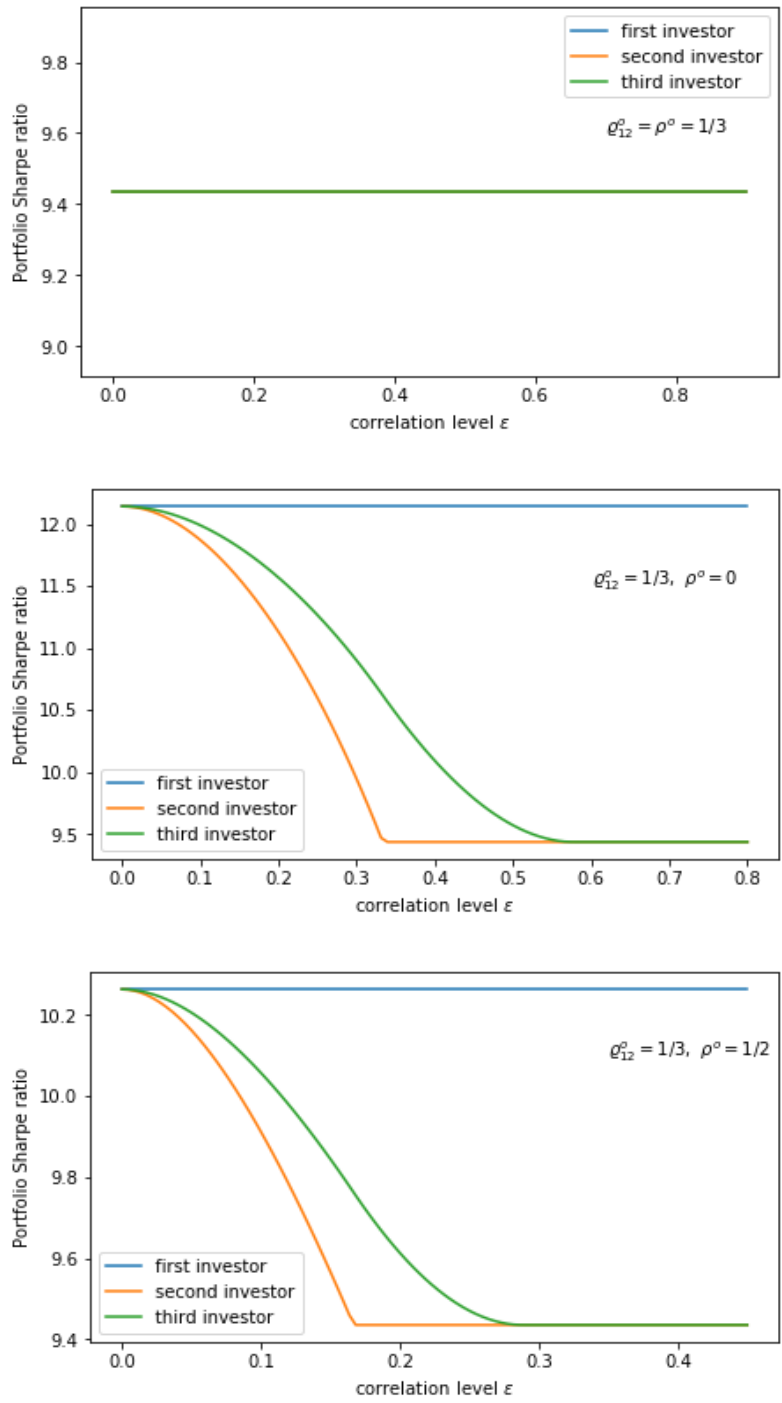


Figure 2: Portfolio Sharpe ratio for different correlation levels ϵ

Appendix

A Separation principle for robust portfolio selection

We show in this section that the separation principle is actually quite general, valid not only for mean-variance problem, but also for other classes in decision making problems, like popular expected utility criteria. It turns out that the proof of the separation principle for standard utility criteria is simpler than for mean-variance problems.

We consider a model uncertainty setting as in Section 2.1. The investor's preferences are represented by utility functions U defined from \mathbb{R} to \mathbb{R} , strictly concave and increasing. The wealth process X^α is defined as in (2.4), and the set \mathcal{A} of admissible controls consists of \mathbb{R}^d -valued \mathbb{F} -progressively measurable process α such that the process $U(X^\alpha)$ is uniformly integrable. The robust portfolio selection problem is then formulated as

$$V_0 = \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} \mathbb{E}_\theta[U(X_T^\alpha)]. \quad (\text{A.31})$$

This is a rather standard min-max stochastic control problem, which is associated by the dynamic programming principle to the Bellman-Isaacs partial differential equation:

$$\begin{cases} \frac{\partial v}{\partial t} + \sup_{a \in \mathbb{R}^d} \inf_{\theta \in \Theta(t)} H_t(v_x, v_{xx}, a, \theta) = 0, & \text{on } [0, T] \times \mathbb{R} \\ v(T, \cdot) = U, \end{cases} \quad (\text{A.32})$$

(assuming that $v(t, x)$ is smooth and strictly concave in x), where for $t \in [0, T]$, H_t is the Hamiltonian function defined on $\mathbb{R} \times (-\infty, 0) \times \mathbb{R}^d \times \Theta(t)$ by

$$H_t(p, M, a, \theta) = pa^\top b + \frac{1}{2}Ma^\top \Sigma(\rho)a, \quad p \in \mathbb{R}, M < 0, a \in \mathbb{R}^d, \theta = (b, \rho) \in \Theta(t).$$

In the no-uncertainty model case, i.e., $\Theta(t)$ is reduced to a singleton $\Theta(t) = \{\theta^o(t) = (b^o(t), \rho^o(t))\}$, $t \in [0, T]$, corresponding to a multi-dimensional Black-Scholes model with deterministic mean return vector $b^o(t)$, covariance matrix $\Sigma^o(t) = \Sigma(\rho^o(t))$, and deterministic risk premium $R^o(t) = R(\theta^o(t)) = (b^o(t))^\top (\Sigma^o(t))^{-1} b^o(t)$, the Bellman-Isaacs equation reduces to the Bellman equation arising in classical expected utility maximization, and called Black-Scholes-Merton Bellman PDE:

$$\begin{cases} \frac{\partial v^o}{\partial t} - \frac{R^o(t)}{2} \frac{(v_x^o)^2}{v_{xx}^o} = 0, & \text{on } [0, T] \times \mathbb{R} \\ v^o(T, \cdot) = U. \end{cases} \quad (\text{A.33})$$

Moreover, when there exists a smooth function $v^o(t, x)$ to (A.33), strictly concave in x , it is known by classical verification theorem (see e.g. Pham (2009)) that the optimal portfolio strategy is given by

$$\alpha_t^{o,*} = \mathcal{R}^o(t, X_t^*) (\Sigma^o(t))^{-1} b^o(t), \quad 0 \leq t \leq T,$$

where $\mathcal{R}^o(t, x) := -\frac{v_x^o(t, x)}{v_{xx}^o(t, x)}$ is the so-called risk tolerance function, X^* is the wealth process associated to α^* . For example, when U is a CRRA utility function, i.e., $U(x) = x^\gamma$, $x > 0$, with $0 < \gamma < 1$, we obtain the famous Merton solution: $v^o(t, x) = \exp\left(\frac{\int_t^T R^o(s) ds}{2} \frac{\gamma}{1-\gamma}\right) U(x)$, and $\mathcal{R}^o(t, x) = x/(1-\gamma)$. When U is of CARA type, i.e., $U(x) = -e^{-\eta x}$, with $\eta > 0$, we have $v^o(t, x) = \exp\left(\frac{\int_t^T R^o(s) ds}{2}\right) U(x)$, and $\mathcal{R}^o(t, x) = 1/\eta$.

In our general model uncertainty setting under $(\mathbf{H}\Theta)$, a key lemma is the following saddle point property:

Lemma A.1 Fix $t \in [0, T]$, and assume that there exists $\theta^*(t) = (b^*(t), \rho^*(t)) \in \arg \min_{\theta \in \Theta(t)} R(\theta)$. Let us denote by

$$a^*(t, p, M) = -\frac{p}{M}(\Sigma(\rho^*(t))^{-1}b^*(t)), \quad p \in \mathbb{R}, M < 0. \quad (\text{A.34})$$

Then, for $p \in \mathbb{R}$, $M < 0$, the pair $(a^*(t, p, M), \theta^*(t))$ is a saddle-point of $(a, \theta) \in \mathbb{R}^d \times \Theta(t) \mapsto H_t(p, M, a, \theta)$, i.e.

$$\begin{aligned} H_t(p, M, a^*(t, p, M), \theta) &\leq H_t(p, M, a^*(t, p, M), \theta^*(t)) = -\frac{1}{2} \frac{p^2}{M} R(\theta^*(t)) \\ &\leq H_t(p, M, a, \theta^*(t)), \quad \forall a \in \mathbb{R}^d, \theta \in \Theta(t). \end{aligned}$$

Proof. For any $n \in \mathbb{N} \setminus \{0\}$, let us introduce the compact set $\mathbb{C}_{n, >+}^d$ of all elements $\rho = (\rho_{ij})_{1 \leq i < j \leq d}$ in the open set $\mathbb{C}_{>+}^d$, such that $|\rho_{ij}| \leq 1 - \frac{1}{n}$, $1 \leq i < j \leq d$. Given the ambiguity set $\Theta(t)$, let us consider the sequence of sets:

$$\Theta_n(t) = \left\{ \theta = (b, \rho) \in \Theta(t) : \rho \in \mathbb{C}_{n, >+}^d \right\}, \quad n > 0,$$

and notice by **(HΘ)** that its closure $\overline{\Theta_n(t)}$ is a compact convex set of $\mathbb{R}^d \times \mathbb{C}_{>+}^d$. For fixed $(t, p, M) \in [0, T] \times \mathbb{R} \times (-\infty, 0)$, it is clear that the function $H_t(p, M, \cdot, \cdot)$ is concave in $a \in \mathbb{R}^d$, and linear (hence convex) in θ lying in the convex-compact set $\overline{\Theta_n(t)}$. By the min-max theorem (see e.g. Theorem 45.8 in Strasser (2011)), we then get the so-called Isaacs relation:

$$\sup_{a \in \mathbb{R}^d} \inf_{\theta \in \Theta_n(t)} H_t(p, M, a, \theta) = \inf_{\theta \in \overline{\Theta_n(t)}} \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta). \quad (\text{A.35})$$

By square completion, we can rewrite the function H_t as:

$$\begin{aligned} H_t(p, M, a, \theta) &= \frac{M}{2} (a - \bar{a}(p, M, \theta))^\top \Sigma(\rho) (a - \bar{a}(p, M, \theta)) - \frac{1}{2} \frac{p^2}{M} R(\theta), \\ \text{with } \bar{a}(p, M, \theta) &:= -\frac{p}{M} \Sigma^{-1}(\rho) b, \quad \theta = (b, \rho) \in \Theta(t), \end{aligned} \quad (\text{A.36})$$

from which we get

$$\sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta) = H_t(p, M, \bar{a}(p, M, \theta), \theta) = -\frac{1}{2} \frac{p^2}{M} R(\theta). \quad (\text{A.37})$$

Observe that for n large enough, $n \geq N^*$, the element $\theta^*(t) = (b^*(t), \rho^*(t)) \in \arg \min_{\theta \in \Theta(t)} R(\theta)$ lies in $\overline{\Theta_n(t)}$, and thus, using also the continuity of $R(\cdot)$ on $\mathbb{R}^d \times \mathbb{C}_{>+}^d$: $\inf_{\theta \in \Theta(t)} R(\theta) = \inf_{\theta \in \overline{\Theta_n(t)}} R(\theta) = R(\theta^*(t))$. We deduce with (A.37) that for $n \geq N^*$,

$$\begin{aligned} H_t^*(p, M) &:= \inf_{\theta \in \Theta(t)} \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta) = \inf_{\theta \in \overline{\Theta_n(t)}} \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta) \\ &= -\frac{1}{2} \frac{p^2}{M} R(\theta^*(t)). \end{aligned} \quad (\text{A.38})$$

On the other hand, we see that the continuous function $a \in \mathbb{R}^d \mapsto \underline{H}_t(p, M, a) := \inf_{\theta \in \overline{\Theta_n(t)}} H_t(p, M, a, \theta)$ is concave in a , and goes to $-\infty$, as $|a|$ goes to infinity (recall that $M < 0$). This implies that $\underline{H}_t(p, M, \cdot)$

attains its supremum at some point $\tilde{a}(t, p, M)$, and we then have

$$\begin{aligned}
\inf_{\theta \in \Theta_n(t)} H_t(p, M, \tilde{a}(t, p, M), \theta) &= \sup_{a \in \mathbb{R}^d} \inf_{\theta \in \Theta_n(t)} H_t(p, M, a, \theta) \\
&= \inf_{\theta \in \Theta_n(t)} \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta) = H_t^*(p, M) \\
&= \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta^*(t)) \geq H_t(p, M, a, \theta^*(t)), \quad \forall a \in \mathbb{R}^d,
\end{aligned}$$

where we used Isaacs condition (A.35) in the second equality, (A.38) in the third equality, and (A.37) for $\theta = \theta^*(t)$ in the fourth one. We then deduce

$$\begin{aligned}
H_t(p, M, \tilde{a}(t, p, M), \theta^*(t)) &\geq \inf_{\theta \in \Theta_n(t)} H_t(p, M, \tilde{a}(t, p, M), \theta) = H_t^*(p, M) \\
&\geq H_t(p, M, a, \theta^*(t)), \quad \forall a \in \mathbb{R}^d.
\end{aligned} \tag{A.39}$$

Similarly, we have for any $n \geq N^*$,

$$\begin{aligned}
\sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta^*(t)) &= \inf_{\theta \in \Theta_n(t)} \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta) \\
&= \sup_{a \in \mathbb{R}^d} \inf_{\theta \in \Theta_n(t)} H_t(p, M, a, \theta) = H_t^*(p, M) \\
&= \inf_{\theta \in \Theta_n(t)} H_t(p, M, \tilde{a}(t, p, M), \theta) \leq H_t(p, M, \tilde{a}(t, p, M), \theta), \quad \forall \theta \in \overline{\Theta_n(t)},
\end{aligned}$$

which implies that

$$\begin{aligned}
H_t(p, M, \tilde{a}(t, p, M), \theta^*(t)) &\leq \sup_{a \in \mathbb{R}^d} H_t(p, M, a, \theta^*(t)) = H_t^*(p, M) \\
&\leq H_t(p, M, \tilde{a}(t, p, M), \theta), \quad \forall \theta \in \Theta(t),
\end{aligned} \tag{A.40}$$

since any $\theta \in \Theta(t)$ lies in $\Theta_n(t)$ for n large enough. Relations (A.39)-(A.40) mean the saddle-point property of the pair $(\tilde{a}(t, p, M), \theta^*(t))$ for the function $(a, \theta) \in \mathbb{R}^d \times \Theta(t) \mapsto H_t(p, M, a, \theta)$, and also imply that $H_t(p, M, \tilde{a}_t(p, M), \theta^*(t)) = H_t^*(p, M)$. Recalling the expression (A.36) of H_t , this is written as

$$\begin{aligned}
&\frac{M}{2} (\tilde{a}(t, p, M) - \bar{a}(p, M, \theta^*(t)))^\top \Sigma(\rho^*(t)) (\tilde{a}(t, p, M) - \bar{a}(p, M, \theta^*(t))) + H_t^*(p, M) \\
&= H_t^*(p, M).
\end{aligned}$$

This proves that $\tilde{a}(t, p, M) = \bar{a}(p, M, \theta^*(t)) = a^*(t, p, M)$ as defined in (A.34), and ends the proof. \square

Proposition A.1 (Separation Principle for utility criteria) *Suppose that there exists a pair $\theta^* = (\theta^*(t))_t = (b^*, \rho^*) = (b^*(t), \rho^*(t))_t \in \Theta$ solution to $\arg \min_{\theta \in \Theta} R(\theta)$, i.e., $\theta^*(t) \in \arg \min_{\theta \in \Theta(t)} R(\theta)$, for all $t \in [0, T]$, and a smooth solution $v(t, x)$, strictly concave in x , to the Black-Scholes Merton Bellman PDE:*

$$\begin{cases} \frac{\partial v}{\partial t} - \frac{R^*(t)}{2} \frac{(v_x)^2}{v_{xx}} = 0, & \text{on } [0, T] \times \mathbb{R} \\ v(T, \cdot) = U, \end{cases} \tag{A.41}$$

where we set $R^*(t) := R(\theta^*(t))$, and satisfying the growth condition $v(t, x) \leq C(1 + |U(x)|)$. Then the robust utility maximization problem (A.31) admits an optimal portfolio strategy given by

$$\alpha_t^* = \mathcal{R}^*(t, X_t^*) (\Sigma^*(t))^{-1} b^*(t), \quad 0 \leq t \leq T, \quad \mathcal{P}^\Theta - q.s.,$$

where $\mathcal{R}^*(t, x) := -\frac{v_x(t, x)}{v_{xx}(t, x)}$, $\Sigma^*(t) := \Sigma(\rho^*(t))$, and X^* is the state process associated to α_t^* . Moreover,

$$\begin{aligned} V_0 = v(0, x_0) &= \sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} \mathbb{E}_\theta[U(X_T^\alpha)] = \mathbb{E}_{\theta^*}[U(X_T^{\alpha^*})] \\ &= \inf_{\theta \in \mathcal{V}_\Theta} \sup_{\alpha \in \mathcal{A}} \mathbb{E}_\theta[U(X_T^\alpha)]. \end{aligned} \quad (\text{A.42})$$

Proof. For any $\alpha \in \mathcal{A}$, and $\theta \in \mathcal{V}_\Theta$, the dynamics of $v(t, X_t^\alpha)$ under \mathbb{P}^θ is given by Itô's formula by

$$dv(t, X_t^\alpha) = D_t^{\alpha, \theta} dt + v_x(t, X_t^\alpha) \alpha_t^\top \sigma(\rho_t) dW_t^\theta,$$

where

$$D_t^{\alpha, \theta} = \frac{\partial v}{\partial t}(t, X_t^\alpha) + H_t(v_x(t, X_t^\alpha), v_{xx}(t, X_t^\alpha), \alpha_t, \theta_t), \quad 0 \leq t \leq T.$$

Observe that $\alpha_t^* = a^*(t, v_x(t, X_t^*), v_{xx}(t, X_t^*))$ as defined in Lemma A.1, and the Black-Scholes Merton Bellman PDE for v is written as

$$\frac{\partial v}{\partial t}(t, x) + H_t(v_x(t, x), v_{xx}(t, x), a^*(t, v_x(t, x), v_{xx}(t, x)), \theta^*(t)) = 0, \quad (t, x) \in [0, T] \times \mathbb{R}.$$

From the saddle point property in this Lemma A.1., we then have

$$D_t^{\alpha, \theta^*} \leq 0 \leq D_t^{\alpha^*, \theta}, \quad 0 \leq t < T, \quad \forall \alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta.$$

This implies that the process $(v(t, X_t^\alpha))_t$ is a local supermartingale under \mathbb{P}^{θ^*} , for any $\alpha \in \mathcal{A}$, while $(v(t, X_t^{\alpha^*}))_t$ is a local submartingale under \mathbb{P}^θ , for any $\theta \in \mathcal{V}_\Theta$. By considering a sequence of localizing stopping times $(\tau_n)_n$ converging to T as n goes to infinity, we then have

$$\mathbb{E}_{\theta^*}[v(\tau_n, X_{\tau_n}^\alpha)] \leq v(0, X_0) \leq \mathbb{E}_\theta[v(\tau_n, X_{\tau_n}^{\alpha^*})], \quad \forall \alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta.$$

From the growth condition on v , and as $U(X^\alpha)$ is uniformly integrable for $\alpha \in \mathcal{A}$, we deduce by sending n to infinity, and recalling that $v(T, x) = U(x)$:

$$\mathbb{E}_{\theta^*}[U(X_T^\alpha)] \leq v(0, x_0) \leq \mathbb{E}_\theta[U(X_T^{\alpha^*})], \quad \forall \alpha \in \mathcal{A}, \theta \in \mathcal{V}_\Theta.$$

As the deterministic process θ^* lies in particular in \mathcal{V}_Θ , and noting that $\sup_{\alpha \in \mathcal{A}} \inf_{\theta \in \mathcal{V}_\Theta} \leq \inf_{\theta \in \mathcal{V}_\Theta} \sup_{\alpha \in \mathcal{A}}$, this above saddle-point relation yields (A.42). \square

Remark A.1 A similar separation principle holds for robust portfolio selection in a discrete-time setting. We first compute at any time $t = 0, \dots, T-1$, the parameter $\theta^*(t) = (b^*(t), \rho^*(t))$, which achieves the minimum of the risk premium function $R(\theta) = b^\top \Sigma(\rho)^{-1} b$, over $\theta = (b, \rho)$ lying in the ambiguity set $\Theta(t)$ at time t . The solution to the robust portfolio selection problem is then given by the solution to the portfolio selection problem in the discrete-time model with mean return $b^*(t)$ and covariance matrix $\Sigma(\rho^*(t))$ at time t . \square

B Proofs of some Lemmas, Propositions and Theorems

Notations, differentiation and characterization of convex function

Let us introduce some notations and state some results which will be used frequently in the proof of some Lemmas and Propositions.

1. We introduce the so-called variance risk ratios

$$\begin{aligned} \hat{\kappa}_t(\rho) &:= \Sigma(\rho)^{-1} \hat{b}(t) = (\hat{\kappa}_t^1(\rho), \dots, \hat{\kappa}_t^d(\rho))^\top, \\ \kappa(b, \rho) &:= \Sigma(\rho)^{-1} b = (\kappa^1(b, \rho), \dots, \kappa^d(b, \rho))^\top. \end{aligned}$$

2. From some matrix calculations (see e.g. corollary 95 and corollary 105 in Dhrymes (1978)), we obtain the explicit expressions of the first partial derivatives of $R(b, \rho) = b^\top \kappa(b, \rho)$ with respect to b_i , ρ_{ij} denoted by $\frac{\partial R(b, \rho)}{\partial b_i}$ and $\frac{\partial R(b, \rho)}{\partial \rho_{ij}}$, $1 \leq i < j \leq d$,

$$\frac{\partial R(b, \rho)}{\partial b_i} = 2\kappa^i(b, \rho), \quad \frac{\partial R(b, \rho)}{\partial \rho_{ij}} = -\sigma_i \sigma_j \kappa^i(b, \rho) \kappa^j(b, \rho). \quad (\text{B.43})$$

We also denote by $\nabla_b R(b, \rho)$ and $\nabla_\rho R(b, \rho)$ the gradients of $R(b, \rho)$ with respect to b and ρ respectively,

$$\begin{cases} \nabla_b R(b, \rho) &= \left(\frac{\partial R(b, \rho)}{\partial b_1}, \dots, \frac{\partial R(b, \rho)}{\partial b_d} \right)^\top \\ \nabla_\rho R(b, \rho) &= \left(\frac{\partial R(b, \rho)}{\partial \rho_{12}}, \dots, \frac{\partial R(b, \rho)}{\partial \rho_{1d}}, \dots, \frac{\partial R(b, \rho)}{\partial \rho_{(d-1)d}} \right)^\top \end{cases} \quad (\text{B.44})$$

3. (*Sufficient and necessary optimality condition*). It is known (see e.g. Lemma 2.2 in Ding et al. (2018)) that $R(b, \rho)$ is jointly convex in b and ρ . Similarly, $R(\hat{b}(t), \rho)$ is convex in ρ . Then $\rho^*(t)$ is a global minimum of $\rho \mapsto R(\hat{b}(t), \rho)$ over $\Gamma(t)$ convex set of \mathbb{C}_+^d if and only if, for any $\rho \in \Gamma(t)$ (see e.g. section 4.2.3 in Boyd & Vandenberghe (2004)),

$$(\rho - \rho^*(t))^\top \nabla_\rho R(\hat{b}(t), \rho^*(t)) = \sum_{j=1}^d \sum_{i=1}^{j-1} \frac{\partial R(\hat{b}(t), \rho^*(t))}{\partial \rho_{ij}} (\rho_{ij} - \rho_{ij}^*) \geq 0,$$

which is written from (B.43) as,

$$\sum_{j=1}^d \sum_{i=1}^{j-1} \sigma_i \sigma_j \hat{\kappa}_t^i \hat{\kappa}_t^j (\rho^*(t)) (\rho_{ij} - \rho_{ij}^*) \leq 0. \quad (\text{B.45})$$

B.1 Proof of Lemma 3.2

Notice that if there exists $(b^*(t), \rho^*(t)) \in \arg \min_{\theta \in \Theta(t)} R(\theta)$, then the first-order condition implies that for any (b, ρ) lying in the convex set $\Theta(t)$:

$$(b - b^*(t))^\top \nabla_b R(\theta^*(t)) + (\rho - \rho^*(t))^\top \nabla_\rho R(\theta^*(t)) \geq 0. \quad (\text{B.46})$$

Recalling the expression of $H_t(b, \rho)$ in (3.9) and explicit expressions (B.43), (B.44) of $\nabla_b R(\theta^*(t))$ and $\nabla_\rho R(\theta^*(t))$, we have

$$\begin{aligned} H_t(b^*(t), \rho) - H_t(b^*(t), \rho^*(t)) &= \sum_{j=1}^d \sum_{i=1}^{j-1} \kappa^i(b^*(t), \rho^*(t)) \kappa^j(b^*(t), \rho^*(t)) \sigma_i \sigma_j (\rho_{ij} - \rho_{ij}^*(t)) \\ &= (\rho^*(t) - \rho)^\top \nabla_\rho R(\theta^*(t)), \\ 2(H_t(b^*(t), \rho^*(t)) - H_t(b, \rho^*(t))) &= 2 \sum_{i=1}^d (b_i^*(t) - b_i) \kappa^i(b^*(t), \rho^*(t)) \\ &= (b^*(t) - b)^\top \nabla_b R(\theta^*(t)), \end{aligned}$$

where by convention, we set: $\sum_{i=1}^0 \cdot = 0$. It follows immediately from the sum of the above two equalities that

$$\begin{aligned} &H_t(b^*(t), \rho^*(t)) - 2H_t(b, \rho^*(t)) + H_t(b^*(t), \rho) \\ &= (\rho^*(t) - \rho)^\top \nabla_\rho R(\theta^*(t)) + (b^*(t) - b)^\top \nabla_b R(\theta^*(t)) \leq 0, \end{aligned}$$

where we used (B.46) in the last inequality.

□

B.2 Proof of Proposition 4.1

Proposition 4.1 is an immediate combination of Theorem 3.1 and the following Lemma B.1.

Lemma B.1 *Fix $t \in [0, T]$. Let $\Theta(t)$ be an ellipsoidal set as in (4.25) with $p = 1$, and assume that there exists $\rho^*(t) \in \arg \min_{\rho \in \Gamma(t)} R(\hat{b}(t), \rho) = \arg \min_{\rho \in \Gamma(t)} \|\sigma(\rho)^{-1} \hat{b}(t)\|_2$. Then $\theta^*(t) = (b^*(t), \rho^*(t))$ with*

$$b^*(t) = \left(1 - \frac{\delta(t)}{\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2}\right) 1_{\{\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2 > \delta(t)\}} \hat{b}(t), \quad (\text{B.47})$$

$$\text{and } R(\theta^*(t)) = (\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2 - \delta(t))^2 1_{\{\|\sigma(\rho^*(t))^{-1} \hat{b}(t)\|_2 > \delta(t)\}}. \quad (\text{B.48})$$

Proof of Theorem B.1 Due to the dependence of b on ρ in the ellipsoidal set $\Theta(t)$ written as $\Theta(t) = \{(b, \rho) \in \mathbb{R}^d \times \Gamma(t) : b \in \Delta_\rho(t)\}$ where $\Delta_\rho(t) := \{b \in \mathbb{R}^d : \|\sigma(\rho)^{-1}(b - \hat{b}(t))\|_2 \leq \delta(t)\}$, we use a Lagrangian approach.

For fixed $\rho \in \Gamma(t)$, let us first focus on the inner minimization

$$\min_{b \in \Delta_\rho(t)} R(b, \rho). \quad (\text{B.49})$$

The Lagrangian with nonnegative multiplier μ associated to this constrained minimization problem is

$$L_t^1(b, \mu) = R(b, \rho) - \mu \left(|\delta(t)|^2 - (b - \hat{b}(t))^\top \Sigma(\rho)^{-1} (b - \hat{b}(t)) \right), \quad (\text{B.50})$$

and the first-order condition gives

$$\begin{aligned} \frac{\partial L_t^1(b, \mu)}{\partial b} &= 2\Sigma(\rho)^{-1}b + 2\mu\Sigma(\rho)^{-1}(b - \hat{b}(t)) = 0, \\ \frac{\partial L_t^1(b, \mu)}{\partial \mu} &= |\delta(t)|^2 - (b - \hat{b}(t))^\top \Sigma(\rho)^{-1} (b - \hat{b}(t)) = 0. \end{aligned}$$

Solving these two equations for fixed ρ , and recalling that the Lagrange multiplier is nonnegative, yield

$$\begin{cases} \mu_t^*(\rho) &= \left(\frac{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2}{\delta(t)} - 1 \right) 1_{\{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 > \delta(t)\}}, \\ b_t^*(\rho) &= \hat{b}(t) \left(1 - \frac{\delta(t)}{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2} \right) 1_{\{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 > \delta(t)\}}. \end{cases} \quad (\text{B.51})$$

Substituting these expressions into the Lagrangian (B.50), we get

$$L_t^1(b_t^*(\rho), \rho) = R(b_t^*(\rho), \rho) = (\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 - \delta(t))^2 1_{\{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 > \delta(t)\}},$$

and thus, the original problem $\inf_{\Theta(t)} R(\theta)$ is reduced to

$$\begin{aligned} \inf_{\theta=(b,\rho) \in \Theta(t)} R(\theta) &= \inf_{\rho \in \Gamma(t)} \inf_{b \in \Delta_\rho(t)} R(b, \rho) = \inf_{\rho \in \Gamma(t)} R(b_t^*(\rho), \rho) \\ &= \inf_{\rho \in \Gamma(t)} \left\{ (\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 - \delta(t))^2 1_{\{\|\sigma(\rho)^{-1} \hat{b}(t)\|_2 > \delta(t)\}} \right\} \\ &= \left(\inf_{\rho \in \Gamma(t)} \|\sigma(\rho)^{-1} \hat{b}(t)\|_2 - \delta(t) \right)^2 1_{\left\{ \inf_{\rho \in \Gamma(t)} \|\sigma(\rho)^{-1} \hat{b}(t)\|_2 > \delta(t) \right\}}. \end{aligned} \quad (\text{B.52})$$

Therefore, whenever $\rho^*(t) \in \arg \min_{\rho \in \Gamma(t)} \|\sigma(\rho)^{-1} \hat{b}(t)\|_2$ exists, we see from (B.52) that R attains its infimum at $\theta^*(t) = (b^*(t), \rho^*(t))$ with $b^*(t) = b_t^*(\rho^*(t))$ as in (B.51) with $\rho = \rho^*(t)$, which leads to the expressions of $(b^*(t), \rho^*(t))$ and $R(\theta^*(t))$ as described in (B.47) and (B.48). \square

B.3 Proof of Theorem 4.1

From the formula (3.6) of the optimal portfolio strategy in Theorem 3.1, we only need to calculate compute $\int_0^T R(b^*(s), \rho^*(s)) ds$ and vector $\Sigma(\rho^*(t))^{-1}b^*(t)$. The rest of this subsection is to calculate $R(b^*(t), \rho^*(t))$ and $\Sigma(\rho^*(t))^{-1}b^*(t)$.

For any $t \in [0, T]$, let us introduce for $1 \leq l \leq p$,

$$\Theta_l(t) = \{(b, \rho) \in \mathbb{R}^d \times \Gamma(t) : \|\sigma_{J_l}(\rho)^{-1}(b_{J_l} - \hat{b}_{J_l}(t))\|_2 \leq \delta_l\}, \quad (\text{B.53})$$

Lemma B.2 *Let $\Theta_l(t)$ be an ellipsoidal set as in (B.53) with $\Gamma(t) = \mathbb{C}_{>+}^d$, and assume there exists $m_l = m_l(t) \in J_l$ s.t. $|\hat{\beta}_{m_l}(t)| > \max_{j \in J_l, j \neq m_l} |\hat{\beta}_j(t)|$, then we have $\arg \min_{(b, \rho) \in \Theta_l(t)} \|\sigma_{J_l}(\rho)^{-1}b_{J_l}\|_2 \neq \emptyset$ attained at $(b^*(t), \rho^*(t))$ with*

$$\begin{aligned} \rho_{m_l j}^*(t) &= \hat{\rho}_{m_l j}(t), \quad j \in J_l, \quad j \neq m_l, \\ b_{J_l}^*(t) &= \hat{b}_{J_l}(t) \left(1 - \frac{\delta_l(t)}{|\hat{\beta}_{m_l}(t)|}\right) \mathbf{1}_{\{|\hat{\beta}_{m_l}(t)| > \delta_l(t)\}}. \end{aligned}$$

Moreover,

$$\min_{\Theta_l(t)} b_{J_l}^\top \Sigma_{J_l}(\rho)^{-1} b_{J_l} = (|\hat{\beta}_{m_l}(t)| - \delta_l(t))^2 \mathbf{1}_{\{|\hat{\beta}_{m_l}(t)| > \delta_l(t)\}}.$$

Proof. We assume w.l.o.g. that $l = 1$ and $m_l = 1$, i.e., $|\hat{\beta}_1(t)| > \max_{j \in J_1, j \neq 1} |\hat{\beta}_j(t)|$, otherwise we rearrange the assets. By noting that $b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1}$ is actually (squared) risk premium associated to the assets in the subclass J_1 , and that Lemma B.1 is valid for any number of assets. we deduce from Lemma B.1 that

$$\begin{aligned} &\inf_{(b, \rho) \in \Theta_1(t)} b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1} \\ &= \left(\inf_{\rho \in \mathbb{C}_{>+}^d} \|\sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t)\|_2 - \delta_1(t) \right)^2 \mathbf{1}_{\left\{ \inf_{\rho \in \mathbb{C}_{>+}^d} \|\sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t)\|_2 > \delta_1(t) \right\}}. \end{aligned}$$

Let us show that if $|\hat{\beta}_1(t)| > \max_{j \in J_1, j \neq 1} |\hat{\beta}_j(t)|$, then $\inf_{\rho \in \mathbb{C}_{>+}^d} \|\sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t)\|_2$ is attained over $\mathbb{C}_{>+}^d$. The key point is that $(\hat{b}_{J_1}(t))^\top \Sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t)$ is written as the sum of two nonnegative parts by matrix transformations. The procedures are as follows: we express $\Sigma_{J_1}(\rho)$ as the form of block matrix

$$\Sigma_{J_1}(\rho) = \left(\begin{array}{c|c} \sigma_1^2 & \Sigma_{J_1-1,1}(\rho)^\top \\ \hline \Sigma_{J_1-1,1}(\rho) & \Sigma_{J_1-1}(\rho) \end{array} \right),$$

by transforming $\Sigma_{J_1}(\rho)$ to block diagonal matrix, and taking the inverse, we then obtain

$$\Sigma_{J_1}(\rho)^{-1} = \left(\begin{array}{cc} 1 & -\frac{\Sigma_{J_1-1,1}(\rho)^\top}{\sigma_1^2} \\ 0_{J_1-1,1} & I_{J_1-1} \end{array} \right) \left(\begin{array}{cc} \frac{1}{\sigma_1^2} & 0_{J_1-1,1}^\top \\ 0_{J_1-1,1} & A_{J_1-1}(\rho)^{-1} \end{array} \right) \left(\begin{array}{cc} 1 & -\frac{\Sigma_{J_1-1,1}(\rho)^\top}{\sigma_1^2} \\ 0_{J_1-1,1} & I_{J_1-1} \end{array} \right)^\top \quad (\text{B.54})$$

where I_{J_1-1} is $(|J_1| - 1) \times (|J_1| - 1)$ unit matrix and $A_{J_1-1}(\rho) := \Sigma_{J_1-1}(\rho) - \frac{\Sigma_{J_1-1,1}(\rho) \Sigma_{J_1-1,1}(\rho)^\top}{\sigma_1^2}$ lies in $\mathbb{S}_{>+}^d$. From the expression of $\Sigma_{J_1}(\rho)^{-1}$ in (B.54) and by writing $\hat{b}_{J_1}(t)$ in corresponding block form $(\hat{b}_1(t), (\hat{b}_{J_1-1}(t))^\top)^\top$, we get

$$\begin{aligned} &(\hat{b}_{J_1}(t))^\top \Sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t) \\ &= |\hat{\beta}_1(t)|^2 + (\hat{b}_{J_1-1}(t) - \frac{\hat{b}_1(t)}{\sigma_1^2} \Sigma_{J_1-1,1}(\rho)^\top)^\top A_{J_1-1}(\rho)^{-1} (\hat{b}_{J_1-1}(t) - \frac{\hat{b}_1(t)}{\sigma_1^2} \Sigma_{J_1-1,1}(\rho)) \\ &\geq |\hat{\beta}_1(t)|^2, \end{aligned}$$

and thus $\|\sigma_{J_1}(\rho^*(t))^{-1}\hat{b}_{J_1}(t)\|_2 = |\hat{\beta}_1(t)|$ only when the second term is zero

$$\hat{b}_{J_1-1}(t) - \frac{\hat{b}_1(t)}{\sigma_1^2}\Sigma_{J_1-1,1}(\rho^*(t)) = 0, \quad (\text{B.55})$$

which has the explicit solution

$$\rho_{1j}^*(t) = \hat{\varrho}_{1j}(t) \in (-1, 1), \quad j \in J_1, j \neq 1. \quad (\text{B.56})$$

Therefore, we deduce from (B.56) and (B.47) that

$$\rho_{1j}^*(t) = \hat{\varrho}_{1j}(t), j \in J_1, j \neq 1, \quad b_{J_1}^*(t) = \hat{b}_{J_1}(t) \left(1 - \frac{\delta_1(t)}{|\hat{\beta}_1(t)|}\right) \mathbf{1}_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}}. \quad (\text{B.57})$$

Once $(\rho_{1j}^*(t))_{j \in J_1, j \neq 1}$ is given in the above equality, we can complete the other values of $\rho_{ij}^*(t)$ such that $\rho^*(t) \in \mathbb{C}_{>+}^d$. For instance, $\rho_{1j}^*(t) = \hat{\varrho}_{1j}(t)$, $1 < j \leq d$, $\rho_{ij}^*(t) = \hat{\varrho}_{1i}(t)\hat{\varrho}_{1j}(t)$, $1 < i \neq j \leq d$. It is easy to check that such a construction of $\rho^*(t) \in \mathbb{C}_{>+}^d$. Moreover in this case, it follows from (B.54) and (B.55) that

$$\Sigma_{J_1}(\rho^*(t))^{-1}\hat{b}_{J_1}(t) = \left(\frac{\hat{b}_1(t)}{\sigma_1^2}, 0, \dots, 0\right)^\top. \quad (\text{B.58})$$

□

Let us now prove Theorem 4.1: if $|\hat{\beta}_{m_l}(t)| > \max_{j \in J_l, j \neq m_l} |\hat{\beta}_j(t)|$ and $|\hat{\beta}_{m_k}(t)| - \delta_k(t) > \max_{1 \leq l \leq p, l \neq k} (|\hat{\beta}_{m_l}(t)| - \delta_l(t))$, then $\inf_{\Theta(t)} R(b, \rho)$ exists and $(b^*(t), \rho^*(t)) \in \arg \min_{(b, \rho) \in \Theta(t)} R(b, \rho)$ is computed explicitly. Let us consider w.l.o.g. the case of $p = 2$ subsets, and reorder the assets s.t. $J_1 = \{1, \dots, k^o - 1\}$, $J_2 = \{k^o, \dots, d\}$ and $m_1 = 1$, $m_2 = k^o$ for some $1 \leq k^o \leq d$, i.e.

$$|\hat{\beta}_1(t)| > \max_{j \in J_1, j \neq 1} |\hat{\beta}_j(t)|, \quad |\hat{\beta}_{k^o}(t)| > \max_{j \in J_2, j \neq k^o} |\hat{\beta}_j(t)|, \quad |\hat{\beta}_1(t)| - \delta_1(t) > |\hat{\beta}_{k^o}(t)| - \delta_2(t). \quad (\text{B.59})$$

Notice that $R(b, \rho)$ can be expressed as the sum of two nonnegative parts in the same way as $(\hat{b}_{J_1}(t))^\top \Sigma_{J_1}(\rho)^{-1} \hat{b}_{J_1}(t)$ in Lemma B.2. $\Sigma(\rho)$ is written in a form of blocks as follows

$$\Sigma(\rho) = \left(\begin{array}{c|c} \Sigma_{J_1}(\rho) & \Sigma_{J_{21}}(\rho)^\top \\ \hline \Sigma_{J_{21}}(\rho) & \Sigma_{J_2}(\rho) \end{array} \right),$$

and its inverse $\Sigma(\rho)$ is in the form

$$\Sigma(\rho)^{-1} = \begin{pmatrix} I_{J_1} & -\Sigma_{J_1}(\rho)^{-1}\Sigma_{J_{21}}(\rho)^\top \\ 0_{J_{21}} & I_{J_2} \end{pmatrix} \begin{pmatrix} \Sigma_{J_1}(\rho)^{-1} & 0_{J_{12}} \\ 0_{J_{21}} & A_{J_2}(\rho)^{-1} \end{pmatrix} \begin{pmatrix} I_{J_1} & -\Sigma_{J_1}(\rho)^{-1}\Sigma_{J_{21}}(\rho)^\top \\ 0_{J_{21}} & I_{J_2} \end{pmatrix}^\top, \quad (\text{B.60})$$

where I_{J_l} , $l = 1, 2$, is $|J_l| \times |J_l|$ unit matrix, and $A_{J_2}(\rho) := \Sigma_{J_2}(\rho) - \Sigma_{J_{21}}(\rho)\Sigma_{J_1}(\rho)^{-1}\Sigma_{J_{21}}(\rho)^\top$ lies in $\mathbb{S}_{>+}^d$. Recalling that $R(b, \rho) = b^\top \Sigma(\rho)^{-1} b$ and rewriting vector b in corresponding block matrix form $(b_{J_1}^\top | b_{J_2}^\top)^\top$, together with (B.60), we express $R(b, \rho)$ as two nonnegative terms

$$\begin{aligned} & R(b, \rho) \\ &= b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1} + (b_{J_2} - \Sigma_{J_{21}}(\rho)\Sigma_{J_1}(\rho)^{-1} b_{J_1})^\top A_{J_2}(\rho)^{-1} (b_{J_2} - \Sigma_{J_{21}}(\rho)\Sigma_{J_1}(\rho)^{-1} b_{J_1}) \\ &\geq b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1} \geq (|\hat{\beta}_1(t)| - \delta_1(t))^2 \mathbf{1}_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}}, \end{aligned}$$

where we used $A_{J_2}(\rho) \in \mathbb{S}_{>+}^d$ in the first inequality, and the second inequality is from $\Theta(t) = \Theta_1(t) \cap \Theta_2(t)$, hence $\inf_{\Theta(t)} b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1} \geq \inf_{\Theta_1(t)} b_{J_1}^\top \Sigma_{J_1}(\rho)^{-1} b_{J_1}$, and Lemma B.2.

Therefore,

$R(b^*(t), \rho^*(t)) = (|\hat{\beta}_1(t)| - \delta_1(t))^2 \mathbf{1}_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}}$ is minimum if and only if $(b^*(t), \rho^*(t)) \in \Theta(t)$ satisfies (B.57) and

$$b_{J_2}^*(t) - \Sigma_{J_{21}}(\rho^*(t)) \Sigma_{J_1}(\rho^*(t))^{-1} b_{J_1}^*(t) = 0, \quad (\text{B.61})$$

which yields, together with (B.58), the explicit form,

$$\beta_j^*(t) = \beta_1^*(t) \rho_{1j}^*(t) = \hat{\beta}_1(t) \left(1 - \frac{\delta_1(t)}{|\hat{\beta}_1(t)|}\right) \mathbf{1}_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}} \rho_{1j}^*(t), \quad j \in J_2. \quad (\text{B.62})$$

In fact when (B.59) holds, there exists an element $\theta^*(t) = (b^*(t), \rho^*(t)) \in \Theta(t)$ attaining this infimum. For instance, we construct $(b^*(t), \rho^*(t))$ satisfying (B.57)-(B.62) in the form

$$\begin{cases} \beta_{J_1}^*(t) &= (1 - \frac{\delta_1(t)}{|\hat{\beta}_1(t)|}) \mathbf{1}_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}} \hat{\beta}_{J_1}(t), \\ \beta_{J_2}^*(t) &= (1 - \frac{\delta_2(t)}{|\hat{\beta}_{k^o}(t)|}) \mathbf{1}_{\{|\hat{\beta}_{k^o}(t)| > \delta_2(t)\}} \hat{\beta}_{J_2}(t), \end{cases} \quad (\text{B.63})$$

$$\text{and } \begin{cases} \rho_{1j}^*(t) &= \beta_j^*(t) / \beta_1^*(t) = \hat{\rho}_{1j}(t), \quad j \in J_1, j \neq 1, \\ \rho_{k^o j}^*(t) &= \beta_j^*(t) / \beta_{k^o}^*(t) = \hat{\rho}_{k^o j}(t), \quad j \in J_2, j \neq k^o, \\ \rho_{ij}^*(t) &= \hat{\rho}_{1i}(t) \hat{\rho}_{1j}(t), \quad 1 < i \neq j \in J_1, \\ \rho_{ij}^*(t) &= \hat{\rho}_{k^o i}(t) \hat{\rho}_{k^o j}(t), \quad k^o < i \neq j \in J_2, \\ \rho_{1j}^*(t) &= \beta_j^*(t) / \beta_1^*(t) = \frac{(1 - \frac{\delta_2(t)}{|\hat{\beta}_{k^o}(t)|}) \hat{\beta}_j(t)}{(1 - \frac{\delta_1(t)}{|\hat{\beta}_1(t)|}) \hat{\beta}_1(t)} \in (-1, 1), \quad j \in J_2, \\ \rho_{ij}^*(t) &= \rho_{1i}^*(t) \rho_{1j}^*(t), \quad \text{otherwise,} \end{cases} \quad (\text{B.64})$$

The rest of the proof is to check that $(b^*(t), \rho^*(t))$ given in (B.63)-(B.64) belongs to $\Theta(t)$, i.e., $\rho^*(t) \in \mathbb{C}_{>+}^d$ and $\|\sigma_{J_l}(\rho^*(t))^{-1} (b_{J_l}^*(t) - \hat{b}_{J_l}(t))\|_2 \leq \delta_l(t)$, $l = 1, 2$.

• Let us first check that $\rho^*(t)$ in (B.64) belongs to $\mathbb{C}_{>+}^d$. In this case, $C(\rho^*(t))$ is written in the form (in what follows we often omit the dependence in t of $\beta^* = \beta^*(t)$ and $\rho^* = \rho^*(t)$),

$$C(\rho^*) = \left(\begin{array}{cccc|cccc} 1 & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1k^o-1} & \frac{\beta_{k^o}^*}{\beta_1^*} & \frac{\beta_{k^o+1}^*}{\beta_1^*} & \cdots & \frac{\beta_d^*}{\beta_1^*} \\ \hat{\rho}_{12} & 1 & \cdots & \hat{\rho}_{12} \hat{\rho}_{1k^o-1} & \frac{\beta_2^* \beta_{k^o}^*}{|\beta_1^*|^2} & \frac{\beta_2^* \beta_{k^o+1}^*}{|\beta_1^*|^2} & \cdots & \frac{\beta_2^* \beta_d^*}{|\beta_1^*|^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \hat{\rho}_{1k^o-1} & \hat{\rho}_{12} \hat{\rho}_{1k^o-1} & \cdots & 1 & \frac{\beta_{k^o-1}^* \beta_{k^o}^*}{|\beta_1^*|^2} & \frac{\beta_{k^o-1}^* \beta_{k^o+1}^*}{|\beta_1^*|^2} & \cdots & \frac{\beta_{k^o-1}^* \beta_d^*}{|\beta_1^*|^2} \\ \hline \frac{\beta_{k^o}^*}{\beta_1^*} & \frac{\beta_2^* \beta_{k^o}^*}{|\beta_1^*|^2} & \cdots & \frac{\beta_{k^o-1}^* \beta_{k^o}^*}{|\beta_1^*|^2} & 1 & \hat{\rho}_{k^o k^o+1} & \cdots & \hat{\rho}_{k^o d} \\ \frac{\beta_{k^o+1}^*}{\beta_1^*} & \frac{\beta_2^* \beta_{k^o+1}^*}{|\beta_1^*|^2} & \cdots & \frac{\beta_{k^o-1}^* \beta_{k^o+1}^*}{|\beta_1^*|^2} & \hat{\rho}_{k^o k^o+1} & 1 & \cdots & \hat{\rho}_{k^o k^o+1} \hat{\rho}_{k^o d} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\beta_d^*}{\beta_1^*} & \frac{\beta_2^* \beta_d^*}{|\beta_1^*|^2} & \cdots & \frac{\beta_{k^o-1}^* \beta_{k^o}^*}{|\beta_1^*|^2} & \hat{\rho}_{k^o d} & \hat{\rho}_{k^o k^o+1} \hat{\rho}_{k^o d} & \cdots & 1 \end{array} \right),$$

observe that

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{\beta_2^*}{\beta_1^*} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\beta_d^*}{\beta_1^*} & 0 & \cdots & 1 \end{pmatrix} C(\rho^*) \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{\beta_2^*}{\beta_1^*} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\beta_d^*}{\beta_1^*} & 0 & \cdots & 1 \end{pmatrix}^\top = \left(\begin{array}{c|c} \hat{C}_{J_1}(\rho^*) & 0_{J_1}^\top \\ \hline 0_{J_{21}} & \hat{C}_{J_2}(\rho^*) \end{array} \right)$$

where $\hat{C}_{J_1}(\rho^*)$ is a diagonal matrix in the form $\text{diag}(1, 1 - |\hat{\varrho}_{12}|^2, \dots, 1 - |\hat{\varrho}_{1k^o-1}|^2)$, and

$$\hat{C}_{J_2}(\rho^*) = \begin{pmatrix} 1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2 & \frac{\beta_{k^o+1}^*}{\beta_{k^o}^*} (1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2) & \dots & \frac{\beta_d^*}{\beta_{k^o}^*} (1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2) \\ \frac{\beta_{k^o+1}^*}{\beta_{k^o}^*} (1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2) & 1 - \left| \frac{\beta_{k^o+1}^*}{\beta_1^*} \right|^2 & \dots & \frac{\beta_{k^o+1}^* \beta_d^*}{|\beta_{k^o}^*|^2} - \frac{\beta_{k^o+1}^* \beta_d^*}{|\beta_1^*|^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta_d^*}{\beta_{k^o}^*} (1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2) & \frac{\beta_{k^o+1}^* \beta_d^*}{|\beta_{k^o}^*|^2} - \frac{\beta_{k^o+1}^* \beta_d^*}{|\beta_1^*|^2} & \dots & 1 - \left| \frac{\beta_d^*}{\beta_1^*} \right|^2 \end{pmatrix},$$

then notice that

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & \dots & 0 \\ -\frac{\beta_{k^o+1}^*}{\beta_{k^o}^*} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\beta_d^*}{\beta_{k^o}^*} & 0 & \dots & 1 \end{pmatrix} \hat{C}_{J_2}(\rho^*) \begin{pmatrix} 1 & 0 & \dots & 0 \\ -\frac{\beta_{k^o+1}^*}{\beta_{k^o}^*} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\beta_d^*}{\beta_{k^o}^*} & 0 & \dots & 1 \end{pmatrix}^\top \\ &= \text{diag}(1 - \left| \frac{\beta_{k^o}^*}{\beta_1^*} \right|^2, 1 - |\hat{\varrho}_{k^o k^o+1}|^2, \dots, 1 - |\hat{\varrho}_{k^o d}|^2). \end{aligned}$$

Since $\hat{C}_{J_1}(\rho^*)$ and $\hat{C}_{J_2}(\rho^*)$ are both symmetric positive definite matrices, then we have $\rho^*(t)$ as in (B.64) belongs to $\mathbb{C}_{>+}^d$.

• Then it is easily checked that $\|\sigma_{J_l}(\rho^*(t))^{-1}(b_{J_l}^*(t) - \hat{b}_{J_l}(t))\|_2 \leq \delta_l$, $l = 1, 2$, with $(b^*(t), \rho^*(t))$ given by (B.63)-(B.64) because we have

$$\begin{aligned} \|\sigma_{J_l}(\rho^*(t))^{-1}(b_{J_l}^*(t) - \hat{b}_{J_l}(t))\|_2 &= \frac{\delta_l(t)}{|\hat{\beta}_{m_l}(t)|} \|\sigma_{J_l}(\rho^*(t))^{-1} \hat{b}_{J_l}(t)\|_2 1_{\{|\hat{\beta}_{m_l}(t)| > \delta_l(t)\}} \\ &= \delta_l(t) 1_{\{|\hat{\beta}_{m_l}(t)| > \delta_l(t)\}} \leq \delta_l(t), \end{aligned}$$

where we used $\|\sigma_{J_l}(\rho^*(t))^{-1} \hat{b}_{J_l}(t)\|_2 = |\hat{\beta}_{m_l}(t)|$, $l = 1, 2$ in Lemma B.2.

Consequently, we deduce that $(b^*(t), \rho^*(t)) \in \Theta(t)$ given by (B.63)-(B.64) achieves the minimal risk premium

$$R(b^*(t), \rho^*(t)) = (|\hat{\beta}_1(t)| - \delta_1(t))^2 1_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}}, \quad (\text{B.65})$$

and with (B.60), (B.61) and (B.58) that

$$\Sigma(\rho^*(t))^{-1} b^*(t) = \left(\frac{\hat{b}_1(t)}{\sigma_1^2}, 0, \dots, 0 \right)^\top \left(1 - \frac{\delta_1(t)}{|\hat{\beta}_1(t)|} \right) 1_{\{|\hat{\beta}_1(t)| > \delta_1(t)\}}, \quad (\text{B.66})$$

which completes the proof. \square

B.4 Proof of Theorem 4.2

In light of formula (4.27) of the optimal portfolio strategy in Proposition 4.1, we only need to compute $\int_0^T (\|\sigma(\rho^*(s))^{-1} \hat{b}(s)\|_2 - \delta(s))^2 1_{\{\|\sigma(\rho^*(s))^{-1} \hat{b}(s)\|_2 > \delta(s)\}} ds$ and vector $\hat{\kappa}_t(\rho^*(t)) = \Sigma(\rho^*(t))^{-1} \hat{b}(t)$.

As $\Gamma(t) = [\underline{\rho}(t), \bar{\rho}(t)]$ is compact for fixed $t \in [0, T]$, we know that $\rho^*(t) = \arg \min_{\rho \in \Gamma(t)} R(\hat{b}(t), \rho)$ exists, and

from Lemma B.1, we only need to compute the minimum of the function $\rho \mapsto R(\hat{b}(t), \rho)$ over $\Gamma(t)$. From (B.45) with $d = 2$, we obtain the sufficient and necessary condition of $\rho^*(t)$ for being global minimum of $R(\hat{b}(t), \rho)$ over $\Gamma(t)$:

$$\sigma_1 \sigma_2 \hat{\kappa}_t^1(\rho^*(t)) \hat{\kappa}_t^2(\rho^*(t)) (\rho - \rho^*(t)) \leq 0, \quad \text{for all } \rho \in [\underline{\rho}(t), \bar{\rho}(t)], \quad (\text{B.67})$$

where $\hat{\kappa}_t(\rho)$ is explicitly written as

$$\hat{\kappa}_t(\rho) = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{\hat{b}_1(t)}{\sigma_1^2} - \frac{\hat{b}_2(t)}{\sigma_1\sigma_2}\rho \\ \frac{\hat{b}_2(t)}{\sigma_2^2} - \frac{\hat{b}_1(t)}{\sigma_1\sigma_2}\rho \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{\hat{\beta}_1(t) - \hat{\beta}_2(t)\rho}{\sigma_1} \\ \frac{\hat{\beta}_2(t) - \hat{\beta}_1(t)\rho}{\sigma_2} \end{pmatrix}. \quad (\text{B.68})$$

From (B.67), we have following possible cases for fixed t :

- $\hat{\kappa}_t^1(\rho^*(t))\hat{\kappa}_t^2(\rho^*(t)) = 0$. From the explicit expression (B.68) of $\hat{\kappa}(\rho^*(t))$ and definition of $\hat{\varrho}_{12}(t)$ in (4.26), we obtain $(\hat{\varrho}_{12}(t) - \rho^*(t))(1 - \hat{\varrho}_{12}(t)\rho^*(t)) = 0$, and as $\rho^*(t)$ has to belong to $[\underline{\rho}(t), \bar{\rho}(t)] \subset (-1, 1)$, we obtain $\rho^*(t) = \hat{\varrho}_{12}(t)$, and so $R(\hat{b}(t), \rho^*(t)) = (\max(|\hat{\beta}_1(t)|, |\hat{\beta}_2(t)|))^2$.
- $\hat{\kappa}_t^1(\rho^*(t))\hat{\kappa}_t^2(\rho^*(t)) > 0$. Then (B.67) is satisfied iff $\rho^*(t) = \bar{\rho}(t)$. Moreover, from the above explicit expression of $\hat{\kappa}_t(\rho^*(t))$, we obtain $\bar{\rho}(t) < \hat{\varrho}_{12}(t)$.
- $\hat{\kappa}_t^1(\rho^*(t))\hat{\kappa}_t^2(\rho^*(t)) < 0$. Then (B.67) is satisfied iff $\rho^*(t) = \underline{\rho}(t)$. Moreover, from the explicit expression of $\hat{\kappa}_t(\rho^*(t))$, we obtain $\underline{\rho}(t) > \hat{\varrho}_{12}(t)$.

We obtain $\alpha^*(t)$ described as in Theorem 4.2. □

B.5 Proof of Theorem 4.3

In view of formula (4.27) of the optimal portfolio strategy in Proposition 4.1, we only need to compute $\hat{\kappa}(\rho^*) = \Sigma(\rho^*)^{-1}\hat{b}$, and $\|\sigma(\rho^*)^{-1}\hat{b}\|_2$, i.e., $R(\hat{b}, \rho^*)$.

As $\Gamma = \prod_{j=1}^3 \prod_{i=1}^{j-1} [\underline{\rho}_{ij}, \bar{\rho}_{ij}]$ is compact, we already know that $\rho^* = \arg \min_{\rho \in \Gamma} R(\hat{b}, \rho)$ exists. From Lemma B.1,

we only need to compute the minimum of the function $\rho \mapsto R(\hat{b}, \rho)$ over Γ by applying the optimality principle (B.45) when $d = 3$,

$$\sum_{j=1}^3 \sum_{i=1}^{j-1} \sigma_i \sigma_j \hat{\kappa}^i(\rho^*) \hat{\kappa}^j(\rho^*) (\rho_{ij} - \rho_{ij}^*) \leq 0 \quad \text{for any } \rho \in \Gamma. \quad (\text{B.69})$$

Observe from (B.69) that, each ρ_{ij}^* , $1 \leq i < j \leq 3$ may be lower bound $\underline{\rho}_{ij}$, upper bound $\bar{\rho}_{ij}$, or an interior point in $(\underline{\rho}_{ij}, \bar{\rho}_{ij})$, which corresponds to $\kappa^i \kappa^j(\rho^*) > 0$, $\kappa^i \kappa^j(\rho^*) < 0$, or $\kappa^i \kappa^j(\rho^*) = 0$ respectively. Therefore, let us consider the following possible exclusive cases depending on the number of zero components in $\hat{\kappa}(\rho^*)$:

1. $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*) = 0$, $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*) = 0$, $\hat{\kappa}^2 \hat{\kappa}^3(\rho^*) = 0$.

In this case, (B.69) is immediately satisfied. As we assume that $\hat{b} \neq 0$, $\hat{\kappa}(\rho^*)$ is not zero, i.e., at least one component of $\hat{\kappa}(\rho^*)$ is nonzero. Then, two components of $\hat{\kappa}(\rho^*)$ are zero. Under the assumption that $|\hat{\beta}_1| \geq |\hat{\beta}_2| \geq |\hat{\beta}_3|$, (B.66), (B.65) in Section B.3 yield the explicit expressions of ρ^* , $\hat{\kappa}(\rho^*)$ and $R(\hat{b}, \rho^*)$

$$\rho_{12}^* = \hat{\varrho}_{12} \in [\underline{\rho}_{12}, \bar{\rho}_{12}], \quad \rho_{13}^* = \hat{\varrho}_{13} \in [\underline{\rho}_{13}, \bar{\rho}_{13}], \quad \text{any } \rho_{23}^* \in [\underline{\rho}_{23}, \bar{\rho}_{23}]$$

and

$$\hat{\kappa}(\rho^*) = \left(\frac{\hat{b}_1}{\sigma_1^2}, 0, 0 \right)^\top, \quad R(\hat{b}, \rho^*) = |\hat{\beta}_1|^2. \quad (\text{B.70})$$

Let us show that $|\hat{\beta}_1|^2$ in (B.70) is strict minimum value in the sense that $R(\hat{b}, \rho^*) = |\hat{\beta}_1|^2$ if and only if $\rho_{12}^* = \hat{\varrho}_{12} \in [\underline{\rho}_{12}, \bar{\rho}_{12}]$, $\rho_{13}^* = \hat{\varrho}_{13} \in [\underline{\rho}_{13}, \bar{\rho}_{13}]$ and any $\rho_{23}^* \in [\underline{\rho}_{23}, \bar{\rho}_{23}]$. We express $\Sigma(\rho)$ as the following block matrix

$$\Sigma(\rho) = \begin{pmatrix} \sigma_1^2 & C_1^\top \\ C_1 & \Sigma_{-1}(\rho_{23}) \end{pmatrix},$$

where the vector $C_1 = (\sigma_1\sigma_2\rho_{12}, \sigma_1\sigma_3\rho_{13})^\top$.

Noting that

$$\begin{pmatrix} 1 & \mathbf{0}_{1 \times 2} \\ -\frac{C_1}{\sigma_1^2} & I_{2 \times 2} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & C_1^\top \\ C_1 & \Sigma_{-1}(\rho_{23}) \end{pmatrix} \begin{pmatrix} 1 & -\frac{C_1^\top}{\sigma_1^2} \\ \mathbf{0}_{2 \times 1} & I_{2 \times 2} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{2 \times 1} & A \end{pmatrix}, \quad (\text{B.71})$$

where $I_{2 \times 2}$ denotes 2×2 identity matrix and $A = \Sigma_{-1}(\rho_{23}) - \frac{C_1 C_1^\top}{\sigma_1^4}$ is 2×2 positive definite matrix, and inverting on both sides of (B.71), we get

$$\Sigma^{-1}(\rho) = \begin{pmatrix} 1 & -\frac{C_1^\top}{\sigma_1^2} \\ \mathbf{0}_{2 \times 1} & I_{2 \times 2} \end{pmatrix} \begin{pmatrix} \sigma_1^{-2} & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{2 \times 1} & A^{-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}_{1 \times 2} \\ -\frac{C_1}{\sigma_1^2} & I_{2 \times 2} \end{pmatrix}. \quad (\text{B.72})$$

We express \hat{b} as $(\hat{b}_1, \hat{b}_{-1}^\top)^\top$ and then write $R(\hat{b}, \rho)$ as two nonnegative decompositions from (B.72),

$$\begin{aligned} R(\hat{b}, \rho) &= |\hat{\beta}_1|^2 + (\hat{b}_{-1} - \frac{\hat{b}_1}{\sigma_1^2} C_1)^\top A^{-1} (\hat{b}_{-1} - \frac{\hat{b}_1}{\sigma_1^2} C_1) \\ &\geq |\hat{\beta}_1|^2, \end{aligned} \quad (\text{B.73})$$

where in the last inequality, '=' holds if and only if $\hat{b}_{-1} - \frac{\hat{b}_1}{\sigma_1^2} C_1 = 0$, i.e., $\rho_{12}^* = \hat{\rho}_{12}$, $\rho_{13}^* = \hat{\rho}_{13}$. This corresponds to case **1.** of Theorem 4.3.

2. $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*) \neq 0$, $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*) = 0$, $\hat{\kappa}^2 \hat{\kappa}^3(\rho^*) = 0$.

In this case, we express $\Sigma(\rho)$ as the following block-matrix form for convenience,

$$\Sigma(\rho) = \begin{pmatrix} \Sigma_{-3}(\rho_{12}) & C_3 \\ C_3^\top & \sigma_3^2 \end{pmatrix},$$

where the vector $C_3 = (\sigma_1\sigma_3\rho_{13}, \sigma_2\sigma_3\rho_{23})^\top$.

By first transforming $\Sigma(\rho)$ to block diagonal matrix as (B.72) and then taking inverse, we obtain

$$\begin{aligned} \Sigma(\rho)^{-1} &= \begin{pmatrix} I_{2 \times 2} & -\Sigma_{-3}(\rho_{12})^{-1} C_3 \\ \mathbf{0}_{1 \times 2} & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{-3}(\rho_{12})^{-1} & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{1 \times 2} & a(\rho)^{-1} \end{pmatrix} \\ &\quad \begin{pmatrix} I_{2 \times 2} & \mathbf{0}_{2 \times 1} \\ -C_3^\top \Sigma_{-3}(\rho_{12})^{-1} & 1 \end{pmatrix}, \end{aligned} \quad (\text{B.74})$$

where $a(\rho) = \sigma_3^2 - C_3^\top \Sigma_{-3}(\rho_{12})^{-1} C_3$ is positive.

Recalling the definition of $\kappa(\hat{b}, \rho)$ and $R(\hat{b}, \rho)$, we obtain from (B.74)

$$\begin{cases} \begin{pmatrix} \hat{\kappa}^1(\rho) \\ \hat{\kappa}^2(\rho) \end{pmatrix} &= \Sigma_{-3}(\rho_{12})^{-1} \hat{b}_{-3} - \hat{\kappa}^3(\rho) \Sigma_{-3}(\rho_{12})^{-1} C_3 \\ \hat{\kappa}^3(\rho) &= \frac{1}{a(\rho)} (\hat{b}_3 - C_3^\top \Sigma_{-3}(\rho_{12})^{-1} \hat{b}_{-3}), \end{cases} \quad (\text{B.75})$$

and

$$R(\hat{b}, \rho) = \hat{b}_{-3}^\top \Sigma_{-3}(\rho_{12})^{-1} \hat{b}_{-3} + a(\rho) |\hat{\kappa}^3(\rho)|^2. \quad (\text{B.76})$$

In the following, we write $\hat{b}_{-3}^\top \Sigma_{-3}(\rho_{12})^{-1} \hat{b}_{-3}$ as $R(\hat{b}_{-3}, \rho_{12})$.

As $\hat{\kappa}^3(\rho^*) = 0$, we obtain from (B.69) that

$$\sigma_1 \sigma_2 \hat{\kappa}^1 \hat{\kappa}^2(\rho_{12}^*)(\rho_{12} - \rho_{12}^*) \leq 0 \quad \text{for all } \rho_{12} \in [\underline{\rho}_{12}, \bar{\rho}_{12}] \quad (\text{B.77})$$

and from (B.75) and (B.76) that

$$\begin{cases} \begin{pmatrix} \hat{\kappa}^1(\rho_{12}^*) \\ \hat{\kappa}^2(\rho_{12}^*) \end{pmatrix} = \Sigma_{-3}(\rho_{12}^*)^{-1} \hat{b}_{-3} \\ R(\hat{b}, \rho^*) = R(\hat{b}_{-3}, \rho_{12}^*). \end{cases} \quad (\text{B.78})$$

This is the case of ambiguous correlation in the two risky assets: the first and second assets with ambiguous correlation ρ_{12} in $[\underline{\rho}_{12}, \bar{\rho}_{12}]$. In this case, $\hat{\kappa}^1(\rho^*)$ and $\hat{\kappa}^2(\rho^*)$ are not zero, therefore we have that from Theorem 4.2

$$\rho_{12}^* = \bar{\rho}_{12} 1_{\{\bar{\rho}_{12} < \hat{\rho}_{12}\}} + \underline{\rho}_{12} 1_{\{\underline{\rho}_{12} > \hat{\rho}_{12}\}}. \quad (\text{B.79})$$

By setting $g(\rho_{12}^*, \rho_{13}, \rho_{23}) := a(\rho_{12}^*, \rho_{13}, \rho_{23}) \kappa^3(\rho_{12}^*, \rho_{13}, \rho_{23})$ for fixed ρ_{12}^* in (B.79), we deduce from (B.75) that the function

$$(\rho_{13}, \rho_{23}) \mapsto g(\rho_{12}^*, \rho_{13}, \rho_{23}) = \hat{b}_3 - \sigma_1 \sigma_3 \hat{\kappa}^1(\rho_{12}^*) \rho_{13} - \sigma_2 \sigma_3 \hat{\kappa}^2(\rho_{12}^*) \rho_{23},$$

is linear in $(\rho_{13}, \rho_{23}) \in [\underline{\rho}_{13}, \bar{\rho}_{13}] \times [\underline{\rho}_{23}, \bar{\rho}_{23}]$, and has the same sign as $\hat{\kappa}^3(\rho_{12}^*, \rho_{13}, \rho_{23})$ due to the positiveness of $a(\rho_{12}^*, \rho_{13}, \rho_{23})$. To study the condition of $\kappa^3(\rho^*) = 0$ on Γ , we discuss it in the following two cases:

- (i) if $\bar{\rho}_{12} < \hat{\rho}_{12}$, then $\hat{\kappa}^1 \hat{\kappa}^2(\bar{\rho}_{12}) > 0$, the function $(\rho_{13}, \rho_{23}) \mapsto g(\bar{\rho}_{12}, \rho_{13}, \rho_{23})$ has the same monotonicity with respect to ρ_{13}, ρ_{23} . Therefore, to ensure that the function $g(\bar{\rho}_{12}, \rho_{13}, \rho_{23})$ has a root in $[\underline{\rho}_{13}, \bar{\rho}_{13}] \times [\underline{\rho}_{23}, \bar{\rho}_{23}]$, we need $g(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) g(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23}) \leq 0$, or equivalently $\hat{\kappa}^3(\bar{\rho}_{12}, \underline{\rho}_{13}, \underline{\rho}_{23}) \hat{\kappa}^3(\bar{\rho}_{12}, \bar{\rho}_{13}, \bar{\rho}_{23}) \leq 0$.
- (ii) if $\underline{\rho}_{12} > \hat{\rho}_{12}$, then $\hat{\kappa}^1 \hat{\kappa}^2(\underline{\rho}_{12}) < 0$, the function $(\rho_{13}, \rho_{23}) \mapsto g(\underline{\rho}_{12}, \rho_{13}, \rho_{23})$ has the opposite monotonicity with respect to ρ_{13}, ρ_{23} . Therefore, when $g(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) g(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23}) \leq 0$, or equivalently $\hat{\kappa}^3(\underline{\rho}_{12}, \bar{\rho}_{13}, \underline{\rho}_{23}) \hat{\kappa}^3(\underline{\rho}_{12}, \underline{\rho}_{13}, \bar{\rho}_{23}) \leq 0$, the function $g(\underline{\rho}_{12}, \rho_{13}, \rho_{23})$ has a root in $[\underline{\rho}_{13}, \bar{\rho}_{13}] \times [\underline{\rho}_{23}, \bar{\rho}_{23}]$.

Therefore, we deduce that $R(\hat{b}, \rho) \geq R(\hat{b}_{-3}, \rho_{12}) \geq R(\hat{b}_{-3}, \bar{\rho}_{12} 1_{\{\bar{\rho}_{12} < \hat{\rho}_{12}\}} + \underline{\rho}_{12} 1_{\{\underline{\rho}_{12} > \hat{\rho}_{12}\}})$ and that ‘=’ holds if and only if $\rho_{12}^* = \bar{\rho}_{12} 1_{\{\bar{\rho}_{12} < \hat{\rho}_{12}\}} + \underline{\rho}_{12} 1_{\{\underline{\rho}_{12} > \hat{\rho}_{12}\}}$ and ρ_{13}^*, ρ_{23}^* satisfies $\hat{\kappa}^3(\rho_{12}^*, \rho_{13}^*, \rho_{23}^*) = 0$. This corresponds to subcases **2.(i)** and **2.(ii)** of Theorem 4.3.

- 3.** $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*) = 0$, $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*) \neq 0$, $\hat{\kappa}^2(\rho^*) \hat{\kappa}^3(\rho^*) = 0$.

In this case, we make permutations as follows,

$$\begin{pmatrix} \hat{\kappa}^{-2}(\rho) \\ \hat{\kappa}^2(\rho) \end{pmatrix} = \begin{pmatrix} \Sigma_{-2}(\rho_{13}) & C_2 \\ C_2^\top & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{b}_{-2} \\ \hat{b}_2 \end{pmatrix}, \quad (\text{B.80})$$

where $\hat{\kappa}^{-2}(\rho) := (\hat{\kappa}^1(\rho), \hat{\kappa}^3(\rho))^\top$ and $C_2 := (\sigma_1 \sigma_2 \rho_{12}, \sigma_2 \sigma_3 \rho_{23})^\top$. Using (B.80) and proceeding with the same arguments as in the case **2.**, we obtain the result of $\hat{\kappa}^2(\rho^*) = 0$, $\hat{\kappa}^1(\rho^*) \hat{\kappa}^3(\rho^*) \neq 0$ as described in the subcases **3.(i)** and **3.(ii)** of Theorem 4.3.

- 4.** $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*) = 0$, $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*) = 0$, $\hat{\kappa}^2(\rho^*) \hat{\kappa}^3(\rho^*) \neq 0$.

Notice that

$$\begin{pmatrix} \hat{\kappa}^{-1}(\rho) \\ \hat{\kappa}^1(\rho) \end{pmatrix} = \begin{pmatrix} \Sigma_{-1}(\rho_{23}) & C_1 \\ C_1^\top & \sigma_1^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat{b}_{-1} \\ \hat{b}_1 \end{pmatrix}, \quad (\text{B.81})$$

where $\hat{\kappa}^{-1}(\rho) := (\hat{\kappa}^2(\rho), \hat{\kappa}^3(\rho))^\top$ and $C_1 := (\sigma_1 \sigma_2 \rho_{12}, \sigma_1 \sigma_3 \rho_{13})^\top$. Using (B.81) and proceeding with the same arguments as in the case **2.**, we obtain the result of $\hat{\kappa}^1(\rho^*) = 0$, $\hat{\kappa}^2(\rho^*) \hat{\kappa}^3(\rho^*) \neq 0$ as described in subcases **4.(i)** and **4.(ii)** of Theorem 4.3.

5. $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*) \neq 0$, $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*) \neq 0$, $\hat{\kappa}^2(\rho^*) \hat{\kappa}^3(\rho^*) \neq 0$.

In this case, we see from (B.69) that each ρ_{ij}^* takes value in $\{\underline{\rho}_{ij}, \bar{\rho}_{ij}\}$ relying on the sign of $\hat{\kappa}^i \hat{\kappa}^j(\rho^*)$. Notice that once the signs of $\hat{\kappa}^1 \hat{\kappa}^2(\rho^*)$ and $\hat{\kappa}^1 \hat{\kappa}^3(\rho^*)$ are known, the sign of $\hat{\kappa}^2(\rho^*) \hat{\kappa}^3(\rho^*)$ is determined. Therefore, by combination, there are 4 possible sub-cases as described in the case 5. of Theorem 4.3. As $\hat{\kappa}^i(\rho^*) \hat{\kappa}^j(\rho^*) \neq 0$ in each subcase, left right hand of (B.69) is strictly negative for any $\rho \in \Gamma \setminus \{\rho^*\}$. From the first-order characterization for convexity of $R(\hat{b}, \rho)$ (see e.g. Section 3.1.3 in Boyd & Vandenberghe (2004)) and (B.43), we obtain for any $\rho \in \Gamma \setminus \{\rho^*\}$,

$$\begin{aligned} R(\hat{b}, \rho) &\geq R(\hat{b}, \rho^*) + (\rho - \rho^*)^\top \nabla_\rho R(\hat{b}, \rho^*) \\ &= R(\hat{b}, \rho^*) - \sum_{j=1}^3 \sum_{i=1}^{j-1} \sigma_i \sigma_j \hat{\kappa}^i(\rho^*) \hat{\kappa}^j(\rho^*) (\rho_{ij} - \rho_{ij}^*) \\ &> R(\hat{b}, \rho^*), \end{aligned}$$

which indicates that ρ^* in each sub-case of case 5. in Theorem 4.3 is a strict minimum of $R(\hat{b}, \rho)$.

As $R(\hat{b}, \rho^*)$ in each subcase is strict minimum value, we conclude that each subcase in Theorem 4.3 is exclusive. By combining this with Lemma B.1, we obtain b^* described as in Theorem 4.3. \square

References

- Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton university press.
- Biagini, S., & Pinar, M. Ç. (2017). The robust Merton problem of an ambiguity averse investor. *Mathematics and Financial Economics*, 11(1), 1–24.
- Bielecki, T., Chen, T., & Cialenco. (2017). Recursive construction of confidence regions. *Electronic Journal of Statistics*, 11(2), 4674–4700.
- Bielecki, T. R., Chen, T., Cialenco, I., Cousin, A., & Jeanblanc, M. (2019). Adaptive robust control under model uncertainty. *SIAM Journal on Control and Optimization*, 57(2), 925–946.
- Björk, T., Khapko, M., & Murgoci, A. (2017). On time-inconsistent stochastic control in continuous time. *Finance and Stochastics*, 21(2), 331–360.
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Boyle, P., Garlappi, L., Uppal, R., & Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, 58(2), 253–272.
- Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5), 707–747.
- Chacko, G., & Viceira, L. M. (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, 116(1-2), 259–292.
- Cizeau, P., Potters, M., & Bouchaud, J.-P. (2001). Correlation structure of extreme stock returns. *Quantitative Finance*, 1(2), 217–222.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.

- Cooper, I., & Kaplanis, E. (1994). Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *The Review of Financial Studies*, 7(1), 45–60.
- Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3), 595–612.
- Denis, L., & Kervarec, M. (2013). Optimal investment under model uncertainty in nondominated models. *SIAM Journal on control and optimization*, 51(3), 1803–1822.
- Dhrymes, P. J. (1978). *Mathematics for econometrics* (Tech. Rep.). Springer.
- Ding, K. W., Chen, Z. Y., & Huang, N. J. (2018). Robust mean variance optimization problem under rényi divergence information. *Optimization*, 67(2), 287–307.
- Dow, J., & da Costa Werlang, S. R. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica: Journal of the Econometric Society*, 197–204.
- El Karoui, N. (1981). Les aspects probabilistes du contrôle stochastique. *École d'été de Probabilités de Saint-Flour IX-1979*, 73–238.
- Fischer, M., & Livieri, G. (2016). Continuous time mean-variance portfolio optimization through the mean field approach. *ESAIM: Probability and Statistics*, 20, 30–44.
- Fouque, J.-P., Pun, C. S., & Wong, H. Y. (2016). Portfolio optimization with ambiguous correlation and stochastic volatilities. *SIAM Journal on Control and Optimization*, 54(5), 2309–2338.
- French, K. R., & Poterba, J. M. (1991). Investor diversification and international equity markets. *The American Economic Review*, 81(2), 222–226.
- Fryszkowski, A. (2004). *Fixed point theory for decomposable sets* (Vol. 2). Springer Science & Business Media.
- Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1), 41–81.
- Ghaoui, L. E., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research*, 51(4), 543–556.
- Glasserman, P., & Xu, X. (2013). Robust portfolio control with stochastic factor dynamics. *Operations Research*, 61(4), 874–893.
- Goh, J., & Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1), 902–917.
- Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of operations research*, 28(1), 1–38.
- Guidolin, M., & Liu, H. (2016). Ambiguity aversion and underdiversification. *Journal of Financial and Quantitative Analysis*, 1297–1323.
- Hanasusanto, G. A., & Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 66(3), 849–869.
- Hu, Y., Imkeller, P., & Müller, M. (2005). Utility maximization in incomplete markets. *The Annals of Applied Probability*, 15(3), 1691–1712.

- Hu, Y., Jin, H., & Zhou, X. Y. (2012). Time-inconsistent stochastic linear–quadratic control. *SIAM journal on Control and Optimization*, 50(3), 1548–1572.
- Huang, H. H., Zhang, S., & Zhu, W. (2017). Limited participation under ambiguity of correlation. *Journal of Financial Markets*, 32, 97–143.
- Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, 14(3), 659–680.
- Ismail, A., & Pham, H. (2019). Robust markowitz mean-variance portfolio selection under ambiguous covariance matrix. *Mathematical Finance*, 29(1), 174–207.
- Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4), 1651–1683.
- Jiang, J., & Tian, W. (2016). Correlation uncertainty, heterogeneous beliefs and asset prices. *mimeo, University of North Carolina*.
- Jiang, R., & Guan, Y. (2018). Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research*, 66(5), 1390–1405.
- Jin, H., & Zhou, X. Y. (2015). Continuous-time portfolio selection under ambiguity. *Mathematical Control & Related Fields*, 5(3), 475.
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2), 365–411.
- Li, D., & Ng, W. L. (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical finance*, 10(3), 387–406.
- Li, D., & Zhou, X. Y. (2000). Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Applied Mathematics and Optimization*, 42(1), 19–33.
- Liechty, J. C., Liechty, M. W., & Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, 91(1), 1–14.
- Lin, Q., & Riedel, F. (2014). Optimal consumption and portfolio choice with ambiguity. *arXiv preprint arXiv:1401.1639*.
- Liu, J., & Zeng, X. (2017). Correlation ambiguity and under-diversification. *preprint, available at SSRN: <https://ssrn.com/abstract=2692692>*.
- Markowitz, H. (1959). *Portfolio selection* (Vol. 7). Yale University Press New Haven.
- Matoussi, A., Possamai, D., & Zhou, C. (2015). Robust utility maximization in nondominated models with 2bsde: the uncertain volatility model. *Mathematical Finance*, 25(2), 258–287.
- Mitton, T., & Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *The Review of Financial Studies*, 20(4), 1255–1288.
- Natarajan, K., Pachamanova, D., & Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3), 573–585.
- Natarajan, K., Sim, M., & Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(4), 695–731.

- Neufeld, A., & Nutz, M. (2018). Robust utility maximization with lévy processes. *Mathematical Finance*, 28(1), 82–105.
- Pham, H. (2009). *Continuous-time stochastic control and optimization with financial applications* (Vol. 61). Springer Science & Business Media.
- Pham, H., & Wei, X. (2017). Dynamic programming for optimal control of stochastic McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(2), 1069–1101.
- Rapisarda, F., Brigo, D., & Mercurio, F. (2007). Parameterizing correlations: a geometric interpretation. *IMA Journal of Management Mathematics*, 18(1), 55–73.
- Rouge, R., & El Karoui, N. (2000). Pricing via utility maximization and entropy. *Mathematical Finance*, 10(2), 259–276.
- Schied, A. (2007). Optimal investments for risk-and ambiguity-averse preferences: a duality approach. *Finance and Stochastics*, 11(1), 107–129.
- Schultz, E. (1996). Workers put too much in their employer’s stock. *Wall Street Journal*, 1.
- Strasser, H. (2011). *Mathematical theory of statistics: statistical experiments and asymptotic decision theory* (Vol. 7). Walter de Gruyter.
- Uppal, R., & Wang, T. (2003). Model misspecification and underdiversification. *The Journal of Finance*, 58(6), 2465–2486.
- Van Nieuwerburgh, S., & Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2), 779–805.
- Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6), 1358–1376.
- Yang, Z., Liang, G., & Zhou, C. (2019). Constrained portfolio-consumption strategies with uncertain parameters and borrowing costs. *Mathematics and Financial Economics*, 13(3), 393–427.