Ann Oper Res (2010) 176: 191–220
DOI 10.1007/s10479-009-0515-6

Robust portfolios: contributions from operations research and finance

Frank J. Fabozzi · Dashan Huang · Guofu Zhou

Published online: 29 January 2009

Abstract In this paper we provide a survey of recent contributions to robust portfolio strategies from operations research and finance to the theory of portfolio selection. Our survey covers results derived not only in terms of the standard mean-variance objective, but also in terms of two of the most popular risk measures, mean-VaR and mean-CVaR developed recently. In addition, we review optimal estimation methods and Bayesian robust approaches.

Keywords Robust portfolio · Mean-variance · Mean-VaR · Mean-CVaR · Parameter uncertainty · Model uncertainty

1 Introduction

Since Markowitz's (1952) seminal work, the mean-variance framework has become the major model used in practice today in asset allocation and active portfolio management despite many sophisticated models developed by academics. Grinold and Kahn (2000), Litterman (2003), and Meucci (2005) are important monographs detailing the practical applications of the mean-variance framework. Because of its enormous theoretical and practical importance, a substantial body of literature has developed under the mean-variance framework, and it is not surprising to find several insightful reviews that have been written on the related portfolio selection problem. To mention a few, Steinbach (2001) provides an extensive review of the interplay between objectives and constraints with different single-period variants

F.J. Fabozzi (✉)
School of Management, Yale University, New Haven, CT 06520, USA
e-mail: frank.fabozzi@yale.edu

D. Huang · G. Zhou
Olin School of Business, Washington University, St. Louis, MO 63130, USA

D. Huang
e-mail: huangdas@wustl.edu

G. Zhou
e-mail: zhou@wustl.edu

in the mean-variance framework. Brandt (2004) details the various approaches in the finance literature to portfolio choice. Konno and Koshizuka (2005) survey the mean-absolute deviation model and its relation to the capital asset pricing model, as well as its application to asset-liability management and mortgage-backed securities portfolio optimization.

In this paper we provide a survey of recent contributions from operations research and finance to the theory of robust portfolio selection. In contrast to existing surveys, our paper focuses on one of the most rapid and important areas, the construction of robust portfolios. We survey robust portfolio strategies not only in terms of the standard mean-variance objective, but also in terms of two of the most popular measures that have been recently developed: mean-VaR (value-at-risk) and mean-CVaR (conditional value-at-risk). We provide a unified framework to understand the various advances. In addition, we survey optimal estimation methods and Bayesian robust approaches.[1]

On the basis of return-risk tradeoff, we address two aspects of portfolio selection problem: 1) what is a proper risk measure that can adapt to the vicissitudes of the financial market when vagaries occur more often than ever; and 2) how to make this proper risk measure work well in practice. In particular, we discuss in some detail two downside risk measures, VaR and CVaR, that have been proposed to deal with the negative effects of extreme portfolio downside movements. Although VaR has become the industry benchmark for risk management due to its intuitive appeal and acceptance by bank regulators (Jorion 2006), CVaR, first used in the insurance industry (Embrechts et al. 1997), has better theoretical properties as a risk measure than VaR, and hence can be regarded as a good supplement. On the other hand, all these models, no matter mean-variance, mean-VaR, and mean-CVaR, require parameter inputs, that usually suffer from data inadequacy, contamination, or estimation error, and hence lead to counterintuitive portfolios. There is a long history of how to improve the parameter estimation and a plethora of methods have been attempted.

Since Markowitz (1952) formulates the idea of diversification of investments by incorporating into portfolio selection the variance as a risk measure that can be reduced (but not necessarily eliminated) without changing expected portfolio return, the mean-variance framework is so intuitive and powerful that it has been continually applied to different areas within finance and risk management. Indeed, numerous innovations within finance have either been an application of the concept of mean-variance analysis or an extension of the methodology to alternative portfolio risk measures (Fabozzi et al. 2002). Unfortunately, the omnipotence of mean-variance in theory does not make its employment in practice as ubiquitous as expected.

A critical weakness of mean-variance analysis is the use of variance as a measure of risk. In some sense, risk is a subjective concept and different investors adopt different investment strategies in seeking to realize their investment objectives (Holton 1997; Siu et al. 2001; and Boyle et al. 2002), and hence the exogenous characteristics of investors mean that probably no unique risk measure exists that can accommodate every investor's problem (Ortobelli et al. 2005). For example, one investor may be concerned about dramatic market fluctuation no matter whether this movement is upside or downside, whereas another investor may be more concerned with the downside movements, which usually imply severe loss consequences. In this case, the variance is obviously not sufficient to express or measure the investors' risk. On the other hand, the market may change in nature. The introduction of new derivatives and new investment strategies may require the formulation of an alternative risk measure more

[1] We do not cover in this survey the treatment of higher-order moments (skewness and kurtosis) and loss-aversion preferences. Models for dealing with these two topics are provided by Gul (1991), Harvey et al. (2003), Ang et al. (2005), Hong et al. (2007), and Guidolin and Timmermann (2008).

appropriate for different investors. This is because the portfolio distribution with derivatives such as futures and options is skewed and heavy-tailed, which calls for a risk measure to respond to downside deviation and upside deviation asymmetrically (Siu et al. 2001; Boyle et al. 2002).

The importance of measuring the downside risk of a portfolio rather than variance has long been recognized by academics and practitioners. This can be traced back to Roy (1952) who proposed a “safety first” strategy to maximize portfolio expected return subject to a downside risk constraint. Probably due to computational difficulties, this downside risk measure failed to be adopted extensively in the industry until J.P. Morgan’s decision in 1994 to make its RiskMetrics publicly available. In RiskMetrics, the concept of VaR was first explicitly proposed to materialize the capital adequacy and market risk of commercial banks.[2] Two other reasons that sparked the greater use of VaR were 1) the fact that financial disasters were occurring with greater frequency than prior periods (e.g. Orange County, Barings, and Long-Term Capital Management); and 2) central bank regulators adopted VaR to calculate a bank’s required capital.[3]

Due to its conceptual simplicity, VaR has become a very popular risk management tool and an industry benchmark in many different types of organizations (Jorion 2006). However, VaR has important theoretical limitations. For example, it is not a coherent risk measure in general and it ignores extreme losses beyond itself (Artzner et al. 1999). To overcome these hurdles, Rockafellar and Uryasev (2000, 2002) propose an alternative risk measure to VaR, CVaR. CVaR has tractable properties: it is a coherent risk measure, it is easy to implement, and it takes into consideration the entire tail that exceeds VaR on average.

There is broad consensus in the investment community as to what the challenge of investment theory is today. Merton (2003) points out that the challenge is to put “the rich set of tools” into practice, and concludes that “I see this as a tough engineering problem, not one of new science”. Indeed, Merton (2003) addresses an ongoing problem since the portfolio selection problem was first formulated by Markowitz in 1952 of how to make the theoretical model do a good job in practice. One major obstacle to successful implementation and industry adoption of the mean-variance strategy in practice is that the “optimal” portfolios resulting from these portfolio optimization models are extremely sensitive to the inputs, which are typically estimated from historical data, and often prone to be imprecise. As a result, the “optimal” portfolios frequently have extreme or counterintuitive weights for some assets (Best and Grauer 1991; Broadie 1993; Chopra and Ziemba 1993). This obstacle also applies to strategies using risk measures other than variance such as VaR and CVaR (El Ghaoui et al. 2003; Ceria and Stubbs 2006; Zhu and Fukushima 2008; Huang et al. 2008).

There are two standard methods extensively adopted in the literature to deal with estimation errors. The first is the robust estimation methods, such as moment estimation, which can be quite robust to distributional assumptions. The second is the Bayesian approach that is neutral to uncertainty in the sense of Knight (1921) because it assumes a single prior on the portfolio distribution (Garlappi et al. 2007). An extension of

[2] Prior to 1994, a more elegant theoretical risk measure than safety first, the lower partial moment measure, was proposed by Bawa (1975). This measure can accommodate a significant number of the known Von Neumann-Morgenstern utility functions, from risk loving or engaging to risk neutral to risk aversion.

[3] VaR can be informally explained in this way: suppose that we believe with β ($= 95\%$ or 99%) certainty that we will not lose more than V dollars in the next N ($= 1$ or 10) days. The variable V is the VaR. In defining a bank’s required capital, typically regulators use $N = 10$ and $\beta = 99\%$. The required capital is therefore the losses over this 10-day period that are expected to happen 1% of the time.

this is the multiple prior approach. This robust technique has obtained prodigious success since the late 1990s, especially in the field of optimization and control with uncertainty parameters (Ben-Tal and Nemirovski 1998, 1999; El Ghaoui and Lebret 1997; Goldfarb and Iyengar 2003a). With respect to portfolio selection, the major contributions have come in the 21st century (see, for example, Rustem et al. 2000; Costa and Paiva 2002; Ben-Tal et al. 2002; Goldfarb and Iyengar 2003b; El Ghaoui et al. 2003; Tütüncü and Koenig 2004; Pinar and Tütüncü 2005; Lutgens and Schotman 2006; Natarajan et al. 2009; Garlappi et al. 2007; Pinar 2007; Calafiore 2007; Huang et al. 2008; Natarajan et al. 2008a; Brown and Sim 2008; Natarajan et al. 2008b; Shen and Zhang 2008; Elliott and Siu 2008; Zhu and Fukushima 2008). For a complete discussion of robust portfolio management and the associated solution methods, see Fabozzi et al. (2007), Föllmer et al. (2008), and the references therein.

The paper is organized as follows. Section 2 reviews the mean-VaR and mean-CVaR models for portfolio selection, as well as a brief introduction of mean-variance analysis. Section 3 surveys the application of robust optimization techniques in portfolio selection. Section 4 discusses how to improve the robustness of portfolio selection by utilizing appropriate statistical estimators. Section 5 reviews the recent advances of Bayesian approach in portfolio selection, which is followed by a simple conclusion on future directions (Sect. 6).

2 The models

In this section, we review first the classic mean-variance framework of Markowitz (1952, 1959). Portfolio optimization models based on alternative risk measures have been introduced in the literature such as semivariance (Markowitz 1959), lower partial moment (Bawa 1975), mean absolute deviation (Konno and Yamazaki 1991), minimax (Young 1998), VaR (Ahn et al. 1999; Basak and Shapiro 2001), CVaR (Rockafellar and Uryasev 2000), coherent risk measures (Artzner et al. 1999), convex risk measures (Föllmer and Schied 2002; Frittelli and Rosazza Gianin 2002), generalized deviation measures (Rockafellar et al. 2006), proper and ideal risk measures (Stoyanov et al. 2007; Rachev et al. 2008a), and have been widely applied in practice (Dembo and Rosen 2000; Ortobelli et al. 2005). Among these, we address in this section the minimization of VaR and CVaR in portfolio selection. Particularly, we explain in some detail the difference between both mean-VaR and mean-CVaR and mean-variance analysis, and how far away the efficient frontiers are from each other.

2.1 The classic mean-variance problem

We consider in this section a one-period portfolio selection problem. Let the random vector $\boldsymbol{r} = (r_1, \dots, r_n)^\top \in \Re^n$ denote random returns of the n risky assets, and $\boldsymbol{x} = (x_1, \dots, x_n)^\top \in X = \{\boldsymbol{x} \in \Re^n : \mathbf{1}_n^\top \boldsymbol{x} = 1\}$ denote the proportion of the portfolio to be invested in the n risky assets, where $^\top$ means transposition and $\mathbf{1}_n$ denotes a vector of all ones. Suppose that $\boldsymbol{r}$ has a probability distribution $p(\boldsymbol{r})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.[4] Then the target of the investor is to choose an optimal portfolio $\boldsymbol{x}$ that lies on the mean-risk efficient frontier. In the Markowitz model, the "mean" of a portfolio is defined as the expected value of the portfolio return, $\boldsymbol{\mu}^\top \boldsymbol{x}$, and the "risk" is defined as the variance of the portfolio return, $\boldsymbol{x}^\top \Sigma \boldsymbol{x}$.

[4] Later we may denote a random variable and its related deterministic variable/constant as the same symbol since they can be distinguished clearly by the context.

Mathematically, minimizing the variance subject to target and budget constraints leads to a formulation like:

$$\min_{\boldsymbol{x}}\left\{\boldsymbol{x}^\top\Sigma\boldsymbol{x}:\ \boldsymbol{\mu}^\top\boldsymbol{x}\geq\mu_0,\ \mathbf{1}_n^\top\boldsymbol{x}=1\right\},\tag{1}$$

where μ_0 is the minimum expected return. There are two implicit assumptions in this formulation: 1) the first two moments of portfolio return exist and 2) the initial wealth is normalized to be 1 without loss of generality.

If the moment parameters are known, the analytical solution to the above formulation is straightforward to apply, and the above problem can be solved numerically under various practical constraints, such as no-short-selling or position limits. However, the moment parameters are never known in practice and they have to be estimated from an unknown distribution with limited data. Typically, the procedure of minimizing the portfolio variance with a given expected return can be decomposed into three steps: 1) estimate the expected return and covariance, 2) use the above optimization problem to create an efficient frontier, and 3) select a point on the efficient frontier or select a mix of the risk-free asset and the optimal risky asset allocation according to the investor's risk tolerance. This procedure is clearly not optimal, and hence robust procedures for making good use of portfolio theory are called for in the presence of parameter or model uncertainties or both.

2.2 Mean-VaR model

Here we consider the mean-VaR model for portfolio selection. Following Rockafellar and Uryasev (2000), we can define the portfolio loss as the minus return, $-\boldsymbol{r}^\top\boldsymbol{x}$. For the purpose of exposition, we assume that the distribution of $\boldsymbol{r}$ is continuous.

For a given portfolio $\boldsymbol{x}\in X$, the probability of the loss not exceeding a threshold α is given by $\Psi(\boldsymbol{x},\alpha)=\int_{-\boldsymbol{r}^\top\boldsymbol{x}\leq\alpha}p(\boldsymbol{r})d\boldsymbol{r}$. Given a confidence level β, the VaR associated with the portfolio $\boldsymbol{x}$ is defined as $\mathrm{VaR}_\beta(\boldsymbol{x})=\min\{\alpha\in\Re\ :\ \Psi(\boldsymbol{x},\alpha)\geq\beta\}$. When the portfolio return distribution is Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix Σ, then the VaR can be expressed as

$$\mathrm{VaR}_\beta(\boldsymbol{x})=\zeta_\beta\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}}-\boldsymbol{\mu}^\top\boldsymbol{x},\tag{2}$$

where $\zeta_\beta=-\Phi^{-1}(1-\beta)$ and $\Phi(\cdot)$ is the standard normal distribution function. From (2), the computation of VaR involves three components (the mean, the variance, and the distribution of the portfolio return), which in turn implies that an investor who regards VaR as the risk measure can adjust ζ_β and $\boldsymbol{\mu}^\top\boldsymbol{x}$ to reduce the portfolio risk. In this sense the VaR measure is more flexible than the variance measure.

Alexander and Baptista (2002, 2004) point out that, for any portfolio $\boldsymbol{x}$, when $\beta\to1/2$, $\zeta_\beta\to0$, and hence $\lim_{\beta\to1/2}\mathrm{VaR}(\boldsymbol{x})=-\boldsymbol{\mu}^\top\boldsymbol{x}$. This indicates that a risk-neutral investor can choose a confidence level of 50% to maximize the expected return in the mean-VaR framework. On the other hand, if $\beta\to1$, $\zeta_\beta\to+\infty$, then $\lim_{\beta\to1}\mathrm{VaR}(\boldsymbol{x})/\zeta_\beta=\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}}$. In this case, minimizing VaR over $\boldsymbol{x}\in X$ is equivalent to minimizing the variance.

Under regular condition that $0.5<\beta<1$, $\zeta_\beta>0$, then the following optimization problem

$$\min_{\boldsymbol{x}}\left\{\zeta_\beta\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}}-\boldsymbol{\mu}^\top\boldsymbol{x}:\mathbf{1}_n^\top\boldsymbol{x}=1\right\}$$

is convex and there exists a real number μ_0 such that the above problem is equivalent to

$$\min_{\boldsymbol{x}}\left\{\boldsymbol{x}^\top\Sigma\boldsymbol{x} : \boldsymbol{\mu}^\top\boldsymbol{x}\geq\mu_0, \mathbf{1}_n^\top\boldsymbol{x}=1\right\},$$

where μ_0 is a function of ζ_β. Up to this point, the problem

$$\min_{\boldsymbol{x}}\left\{\mathrm{VaR}_\beta(\boldsymbol{x}) \;:\; \mathbf{1}_n^\top\boldsymbol{x}=1\right\}$$

is equivalent to the mean-variance optimization problem

$$\min_{\boldsymbol{x}}\left\{\boldsymbol{x}^\top\Sigma\boldsymbol{x} \;:\; \boldsymbol{\mu}^\top\boldsymbol{x}\geq\mu_0, \mathbf{1}_n^\top\boldsymbol{x}=1\right\}$$

in the sense that the mean-VaR boundary coincides with the mean-variance boundary. More specifically, a portfolio lies on the mean-VaR boundary if and only if it lies on the mean-variance boundary. Denote $a=\mathbf{1}_n^\top\Sigma^{-1}\boldsymbol{\mu}$, $b=\boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}$, $c=\mathbf{1}_n^\top\Sigma^{-1}\mathbf{1}_n$, and $d=bc-a^2$. Alexander and Baptista (2002) show that the minimum VaR portfolio at the β confidence level exists if and only if $\zeta_\beta>\sqrt{d/c}$. Moreover, this minimum VaR portfolio is mean-variance efficient, and can be given by

$$\boldsymbol{x}^*_{\mathrm{VaR}}=g+h\left[\mathbb{E}[r_{\boldsymbol{x}^*_\sigma}]+\sigma[r_{\boldsymbol{x}^*_\sigma}]\sqrt{\frac{d^2/c^2}{(\zeta_\beta)^2-d/c}}\;\right],$$

which yields the optimal VaR as

$$\mathrm{VaR}_\beta(\boldsymbol{x}^*_{\mathrm{VaR}})=\sigma[r_{\boldsymbol{x}^*_\sigma}]\sqrt{(\zeta_\beta)^2-d/c}-\mathbb{E}[r_{\boldsymbol{x}^*_\sigma}],$$

where $g=(1/d)[b(\Sigma^{-1}\mathbf{1}_n)-a(\Sigma^{-1}\boldsymbol{\mu})]$, $h=(1/d)[c(\Sigma^{-1}\boldsymbol{\mu})-a(\Sigma^{-1}\mathbf{1}_n)]$, $r_{\boldsymbol{x}^*_\sigma}$ is the return of the minimum variance portfolio $\boldsymbol{x}^*_\sigma$, and $\sigma[r_{\boldsymbol{x}^*_\sigma}]$ is its corresponding standard deviation.

2.3 Mean-CVaR model

Motivated by the theoretical limitations of VaR, Rockafellar and Uryasev (2000) propose an alternative risk measure, CVaR, that is defined as the conditional expectation of the loss of the portfolio exceeding or equal to VaR,[5] that is,

$$\mathrm{CVaR}_\beta(\boldsymbol{x})=\frac{1}{1-\beta}\int_{-\boldsymbol{r}^\top\boldsymbol{x}\geq\mathrm{VaR}_\beta(\boldsymbol{x})}-\boldsymbol{r}^\top\boldsymbol{x}\,p(\boldsymbol{r})dr.$$

Moreover, Rockafellar and Uryasev (2000) prove that CVaR is sub-additive and can be casted into the following convex optimization problem: $\mathrm{CVaR}_\beta(\boldsymbol{x})=\min_{\alpha\in\Re}F_\beta(\boldsymbol{x},\alpha)$, where $F_\beta(\boldsymbol{x},\alpha)$ is expressed as

$$F_\beta(\boldsymbol{x},\alpha)=\alpha+\frac{1}{1-\beta}\int_{\boldsymbol{r}\in\Re^n}[-\boldsymbol{r}^\top\boldsymbol{x}-\alpha]^+p(\boldsymbol{r})d\boldsymbol{r}, \tag{3}$$

[5]With more delicate assumptions on portfolio returns, CVaR is also called Extreme Value Theory VaR, Mean-Excess Loss, Mean Shortfall, Tail VaR, Expected Shortfall, or Conditional Tail Expectation (Embrechts et al. 1997; Artzner et al. 1999).

where $[\cdot]^+$ is defined as $[t]^+ = \max\{0, t\}$ for any $t \in \Re$. Thus, minimizing CVaR over $\boldsymbol{x} \in X$ is equivalent to minimizing $F_\beta(\boldsymbol{x}, \alpha)$ over $(\boldsymbol{x}, \alpha) \in X \times \Re$, i.e.,

$$\min_{\boldsymbol{x}\in X} \mathrm{CVaR}_\beta(\boldsymbol{x}) = \min_{(\boldsymbol{x},\alpha)\in X\times\Re} F_\beta(\boldsymbol{x}, \alpha),$$

which implies that a pair $(\boldsymbol{x}^*, \alpha^*)$ solves $\min_{(\boldsymbol{x},\alpha)\in X\times\Re} F_\beta(\boldsymbol{x}, \alpha)$ if and only if $\boldsymbol{x}^*$ solves $\min_{\boldsymbol{x}\in X} \mathrm{CVaR}_\beta(\boldsymbol{x})$. Moreover, because $F_\beta(\boldsymbol{x}, \alpha)$ is convex with respect to $(\boldsymbol{x}, \alpha)$ and $\mathrm{CVaR}_\beta(\boldsymbol{x})$ is convex with respect to $\boldsymbol{x}$, the joint minimization is a convex programming problem, which can be efficiently solved by an interior algorithm.

In particular, Rockafellar and Uryasev (2000) show that, under the mild assumption of $\beta > 0.5$ and normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ, CVaR reduces to

$$\mathrm{CVaR}_\beta(\boldsymbol{x}) = \kappa_\beta\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}} - \boldsymbol{\mu}^\top\boldsymbol{x},$$

where $\kappa_\beta = \frac{-\int_{-\infty}^{\Phi^{-1}(1-\beta)} t\phi(t)dt}{1-\beta}$ and $\phi(\cdot)$ is the standard normal density function. Recall that the parameter ζ_β of VaR in the case of a normal distribution, $\kappa_\beta > \zeta_\beta$ when $1/2 < \beta < 1$, we have $\mathrm{CVaR}_\beta(\boldsymbol{x}) > \mathrm{VaR}_\beta(\boldsymbol{x})$. Hence, if the minimum VaR portfolio at the β confidence level exists, then it is mean-CVaR efficient at the β confidence level.

In the more general case where the distribution is not necessarily normal, to deal with the calculation of the integral of the multivariate and non-smooth function in (3), Rockafellar and Uryasev (2000) present an approximation approach via sampling method:

$$F_\beta(\boldsymbol{x}, \alpha) \approx \alpha + \frac{1}{T(1-\beta)}\sum_{t=1}^{T}[-(\boldsymbol{r}^t)^\top\boldsymbol{x} - \alpha]^+,$$

where T denotes the number of samples with respect to the portfolio distribution and $\boldsymbol{r}^t$ denotes the t-th sample.

Finally, we are in a position to present an interesting conclusion found by Alexander and Baptista (2004) that in the case of the normal distribution, the mean-VaR, mean-CVaR, and mean-variance efficient frontiers have the following relationship: 1) If $\kappa_\beta \le \sqrt{d/c}$, then the mean-VaR and mean-CVaR efficient frontiers are empty; 2) If $\zeta_\beta \le \sqrt{d/c} < \kappa_\beta$, then the mean-VaR efficient frontier is empty but the mean-CVaR efficient frontier is a nonempty proper subset of the mean-variance efficient frontier; and 3) If $\zeta_\beta > \sqrt{d/c}$, then a portfolio belongs to the mean-VaR efficient frontier if and only if it belongs to the mean-CVaR efficient frontier and $\mathbb{E}[r_{\boldsymbol{x}^*_{\mathrm{VaR}}}] > \mathbb{E}[r_{\boldsymbol{x}^*_{\mathrm{CVaR}}}] > \mathbb{E}[r_{\boldsymbol{x}^*_\sigma}]$, where c and d are defined in the VaR section.

3 Robust portfolios

In this section, we survey some recent advances in portfolio selection with parameter uncertainty. Suppose $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ represent the true and estimated input parameters in a portfolio selection model, respectively. For example, $\boldsymbol{\theta}$ denotes the mean $\boldsymbol{\mu}$ and the covariance Σ in the mean-variance model, whereas in the mean-VaR or mean-CVaR model, it represents the distribution of portfolio return. Typically, $\boldsymbol{\theta}$ is unobservable but is believed to belong to a certain set $\mathcal{P}$ which is generated from the estimated parameter $\hat{\boldsymbol{\theta}}$, i.e., $\boldsymbol{\theta} \in \mathcal{P} = \mathcal{P}_{\hat{\boldsymbol{\theta}}}$ (we will discuss in some detail how to generate $\mathcal{P}_{\hat{\theta}}$ from real data later). We aim at constructing a portfolio so that the risk is as small as possible with respect to the worst-case scenario of the uncertain parameters in this set $\mathcal{P}$.

3.1 Portfolio with known moments

We consider a general portfolio optimization model where the investor seeks to maximize the expectation of his utility $u(\cdot)$. The investor solves the following general stochastic mathematical program:

$$\max_{\boldsymbol{x}\in X} \mathbb{E}[u(\boldsymbol{r}^\top \boldsymbol{x})], \tag{4}$$

where $X = \{\boldsymbol{x} \in \Re^n : \mathbf{1}_n^\top \boldsymbol{x} = 1\}$.[6]

When the distribution of the portfolio return $\boldsymbol{r}$ is exactly known, problem (4) is a general one-stage stochastic optimization problem without recourse. Particularly, if $u(\cdot)$ is a quadratic utility $u(r) = c_2 r^2 + c_1 r + c_0$, problem (4) does not depend on the actual distribution of $\boldsymbol{r}$, except its mean $\boldsymbol{\mu}$ and covariance Σ, and amounts to: $\max_{\boldsymbol{x}\in X} c_2 \boldsymbol{x}^\top(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top)\boldsymbol{x} + c_1 \boldsymbol{\mu}^\top \boldsymbol{x} + c_0$. This only applies to the case of quadratic utility functions, partially explaining why it is particularly favored in economics and finance models, especially for portfolio selection.

Instead, in the general case where $u(\cdot)$ is not quadratic and the distribution $p(\boldsymbol{r})$ is partially known, we formulate the portfolio selection problem with robust optimization technique as follows:

$$\max_{\boldsymbol{x}\in X} \min_{\boldsymbol{r}\sim(\boldsymbol{\mu},\Sigma)} \mathbb{E}[u(\boldsymbol{r}^\top \boldsymbol{x})], \tag{5}$$

where $\boldsymbol{r} \sim (\boldsymbol{\mu}, \Sigma)$ indicates the fact that the distribution of $\boldsymbol{r}$ belongs to the class $\mathscr{P}_{(\boldsymbol{\mu},\Sigma)}$ of n-variate distributions with mean $\boldsymbol{\mu}$ and covariance Σ.

For a given vector $\boldsymbol{x}$, denote the inner optimization problem of (5) as

$$U(\boldsymbol{x}) = \min_{\boldsymbol{r}\sim(\boldsymbol{\mu},\Sigma)} \mathbb{E}[u(\boldsymbol{r}^\top \boldsymbol{x})]. \tag{6}$$

Popescu (2007) shows that problem (6) is equivalent to an optimization problem of univariate distributions with a given mean and variance; that is,

$$U(\boldsymbol{x}) = \min_{\boldsymbol{r}\sim(\boldsymbol{\mu},\Sigma)} \mathbb{E}[u(\boldsymbol{r}^\top \boldsymbol{x})] = \min_{r_{\boldsymbol{x}}\sim(\mu_{\boldsymbol{x}},\sigma_{\boldsymbol{x}}^2)} \mathbb{E}[u(r_{\boldsymbol{x}})],$$

which can be further reduced to optimizing over a restricted class of distribution with at most two support points if u is differentiable such that the first-order derivative of u is inverse S-shaped with finite limits at $\pm\infty$. That is, the objective $U(\boldsymbol{x})$ can be casted as a deterministic optimization problem with at most two variables, reducing the robust optimization problem (5) to a deterministic bi-criteria problem that only depends on the mean and variance of the random variable $\boldsymbol{r}^\top \boldsymbol{x}$.[7] More specifically, if the robust objective function

[6] We abuse *max* and *min* in the wide sense of *sup* and *inf*, respectively, to account for the case when the optimum is not attained.

[7] A function f defined on $\Re$ is convex-concave if there exists $x_0 \in \Re$ such that f is convex on $(-\infty, x_0)$ and concave on (x_0, ∞). A function f defined on $\Re$ is concave-convex if $-f$ is convex-concave. A function f is S-shaped if it is increasing and convex-concave. A function f is inverse S-shaped if $-f$ is S-shaped. A function u satisfies the two-point support property with respect to (μ, σ^2) if there exists a quadratic function that supports u from below and intersects it at two points a, b such that a feasible (μ, σ^2) distribution with support $\{a, b\}$ exists. A function u satisfies the two-point support property if it satisfies the property with respect to any (μ, σ^2).

$U(\boldsymbol{x}) = \min_{r_{\boldsymbol{x}} \sim (\mu_{\boldsymbol{x}}, \sigma_{\boldsymbol{x}}^2)} \mathbb{E}[u(r_{\boldsymbol{x}})]$ is continuous, non-decreasing in $\mu_{\boldsymbol{x}}$, non-increasing in $\sigma_{\boldsymbol{x}}$ and quasi-concave, then problem (5) over X is equivalent to solving the following parametric quadratic program:

$$\max_{\boldsymbol{x} \in X} \gamma \boldsymbol{\mu}^\top \boldsymbol{x} - (1-\gamma) \boldsymbol{x}^\top \Sigma \boldsymbol{x}. \tag{7}$$

Thus, if $\boldsymbol{x}^*(\gamma)$ denotes the optimal solution of problem (7), then the function $U(\boldsymbol{x}^*(\gamma))$ is continuous and unimodal in $\gamma \in [0, 1]$. Moreover, if γ^* is its maximum, then $\boldsymbol{x}^*(\gamma^*)$ is optimal for problem (5). With simple formulation, problem (7) can be written as (1) with $\mu_0 = \mu_0(\gamma)$.

Problem (7) implies that even if the distribution/input is unknown, under mild assumptions, we can still transform the portfolio selection problem with expected utility maximization into a tractable quadratic problem without increasing the level of complexity relative to the conventional problem with known input.

3.2 Portfolio with unknown mean

In this part, we follow Garlappi et al. (2007) by discussing the robust version of the mean-variance portfolio problem when uncertainty is assumed to be present only in the expected return (Σ is known), i.e., $\boldsymbol{\theta} = \boldsymbol{\mu} \in \mathcal{P}_{\hat{\boldsymbol{\mu}}}$.

3.2.1 Box uncertainty on mean

The simplest choice for the uncertain set of $\boldsymbol{\mu}$ is box:

$$\mathcal{P}_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : |\mu_i - \hat{\mu}_i| \le \delta_i, i = 1, \ldots, n\}.$$

The δ_is could be related to some confidence interval around the estimated expected return. For example, if the individual return of the risky assets is normally distributed, then $\frac{\mu_i - \hat{\mu}_i}{\sigma_i/\sqrt{T_j}}$ follows a standard normal distribution, and a 95% confidence interval for μ_i can be obtained by setting $\delta_i = 1.96\sigma_i/\sqrt{T_i}$, where T_i is the sample size used in the estimation and σ_i is the standard deviation of asset i.

The robust portfolio problem can be formulated as

$$\min_{\boldsymbol{x} \in X} \left\{ \boldsymbol{x}^\top \Sigma \boldsymbol{x} : \min_{\boldsymbol{\mu}} \boldsymbol{\mu}^\top \boldsymbol{x} \ge \mu_0, |\mu_i - \hat{\mu}_i| \le \delta_i, i = 1, \ldots, n \right\},$$

which can be further formulated as

$$\min_{\boldsymbol{x} \in X} \left\{ \boldsymbol{x}^\top \Sigma \boldsymbol{x} : (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_\delta)^\top \boldsymbol{x} \ge \mu_0 \right\}, \tag{8}$$

where $\boldsymbol{\mu}_\delta = (\mathrm{sign}(x_1)\delta_1, \ldots, \mathrm{sign}(x_n)\delta_n)^\top$. Here $\mathrm{sign}(x)$ is the sign function equal to 1 if $x \ge 0$ and 0 otherwise.

There are two explanations for (8). First, $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_\delta$ can be regarded as a shrinkage estimator of the expectation of portfolio returns, i.e., constructing a robust portfolio for $\boldsymbol{\mu}$ from $\hat{\boldsymbol{\mu}}$ is equivalent to constructing a conventional portfolio from $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_\delta$. If the weight of asset i in the portfolio is negative, the expected return on this asset is increased, $\mu_i + \delta_i$, while if it is positive, the expected return on this asset is decreased, $\mu_i - \delta_i$.

Second, by using the equality, $x_i \mathrm{sign}(x_i)\delta_i = x_i \frac{x_i}{|x_i|}\delta_i = \frac{x_i}{\sqrt{|x_i|}}\delta_i \frac{x_i}{\sqrt{|x_i|}}$, we can rewrite the problem as $\min_{\boldsymbol{x}\in X}\{\hat{\boldsymbol{\mu}}^\top \boldsymbol{x} - \hat{\boldsymbol{x}}^\top \Delta \hat{\boldsymbol{x}} - \gamma \boldsymbol{x}^\top \Sigma \boldsymbol{x}\}$, where $\hat{\boldsymbol{x}} = (\frac{x_1}{\sqrt{|x_1|}}, \dots, \frac{x_n}{\sqrt{|x_n|}})^\top$, $\Delta = \mathrm{diag}(\delta_1, \dots, \delta_n)$, and $\gamma = \gamma(\mu_0)$. Interestingly, the term $\hat{\boldsymbol{x}}^\top \Delta \hat{\boldsymbol{x}}$ can be interpreted as a risk adjustment performed by an investor who is averse to estimation error, and the magnitude of this "risk-like" term is determined by the values of δ_is.

3.2.2 *Ellipsoidal uncertainty on mean*

Here we consider another uncertain set for $\boldsymbol{\mu}$:

$$\min_{\boldsymbol{x}\in X}\left\{\boldsymbol{x}^\top \Sigma \boldsymbol{x} : \min_{\boldsymbol{\mu}} \boldsymbol{\mu}^\top \boldsymbol{x} \geq \mu_0, (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \Sigma^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \chi^2\right\}, \tag{9}$$

where χ^2 is a non-negative number (χ^2 is related to a chosen quantile for chi-square distribution with n degrees of freedom if the portfolio returns are multivariate and Σ is known).

For any feasible $\boldsymbol{x} \in X$, let us look at the following optimization problem with respect to $\boldsymbol{\mu}$:

$$\min_{\boldsymbol{\mu}\in\Re^n}\left\{\boldsymbol{\mu}^\top \boldsymbol{x} : (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \Sigma^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \chi^2\right\}.$$

Apparently, the objective is linear, the constraint is quadratic, and hence we can use the necessary and sufficient Kuhn-Tucker condition to obtain its optimal solution

$$\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}} - \frac{\chi \Sigma \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \Sigma \boldsymbol{x}}}$$

with unique optimal objective value, $\hat{\boldsymbol{\mu}}^\top x - \chi \sqrt{\boldsymbol{x}^\top \Sigma \boldsymbol{x}}$. Therefore, problem (9) reduces to the following second-order cone program:

$$\min_{x\in X} -\hat{\boldsymbol{\mu}}^\top x + \chi \sqrt{\boldsymbol{x}^\top \Sigma \boldsymbol{x}} + \gamma \boldsymbol{x}^\top \Sigma \boldsymbol{x},$$

where $\gamma = \gamma(\mu_0)$. Since this is a convex programming problem, it is easy to show that there exists a $\tilde{\gamma} = \tilde{\gamma}(\mu_0, \chi) \geq \gamma$ such that the above problem is equivalent to

$$\min_{\boldsymbol{x}\in X} \; -\hat{\boldsymbol{\mu}}^\top \boldsymbol{x} + \tilde{\gamma} \boldsymbol{x}^\top \Sigma \boldsymbol{x}.$$

Similar to the box uncertainty, the investor who is averse to the estimation error becomes more conservative.

3.3 Portfolio with unknown mean and covariance

Now we extend our discussion to a more realistic situation where the covariance Σ is also subject to estimation error. Then, $\boldsymbol{\theta} = (\boldsymbol{\mu}, \Sigma) \in \mathcal{P}_{(\hat{\boldsymbol{\mu}}, \hat{\Sigma})}$.

A few different methods for modeling uncertainty in the covariance matrix are used in practice. Some are superimposed on top of factor models for returns (Goldfarb and Iyengar 2003b), while others consider confidence intervals for the individual covariance matrix entries (Tütüncü and Koenig 2004). Benefits for portfolio performance have been observed even when the uncertainty set is defined simply as a collection of several possible scenarios for the covariance matrix (Rustem et al. 2000; Costa and Paiva 2002).

3.3.1 Factor models

In the robust factor model, the portfolio return $\boldsymbol{r} \in \Re^n$ is given by

$$\boldsymbol{r} = \boldsymbol{\mu} + V^\top \boldsymbol{f} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu} \in \Re^n$ is the vector of mean returns, $\boldsymbol{f} \in \Re^m \sim \mathcal{N}(0, F)$ is the vector of random returns of the m $(< n)$ factors that drive the market, $V \in \Re^{m\times n}$ is the factor loading matrix, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, D)$ is the vector of residual returns. The mean return vector $\boldsymbol{\mu}$, the factor loading matrix V, the covariance matrices of the factor return vector $\boldsymbol{f}$, and the residual error vector $\boldsymbol{\epsilon}$ are known to lie within suitably defined uncertain sets.

In addition, we assume that the vector of residual return $\boldsymbol{\epsilon}$ is independent of the vector of factor returns $\boldsymbol{f}$, the covariance matrix $F \succ 0$ and the covariance matrix $D = \mathrm{diag}(\boldsymbol{d}) \succeq 0$, i.e., $d_i \geq 0, i = 1, \ldots, n$.[8] Thus, the portfolio return $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\mu}, V^\top F V + D)$, in which $\boldsymbol{\mu}$, D, and V are uncertain.

Without loss of generality, we assume the uncertain set $\mathcal{P}$ for $\boldsymbol{\mu}$, D, and V is separable, that is,

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, D, V) \in \mathcal{P}_{(\hat{\boldsymbol{\mu}}, \hat{D}, \hat{V})} = \mathcal{P}_{\hat{\boldsymbol{\mu}}} \times \mathcal{P}_{\hat{D}} \times \mathcal{P}_{\hat{V}}.$$

Following Goldfarb and Iyengar (2003b), the mean returns vector $\boldsymbol{\mu}$ is assumed to lie in the uncertain set $\mathcal{P}_{\hat{\boldsymbol{\mu}}}$ given by

$$\mathcal{P}_{\hat{\boldsymbol{\mu}}} = \{\boldsymbol{\mu} : |\mu_i - \hat{\mu}_i| \leq \delta_i, i = 1, \ldots, n\}. \tag{10}$$

The individual diagonal elements d_i of the covariance matrix D are assumed to lie in an interval $[\underline{d}_i, \overline{d}_i]$, i.e.,

$$\mathcal{P}_{\hat{D}} = \{D : D = \mathrm{diag}(d), d_i \in [\underline{d}_i, \overline{d}_i], i = 1, \ldots, n\}.$$

Finally, the columns of the matrix V are also assumed to be known approximately, i.e., V belongs to an elliptical uncertain set $\mathcal{P}_{\hat{V}}$:

$$\mathcal{P}_{\hat{V}} = \{V : V = \hat{V} + W, \|\boldsymbol{w}_i\|_g \leq \rho_i, i = 1, \ldots, n\},$$

where $\boldsymbol{w}_i$ is the ith column of W and $\|\boldsymbol{w}\|_g = \sqrt{\boldsymbol{w}^\top G \boldsymbol{w}}$ denotes the elliptic norm of $\boldsymbol{w}$ with respect to a symmetric, positive definite matrix G. Then the portfolio return is given by

$$\boldsymbol{r}^\top \boldsymbol{x} = \boldsymbol{\mu}^\top \boldsymbol{x} + \boldsymbol{f}^\top V \boldsymbol{x} + \boldsymbol{\epsilon}^\top \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}^\top \boldsymbol{x}, \boldsymbol{x}^\top (V^\top F V + D)\boldsymbol{x}).$$

The robust mean-variance optimization problem is given by

$$\min_{\boldsymbol{x}\in X} \left\{ \max_{V\in\mathcal{P}_{\hat{V}}, D\in\mathcal{P}_{\hat{D}}} \boldsymbol{x}^\top (V^\top F V + D)\boldsymbol{x} : \min_{\mu\in\mathcal{P}_{\hat{\mu}}} \boldsymbol{\mu}^\top \boldsymbol{x} \geq \mu_0 \right\}.$$

By introducing auxiliary variables ν and ς, Goldfarb and Iyengar (2003b) prove that this problem can be reformulated as

$$\min_{(\boldsymbol{x},\nu,\varsigma,\boldsymbol{\psi},\tau,\eta,\boldsymbol{t},\boldsymbol{s})} \quad \nu + \varsigma$$

[8] For two $n \times n$ symmetric matrices $A, B \in \mathcal{L}_n$, $A \succeq B$ (resp. $A \succ B$) means $A - B$ is positive semidefinite (resp. definite).

$$
\begin{aligned}
\text{s.t. } & \mathbf{1}_n^\top \boldsymbol{x} = 1,\ \hat{\boldsymbol{\mu}}^\top \boldsymbol{x} - \boldsymbol{\delta}^\top \boldsymbol{\psi} \geq \mu_0, \\
& \psi_i \geq x_i, \qquad \psi \geq -x_i, \quad i = 1, \ldots, n, \\
& \tau + \mathbf{1}_m^\top \boldsymbol{t} \leq \nu - \varsigma, \qquad \eta \leq \frac{1}{l_{\max}(H)}, \\
& \left\| \begin{bmatrix} 2\overline{D}^{1/2}\boldsymbol{x} \\ 1 - \varsigma \end{bmatrix} \right\| \leq 1 + \varsigma, \qquad \left\| \begin{bmatrix} 2\boldsymbol{\rho}^\top \boldsymbol{\psi} \\ \eta - \tau \end{bmatrix} \right\| \leq \eta + \tau, \\
& \left\| \begin{bmatrix} 2s_j \\ 1 - nl_j - t_j \end{bmatrix} \right\| \leq 1 - \eta l_j + t_j, \quad j = 1, \ldots, m,
\end{aligned}
$$

where $QLQ^\top$ is the spectral decomposition of $H = G^{-1/2}FG^{-1/2}$, $L = \mathrm{diag}(l_1, \ldots, l_m)$ ($l_{\max}$ is the maximum of these elements), and $\boldsymbol{s} = Q^\top H^{1/2} G^{1/2} \hat{V} \boldsymbol{x}$.

3.3.2 Box uncertainty on the covariance matrix

Instead of using uncertain sets based on estimates from a factor model, Tütüncü and Koenig (2004) directly specify intervals for the individual elements of the covariance matrix:

$$
\Sigma \in \mathcal{P}_{(\underline{\Sigma}, \overline{\Sigma})} = \{\Sigma : \ \underline{\Sigma} \preceq \Sigma \preceq \overline{\Sigma}\},
$$

where $\underline{\Sigma}$ and $\overline{\Sigma}$ are assumed to be positive semidefinite. The uncertainty of the expected return is the same with (10).

Suppose that short sales are not permitted (i.e., $\boldsymbol{x} \geq 0$), then the resulting problem is very simple to formulate. We just need to replace $\hat{\boldsymbol{\mu}}$ by $\hat{\boldsymbol{\mu}} - \boldsymbol{\delta}$ and Σ by $\overline{\Sigma}$ because

$$
\min_{\boldsymbol{x}} \left\{ \boldsymbol{x}^\top \overline{\Sigma} \boldsymbol{x} : \ (\hat{\boldsymbol{\mu}} - \boldsymbol{\delta})^\top \boldsymbol{x} \geq \mu_0, \mathbf{1}_n^\top \boldsymbol{x} = 1, \boldsymbol{x} \geq 0 \right\}
$$

is equivalent in fact to

$$
\min_{\boldsymbol{x}} \left\{ \max_{\Sigma \in \mathcal{P}_{(\underline{\Sigma}, \overline{\Sigma})}} \boldsymbol{x}^\top \Sigma \boldsymbol{x} : \ \min_{\boldsymbol{\mu} \in \mathcal{P}_{\hat{\boldsymbol{\mu}}}} \boldsymbol{\mu}^\top \boldsymbol{x} \geq \mu_0, \mathbf{1}_n^\top \boldsymbol{x} = 1, \boldsymbol{x} \geq 0 \right\}.
$$

In the general situation without short-sale constraints, the robust counterpart remains a convex problem, and is given by

$$
\begin{aligned}
\max_{\boldsymbol{x}, \boldsymbol{x}^+, \boldsymbol{x}^-, \underline{\Gamma}, \overline{\Gamma}} \quad & \langle \overline{\Gamma}, \overline{\Sigma} \rangle - \langle \underline{\Gamma}, \underline{\Sigma} \rangle \\
\text{s.t. } \quad & \mathbf{1}_n^\top \boldsymbol{x} = 1, \\
& \hat{\boldsymbol{\mu}}^\top \boldsymbol{x} - \boldsymbol{\delta}^\top (\boldsymbol{x}^+ + \boldsymbol{x}^-) \geq \mu_0, \\
& \boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-, \quad \boldsymbol{x}^+ \geq 0,\ \boldsymbol{x}^- \geq 0, \\
& \underline{\Gamma} \succeq 0, \quad \overline{\Gamma} \succeq 0, \quad \begin{bmatrix} \overline{\Gamma} - \underline{\Gamma} & \boldsymbol{x} \\ \boldsymbol{x}^\top & 1 \end{bmatrix} \succeq 0,
\end{aligned}
$$

where for any two symmetric matrices A and B, $\langle A, B \rangle$ represents the trace of the matrix product AB, i.e, $\langle A, B \rangle = \mathrm{trace}(AB)$.

3.4 Robust VaR

Recall in Sect. 2 that if the distribution of portfolio return is Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix Σ, the VaR at the β confidence level can be expressed as

$$\mathrm{VaR}_\beta(\boldsymbol{x}) = \zeta_\beta\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}} - \boldsymbol{\mu}^\top\boldsymbol{x}.$$

Now we assume that the true distribution of returns is only partially known. Particularly, in parallel with the robust mean-variance framework, we first consider the case where the distribution of portfolio return belongs to a certain set of distributions, i.e., $p(\boldsymbol{r}) \in \mathcal{P}_{(\boldsymbol{\mu},\Sigma)}$, whose mean and covariance matrix are perfectly known, denoted by $\boldsymbol{\mu}$ and Σ, respectively.

Mathematically, for a given confidence level $\beta \in (1/2, 1]$, and a given portfolio $\boldsymbol{x} \in X$, the robust VaR with respect to the set of probability distributions $\mathcal{P}_{(\boldsymbol{\mu},\Sigma)}$ is defined as

$$\mathrm{RVaR}_\beta(\boldsymbol{x}) = \min\gamma \quad \text{s.t.} \quad \sup_{p(\cdot)\in\mathcal{P}_{(\boldsymbol{\mu},\Sigma)}} \mathbb{P}\{\gamma \le -\boldsymbol{r}^\top\boldsymbol{x}\} \le 1-\beta.$$

El Ghaoui et al. (2003) show that $\sup_{p(\cdot)\in\mathcal{P}_{(\boldsymbol{\mu},\Sigma)}} \mathbb{P}\{\gamma \le -\boldsymbol{r}^\top\boldsymbol{x}\} \le 1-\beta$ has the following four equivalent counterparts:

1.
$$\tilde{\zeta}_\beta\sqrt{\boldsymbol{x}^\top\Sigma\boldsymbol{x}} - \boldsymbol{\mu}^\top\boldsymbol{x} \le \gamma, \tag{11}$$
where $\tilde{\zeta}_\beta = \sqrt{\beta/(1-\beta)}$;
2. There exist a symmetric matrix $M \in \mathcal{L}_{n+1}$ and $\tau \in \Re$ such that
$$\langle M, \Gamma\rangle \le \tau(1-\beta), \quad M \succeq 0,\ \tau \ge 0, \qquad M + \begin{bmatrix} 0 & \boldsymbol{x} \\ \boldsymbol{x}^\top & -\tau + 2\gamma \end{bmatrix} \succeq 0, \tag{12}$$
where $\Gamma = \left[\begin{smallmatrix} \Sigma+\boldsymbol{\mu}\boldsymbol{\mu}^\top & \boldsymbol{x} \\ \boldsymbol{x}^\top & 1 \end{smallmatrix}\right]$;
3. For every $\hat{\boldsymbol{\mu}} \in \Re^n$ such that
$$\begin{bmatrix} \Sigma & \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \\ (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top & \tilde{\zeta}_\beta^2 \end{bmatrix} \succeq 0, \tag{13}$$
we have $-\hat{\boldsymbol{\mu}}^\top\boldsymbol{x} \le \gamma$; or
4. There exist $\Lambda \in \mathcal{L}_n$ and $v \in \Re$ such that
$$\langle \Lambda, \Sigma\rangle + \tilde{\zeta}_\beta^2 v - \boldsymbol{\mu}^\top\boldsymbol{x} \le \gamma, \qquad \begin{bmatrix} \Lambda & \boldsymbol{x}/2 \\ \boldsymbol{x}^\top/2 & v \end{bmatrix} \succeq 0. \tag{14}$$

With these four equivalent formulations, we can easily solve the robust VaR portfolio selection problem with distribution uncertainty, which have the most interesting properties of transforming probability chance constraints into tractable convex inequality constraints.

We now turn to the case when $(\boldsymbol{\mu}, \Sigma)$ are only known to belong to a given convex set $\mathcal{P}$ of $\Re^n \times \mathcal{L}_n$. We denote by $\mathcal{P}_+$ the set $\{(\boldsymbol{\mu}, \Sigma) \in \mathcal{P} | \Sigma \succ 0\}$. Then the portfolio selection problem with robust VaR strategy can be formulated as:

$$\min_{M,\tau,\boldsymbol{x}} \max_{\boldsymbol{\mu},\Sigma} \Bigg\{ \gamma : (\boldsymbol{\mu}, \Sigma) \in \mathcal{P}_+, \langle M, \Gamma\rangle \le \tau(1-\beta), M \succeq 0, \tau \ge 0,$$

$$M + \begin{bmatrix} 0 & \boldsymbol{x} \\ \boldsymbol{x}^\top & -\tau + 2\gamma \end{bmatrix} \succeq 0 \Bigg\},$$

$$\min_{\boldsymbol{x}\in X} \max_{\hat{\boldsymbol{\mu}},\boldsymbol{\mu},\Sigma} \left\{ -\boldsymbol{\mu}^\top \boldsymbol{x} : (\boldsymbol{\mu}, \Sigma) \in \mathcal{P}_+, \begin{bmatrix} \Sigma & \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \\ (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top & \tilde{\zeta}_\beta^2 \end{bmatrix} \succeq 0 \right\},$$

or

$$\min_{\Lambda,v,\boldsymbol{x}} \max_{\boldsymbol{\mu},\Sigma} \left\{ \langle \Lambda, \Sigma \rangle + \tilde{\zeta}_\beta^2 v - \boldsymbol{\mu}^\top \boldsymbol{x} : (\boldsymbol{\mu}, \Sigma) \in \mathcal{P}_+, \begin{bmatrix} \Lambda & \boldsymbol{x}/2 \\ \boldsymbol{x}^\top/2 & v \end{bmatrix} \succeq 0 \right\},$$

which correspond to cases (12), (13), and (14). We omit the case of (11) since its formulation is very similar to the mean-VaR portfolio selection with known distribution. The above three formulations are ready for application once the structure of the uncertain set $\mathcal{P}_+$ is specified. Fortunately, $\mathcal{P}_+$ consists of the information of the mean and covariance of the portfolio return, so it can be specified as in the robust mean-variance section. Here we omit this discussion because of space limitation and refer interested readers to El Ghaoui et al. (2003).

3.5 Robust CVaR

In this section, we consider another type of uncertainty which is directly associated with the distribution of the portfolio return rather than its first- and second-order moments. More specifically, we assume in this section that the density function of the portfolio return is only known to belong to a certain set $\mathcal{P}$ of distributions, i.e., $p(\cdot) \in \mathcal{P}$. Zhu and Fukushima (2008) define the robust CVaR (RCVaR) for fixed $\boldsymbol{x} \in X$ with respect to $\mathcal{P}$ as

$$\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \max_{p(\cdot)\in\mathcal{P}} \mathrm{CVaR}_\beta(\boldsymbol{x}).$$

Moreover, they investigate some special cases of $\mathcal{P}$ such as mixture distribution uncertain set and box uncertain set that meet practical requirements and, at the same time, can be efficiently solved.

3.5.1 Mixture distribution

We assume in our discussion that the density function of $\boldsymbol{r}$ is only known to belong to a set of distributions which consists of all the mixture distributions of some possible distribution scenarios; that is,

$$p(\cdot) \in \mathcal{P}_M = \left\{ \sum_{l=1}^{L} \lambda_l p^l(\cdot) : \sum_{l=1}^{L} \lambda_l = 1,\ \lambda_l \geq 0,\ l = 1, \dots, L \right\},$$

where $p^l(\cdot)$ denotes the l-th distribution scenario, and l denotes the number of possible scenarios. Mixture distributions have already been studied in robust statistics and used in modeling the distribution of financial data (Hall et al. 1989; Peel and McLachlan 2000).

Denote

$$\Lambda = \left\{ \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_L) : \sum_{l=1}^{L} \lambda_l = 1, \lambda_l \geq 0, l = 1, \dots, L \right\},$$

and

$$F_\beta^l(\boldsymbol{x},\alpha) = \alpha + \frac{1}{1-\beta}\int_{\boldsymbol{r}\in\Re^n}[-\boldsymbol{r}^\top\boldsymbol{x}-\alpha]^+ p^l(\boldsymbol{r})d\boldsymbol{r}, \quad l=1,\dots,L.$$

Using the minimax theorem, for each $\boldsymbol{x}$, $\mathrm{RCVaR}_\beta(\boldsymbol{x})$ with respect to $\mathscr{P}_M$ is given by

$$\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha\in\Re}\max_{l\in\mathcal{L}} F_\beta^l(\boldsymbol{x},\alpha),$$

where $\mathcal{L}=\{1,2,\dots,L\}$.

Denote

$$F_\beta^{\mathcal{L}}(\boldsymbol{x},\alpha) = \max_{l\in\mathcal{L}} F_\beta^l(\boldsymbol{x},\alpha).$$

Then, minimizing $\mathrm{RCVaR}_\beta(\boldsymbol{x})$ over X can be achieved by minimizing $F_\beta^{\mathcal{L}}(\boldsymbol{x},\alpha)$ over $X\times\Re$, i.e.,

$$\min_{\boldsymbol{x}\in X}\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \min_{(\boldsymbol{x},\alpha)\in X\times\Re} F_\beta^{\mathcal{L}}(\boldsymbol{x},\alpha).$$

More specifically, if $(\boldsymbol{x}^*,\alpha^*)$ attains the right-hand side minimum, then $\boldsymbol{x}^*$ attains the left-hand side minimum and α^* attains the minimum of $F_\beta^{\mathcal{L}}(\boldsymbol{x}^*,\alpha)$, and vice versa. Furthermore, it can be seen that the RCVaR minimization is equivalent to

$$\min_{(\boldsymbol{x},\alpha,\theta)\in X\times\Re\times\Re}\left\{\theta:\alpha+\frac{1}{1-\beta}\int_{\boldsymbol{r}\in\Re^n}[f(\boldsymbol{x},\boldsymbol{r})-\alpha]^+p^l(\boldsymbol{r})d\boldsymbol{r}\le\theta, \quad l=1,\dots,L\right\}. \tag{15}$$

Thus, the computation of (15) is similar to CVaR. More specifically,

$$\min_{(\boldsymbol{x},\alpha,\theta)\in X\times\Re\times\Re}\left\{\theta:\alpha+\frac{1}{S^l(1-\beta)}\sum_{k=1}^{S^l}[f(\boldsymbol{x},\boldsymbol{r}_{[k]}^l)-\alpha]^+\le\theta, \quad l=1,\dots,L\right\}, \tag{16}$$

where $\boldsymbol{r}_{[k]}^l$ is the k-th sample with respect to the l-th distribution scenario $p^l(\cdot)$, and S^l denotes the number of the corresponding samples. Instead of the simple random sampling method, some improved sampling approaches can be used to approximate the integral. Generally, the approximation of problem (15) can be formulated as

$$\min_{(\boldsymbol{x},\alpha,\theta)\in X\times\Re\times\Re}\left\{\theta:\alpha+\frac{1}{1-\beta}\sum_{k=1}^{S^l}\pi_k^l[f(\boldsymbol{x},\boldsymbol{r}_{[k]}^l)-\alpha]^+\le\theta, \quad l=1,\dots,L\right\}, \tag{17}$$

where π_k^l denotes the probability according to the k-th sample with respect to the l-th distribution scenario $p^l(\cdot)$. If π_k^l is equal to $\frac{1}{S^l}$ for all k, then (17) reduces to (16). In the following, we denote $\boldsymbol{\pi}^l=(\pi_1^l,\dots,\pi_{S^l}^l)^\top$.

Let $m=\sum_{l=1}^{l} S^l$. Then, by introducing an auxiliary vector $\boldsymbol{u}=(\boldsymbol{u}^1;\boldsymbol{u}^2;\dots;\boldsymbol{u}^L)\in\Re^m$, the optimization problem (17) can be rewritten as the following minimization problem with variables $(\boldsymbol{x},\boldsymbol{u},\alpha,\theta)\in\Re^n\times\Re^m\times\Re\times\Re$:

$$\begin{aligned}&\min\ \theta\\ &\text{s.t.}\quad \boldsymbol{x}\in X,\end{aligned}$$

$$\alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^l\right)^\top \boldsymbol{u}^l \le \theta, \qquad (18)$$
$$u_k^l \ge f(\boldsymbol{x}, \boldsymbol{r}_{[k]}^l) - \alpha,\ u_k^l \ge 0, \quad k = 1, \ldots, S^l,\ l = 1, \ldots, L.$$

More specifically, if $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ solves (18), then $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ solves (17). Conversely, if $(\boldsymbol{x}^*, \alpha^*, \theta^*)$ solves (17), then $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ solves (18), where $\boldsymbol{u}^*$ is constructed as

$$u_k^{l*} = [f(\boldsymbol{x}^*, \boldsymbol{r}_{[k]}^l) - \alpha^*]^+, \quad k = 1, \ldots, S^l,\ l = 1, \ldots, l.$$

In the special case where $L = 1$ (i.e., the distribution of $\boldsymbol{r}$ is confirmed to be only one scenario without any other possibility), problem (18) is exactly that of Rockafellar and Uryasev (2000) with $\pi_k^l = \frac{1}{S^1}$ for all k. Although in the above discussion we assume a continuous distribution, it is easy to see from Rockafellar and Uryasev (2002) that the results also hold in the general distribution case. For example, in the discrete distribution case, it is straightforward to interpret the integral as a summation. Moreover, it should be interpreted as a mixture of an integral and a summation in the case of mixed continuous and discrete distributions.

3.5.2 Discrete distribution

In this subsection, we assume that $\boldsymbol{r}$ follows a discrete distribution and discuss the minimization of the robust CVaR under box uncertainty. From a practical viewpoint, this consideration still makes sense for continuous distributions, since we usually use a discretization procedure to approximate the integral resulting from a continuous distribution.

Let the sample space of a random vector $\boldsymbol{r}$ be given by $\{\boldsymbol{r}_{[1]}, \boldsymbol{r}_{[2]}, \ldots, \boldsymbol{r}_{[S]}\}$ with $\Pr\{\boldsymbol{r}_{[k]}\} = \pi_k$ and $\sum_{k=1}^S \pi_k = 1,\ \pi_k \ge 0, k = 1, \ldots, S$. Denote $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_S)^\top$ and define

$$G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}) = \alpha + \frac{1}{1-\beta}\sum_{k=1}^S \pi_k [\boldsymbol{x}^\top \boldsymbol{r}_{[k]} - \alpha]^+.$$

For given $\boldsymbol{x}$ and $\boldsymbol{\pi}$, the corresponding CVaR is then defined as (Rockafellar and Uryasev 2002)

$$\mathrm{CVaR}_\beta(\boldsymbol{x}, \boldsymbol{\pi}) = \min_{\alpha \in \Re} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

We denote $\mathcal{P}$ as $\mathcal{P}_\pi$ in the case of a discrete distribution. Then we may identify $\mathcal{P}_\pi$ as a subset of $\Re^S$ and the robust CVaR for fixed $\boldsymbol{x} \in X$ with respect to $\mathcal{P}_\pi$ is defined as

$$\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \max_{\boldsymbol{\pi} \in \mathcal{P}_\pi} \mathrm{CVaR}_\beta(\boldsymbol{x}, \boldsymbol{\pi}),$$

or equivalently,

$$\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \max_{\boldsymbol{\pi} \in \mathcal{P}_\pi} \min_{\alpha \in \Re} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

Then, for each $\boldsymbol{x}$, we have

$$\mathrm{RCVaR}_\beta(\boldsymbol{x}) = \min_{\alpha \in \Re} \max_{\boldsymbol{\pi} \in \mathcal{P}_\pi} G_\beta(\boldsymbol{x}, \alpha, \boldsymbol{\pi}).$$

Particularly, if $\mathcal{P}_\pi$ is a compact convex set, the problem of minimizing $\mathrm{RCVaR}_\beta(\boldsymbol{x})$ over X can be written as

$$\min_{(\boldsymbol{x},\alpha,\theta)\in X\times\Re\times\Re}\left\{\theta : \max_{\boldsymbol{\pi}\in\mathcal{P}_\pi} \alpha + \frac{1}{1-\beta}\sum_{k=1}^{S}\pi_k[-\boldsymbol{x}^\top \boldsymbol{r}_{[k]} - \alpha]^+ \le \theta\right\}. \tag{19}$$

By introducing an auxiliary vector $\boldsymbol{u}\in\Re^S$, we can show that problem (19) is equivalent to the following minimization problem with variables $(\boldsymbol{x},\boldsymbol{u},\alpha,\theta)\in\Re^n\times\Re^S\times\Re\times\Re$:

$$\min\left\{\theta : \boldsymbol{x}\in X, \max_{\boldsymbol{\pi}\in\mathcal{P}_\pi}\alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^\top\boldsymbol{u}\le\theta, u_k\ge -\boldsymbol{x}^\top\boldsymbol{r}_{[k]}-\alpha, u_k\ge 0, k=1,\dots,S\right\}. \tag{20}$$

In the following, we show that under the box uncertainty in the distributions, (20) can be casted as linear programs.

Suppose $\boldsymbol{\pi}$ belongs to a box set, i.e.,

$$\boldsymbol{\pi}\in\mathcal{P}_\pi^B = \left\{\boldsymbol{\pi} : \boldsymbol{\pi} = \boldsymbol{\pi}^0 + \boldsymbol{\eta}, \mathbf{1}_S^\top\boldsymbol{\eta} = 0, \underline{\boldsymbol{\eta}}\le\boldsymbol{\eta}\le\overline{\boldsymbol{\eta}}\right\},$$

where $\boldsymbol{\pi}^0$ is a nominal distribution which represents the most likely distribution, $\mathbf{1}_S$ denotes an S-dimensional vector of ones, and $\underline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{\eta}}$ are given constant vectors. The condition $\mathbf{1}_S^\top\boldsymbol{\eta}=0$ ensures $\boldsymbol{\pi}$ to be a probability distribution, and the non-negativity constraint $\boldsymbol{\pi}\ge 0$ is included in the binding constraints $\underline{\boldsymbol{\eta}}\le\boldsymbol{\eta}\le\overline{\boldsymbol{\eta}}$.

Since

$$\alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^\top\boldsymbol{u} = \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^\top\boldsymbol{u} + \frac{1}{1-\beta}\boldsymbol{\eta}^\top\boldsymbol{u},$$

we have

$$\max_{\boldsymbol{\pi}\in\mathcal{P}_\pi^B}\alpha + \frac{1}{1-\beta}\boldsymbol{\pi}^\top\boldsymbol{u} = \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^\top\boldsymbol{u} + \frac{\gamma^*(\boldsymbol{u})}{1-\beta},$$

where $\gamma^*(\boldsymbol{u})$ is the optimal value of the following linear program

$$\max_{\boldsymbol{\eta}\in\Re^S}\left\{\boldsymbol{u}^\top\boldsymbol{\eta} : \mathbf{1}_S^\top\boldsymbol{\eta}=0,\ \underline{\boldsymbol{\eta}}\le\boldsymbol{\eta}\le\overline{\boldsymbol{\eta}}\right\}. \tag{21}$$

The dual program of (21) is given by

$$\min_{(z,\boldsymbol{\xi},\boldsymbol{\omega})\in\Re\times\Re^S\times\Re^S}\left\{\overline{\boldsymbol{\eta}}^\top\boldsymbol{\xi} + \underline{\boldsymbol{\eta}}^\top\boldsymbol{\omega} : \mathbf{1}_S z + \boldsymbol{\xi} + \boldsymbol{\omega} = \boldsymbol{u}, \boldsymbol{\xi}\ge 0,\ \boldsymbol{\omega}\le 0\right\}. \tag{22}$$

Consider the following minimization problem over $(\boldsymbol{x},\boldsymbol{u},z,\boldsymbol{\xi},\boldsymbol{\omega},\alpha,\theta)\in\Re^n\times\Re^S\times\Re\times\Re^S\times\Re^S\times\Re\times\Re$:

$$\begin{aligned}
&\min\ \theta\\
&\text{s.t.}\ \boldsymbol{x}\in X,\\
&\qquad \alpha + \frac{1}{1-\beta}\left(\boldsymbol{\pi}^0\right)^\top\boldsymbol{u} + \frac{1}{1-\beta}\left(\overline{\boldsymbol{\eta}}^\top\boldsymbol{\xi} + \underline{\boldsymbol{\eta}}^\top\boldsymbol{\omega}\right)\le\theta, \qquad (23)\\
&\qquad \mathbf{1}_S z + \boldsymbol{\xi} + \boldsymbol{\omega} = \boldsymbol{u},\quad \boldsymbol{\xi}\ge 0,\ \boldsymbol{\omega}\le 0,
\end{aligned}$$

$$u_k \geq -\boldsymbol{x}^\top \boldsymbol{r}_{[k]} - \alpha, \quad u_k \geq 0, \ k = 1, \dots, S.$$

If $(\boldsymbol{x}^*, \boldsymbol{u}^*, z^*, \boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \alpha^*, \theta^*)$ solves (23), then $(\boldsymbol{x}^*, \boldsymbol{u}^*, \alpha^*, \theta^*)$ solves (20) with $\mathcal{P}_\pi = \mathcal{P}_\pi^B$. Conversely, if $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ solves (20) with $\mathcal{P}_\pi = \mathcal{P}_\pi^B$, then $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{u}}^*, \tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*, \tilde{\alpha}^*, \tilde{\theta}^*)$ solves (23), where $(\tilde{z}^*, \tilde{\boldsymbol{\xi}}^*, \tilde{\boldsymbol{\omega}}^*)$ is an optimal solution to (22) with $\boldsymbol{u} = \tilde{\boldsymbol{u}}^*$.

In the special case where $\underline{\boldsymbol{\eta}} = \overline{\boldsymbol{\eta}} = 0$, (23) reduces to the original CVaR minimization problem. Using a similar approach, we can also assume $\boldsymbol{\pi}$ belongs to an ellipsoid and transform the minimization of RCVaR into a second-order cone program.

3.6 Size and shape of the uncertain set

While there is a large body of literature concerned with the formulation of robust portfolio selection, the discussion in the literature that focuses specifically on how to specify the size and shape of the set of the uncertain parameter in an attractive manner is sparse; the size of the set ensures the probability that the uncertain parameter takes on a value in the set, while the shape of the set determines the complexity of the robust optimization problem.

In practice, a distribution is often completely determined by its parameters. In this sense, we may write $p(\cdot)$ as $p(\cdot|\boldsymbol{\theta})$. To identify $\boldsymbol{\theta}$, two approaches have been employed in both theoretical studies and practice. One is estimating them by data: collecting a set of historical data to estimate the parameters of the likelihood distribution. For example, if we consider the returns following a joint normal distribution, the corresponding mean and covariance of the portfolio returns can be estimated immediately from historical returns. However, due to estimation errors, we can only obtain a confidence region Θ of the parameters $\boldsymbol{\theta}$, and we believe with certainty that the real parameters $\boldsymbol{\theta}$ belong to Θ, i.e., $\boldsymbol{\theta} \in \Theta$. Thus, we may choose proper $\boldsymbol{\theta}$ in Θ to construct the likelihood distribution.

The other approach is expert prediction: all the different distributions predicted by a group of experts can be regarded as the likelihood distributions (Lutgens and Schotman 2006). Actually, the success of quantitative approaches to portfolio selection relies on the return model employed. While each alternative return model supplies the portfolio manager with a specific set of parameters such as the mean and covariance of the returns, there is no consensus among the experts on which return model is most appropriate, and hence they will recommend different parameters for the portfolio return distribution. Apart from relying on one particular expert, we may combine the information from a number of experts and consider each of their predictions as a likelihood distribution. The Bayesian model averaging approach reviewed later will be useful for this purpose.

3.6.1 Specification for factor model

In Sect. 3.3.1, we assume that the return vector $\boldsymbol{r}$ is given by the linear model,

$$\boldsymbol{r} = \boldsymbol{\mu} + V^\top \boldsymbol{f} + \boldsymbol{\epsilon}, \tag{24}$$

where $\boldsymbol{\epsilon}$ is the residual return. With robust technique for portfolio selection, in addition to the least squares estimates $(\hat{\boldsymbol{\mu}}, \hat{V})$ of $(\boldsymbol{\mu}, V)$, we need to specify the uncertain structures $\mathcal{P}_{\hat{D}}$, $\mathcal{P}_{\hat{V}}$, and $\mathcal{P}_{\hat{\mu}}$ by choosing the proper matrix G for the elliptic norm $\|\cdot\|_g$ and the bounds $\overline{d}_j$, ρ_j, δ_j, $j = 1, \dots, n$.

We present here the specifications developed by Goldfarb and Iyengar (2003b). Suppose we have historical data on asset returns, $\{\boldsymbol{r}^t, t = 1, \dots, T\}$, for T periods and the corresponding factor returns $\{\boldsymbol{f}^t, t = 1, \ldots, T\}$. According to (24), we have

$$r_i^t = \mu_i^t + \sum_{j=1}^{n} V_{ji} f_j^t + \epsilon_i^t, \quad i = 1, \ldots, n, \; t = 1, \ldots, T.$$

In the regression analysis, we assume $\{\epsilon_i^t, i = 1, \ldots, n, t = 1, \ldots, T\}$ are all independent normal random variables and $\epsilon_i^t \sim \mathcal{N}(0, \sigma_i^2)$ for all $t = 1, \ldots, T$.

Let $R_T = [\boldsymbol{r}^1, \boldsymbol{r}^2 \ldots, \boldsymbol{r}^T] \in \Re^{n\times T}$ be the matrix of asset returns and $Q = [\boldsymbol{f}^1, \boldsymbol{f}^2, \ldots, \boldsymbol{f}^T] \in \Re^{m\times T}$ be the matrix of factor returns. Then we can write the regression equation for asset i as

$$\boldsymbol{r}_i = H z_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{r}_i = (r_i^1, \ldots, r_i^T)^\top$, $H = (\mathbf{1}_T, Q^\top)$, $z_i = (\mu_i, V_{1i}, V_{2i}, \ldots, V_{mi})^\top$, and $\boldsymbol{\epsilon}_i = (\epsilon_i^1, \ldots, \epsilon_i^T)^\top$. With the least-squares regression, we can estimate $(\mu_i, V_{1i}, V_{2i}, \ldots, V_{mi})^\top$ as

$$(\hat{\mu}_i, \hat{V}_{1i}, \hat{V}_{2i}, \ldots, \hat{V}_{mi})^\top = \hat{z}_i = (H^\top H)^{-1} H^\top \boldsymbol{r}_i.$$

In what follows, we describe how to estimate G, ρ_i, δ_i, and $\overline{d}_i$.

Denote

$$\mathcal{Y} = \frac{(z_i - \hat{z}_i)^\top (H^\top H)(z_i - \hat{z}_i)}{(m+1)s_i^2},$$

where $s_i^2 = \frac{\|\boldsymbol{r}_i - A\hat{z}_i\|^2}{T-m-1}$. With the normal assumption on $\boldsymbol{\epsilon}_i$, Anderson (1984) shows that $\mathcal{Y}$ follows an $F(m+1, T-m-1)$ distribution.

Let $q_{m+1}(\beta)$ be the β quantile of $\mathcal{Y}$, i.e.,

$$\mathbb{P}\left\{\frac{(z_i - \hat{z}_i)^\top (H^\top H)(z_i - \hat{z}_i)}{(m+1)s_i^2} \le q_{m+1}(\beta)\right\} = \beta.$$

Then, we can set

$$G = QQ^\top - \frac{1}{T}(Q\mathbf{1}_T)(Q\mathbf{1}_T)^\top,$$

$$\delta_i = \sqrt{(m+1)(H^\top H)_{11}^{-1} q_{m+1}(\beta) s_i^2},$$

$$\rho_i = \sqrt{(m+1) q_{m+1}(\beta) s_i^2},$$

to ensure $(\boldsymbol{\mu}, V) \in \mathcal{P}_{\hat{\mu}} \times \mathcal{P}_{\hat{V}}$ with at least β^n certainty.

Up to now, there is no realistic method for the determination of $\overline{d}_i$ since we cannot find a confidence interval for d_i although s_i^2 is an unbiased estimator of d_i^2. Goldfarb and Iyengar (2003b) suggest a possible solution by using the bootstrap technique.

3.6.2 Specification for mean and covariance

Next, we show how one can directly define a confidence region for mean and covariance statistics such that it is assured with high probability to contain the given statistics of the distribution that generated $(\boldsymbol{\mu}, \Sigma)$.

Given T samples from $\boldsymbol{r}$ and an empirical estimate of the mean $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\Sigma}$, Delage and Ye (2008) show that with probability greater than or equal to β, the true parameter $(\boldsymbol{\mu}, \Sigma)$ belongs to the following set

$$(\boldsymbol{\mu}, \Sigma) \in \mathscr{P}_{(\hat{\mu},\hat{\Sigma})} = \left\{(\boldsymbol{\mu}, \Sigma) : (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top \hat{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \le \chi, \frac{1}{1-q}\hat{\Sigma} \preceq \Sigma \preceq \frac{1}{1-q-\chi}\hat{\Sigma}\right\},$$

if

$$T > \max\left\{(\hat{R}^2+2)^2\left(2+\sqrt{2\ln(4/(1-\beta))}\right)^2, \frac{\left((8+\sqrt{32\ln(4/(1-\beta))})\right)^2}{(\sqrt{\hat{R}+4}-\hat{R})^4}\right\},$$

$$q = (\bar{R}^2/\sqrt{T})\left(\sqrt{1-n/\bar{R}^4}+\sqrt{\ln(4/(1-\beta))}\right),$$

$$\chi = (\bar{R}^2/T)\left(2+\sqrt{2\ln(2/(1-\beta))}\right)^2,$$

$$\bar{R} = \hat{R}\left(1-(\hat{R}^2+2)\frac{2+\sqrt{2\ln(4/(1-\beta))}}{\sqrt{T}}\right)^{-1/2},$$

$$\hat{R} = \max_{t\in\{1,\dots,T\}}\left\{(\boldsymbol{r}^t-\hat{\boldsymbol{\mu}})^\top\hat{\Sigma}^{-1/2}(\boldsymbol{r}^t-\hat{\boldsymbol{\mu}})\right\}.$$

3.7 Hansen and Sargent approach

Hansen and Sargent (2008) summarize a series of their papers and those in the area of their robust approach, which was developed based on robust control theory (Whittle 1990, 1996). In contrast to the robust approaches reviewed in previous sections, the deviations from the true model are captured by their framework. To see how this works, consider an investor who uses an approximate model,

$$\boldsymbol{y}_{t+1} = A\boldsymbol{y}_t + B\boldsymbol{u}_t + C\tilde{\boldsymbol{\epsilon}}_{t+1}, \tag{25}$$

where $\boldsymbol{y}_{t+1}$ is a state vector, $\boldsymbol{u}_t$ a vector of controls, A, B and C are constant matrices, and $\tilde{\boldsymbol{\epsilon}}_t$ is a vector of standardized Gaussian noises. Unsure about the validity of (25), the investor entertains a set of approximate models represented by

$$\boldsymbol{y}_{t+1} = A\boldsymbol{y}_t + B\boldsymbol{u}_t + C(\boldsymbol{\epsilon}_{t+1} + \boldsymbol{w}_{t+1}), \tag{26}$$

where $\boldsymbol{\epsilon}_{t+1}$ is another vector of noises and $\boldsymbol{w}_{t+1}$ can be a very general stochastic process that represents the deviations of those models from (25). One of their key innovations is to study the impact of the discounted future deviations of (26) from (25),

$$R(\boldsymbol{w}) = 2\mathbb{E}_0\sum_{t=0}^{\infty}\beta^{t+1}I(\boldsymbol{w}_{t+2}),$$

where $\mathbb{E}_0$ is the expectation under (26), β is a discount parameter, and $I(\boldsymbol{w}_{t+2})$ is the entropy measure of the deviations each period. Under suitable conditions, the maxmin problem of an objective with an upper bound on $R(\boldsymbol{w})$ can be transferred into a standard problem without misspecification by adding a suitable penalty function into the original objective function, as demonstrated below.

In the mean-variance framework with unknown mean, let $\boldsymbol{\delta} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ with $\hat{\boldsymbol{\mu}}$ as the sample mean and $\boldsymbol{\mu}$ the unknown true mean. Then the entropy measure can be written as

$$I(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^\top \Sigma^{-1}\boldsymbol{\delta}.$$

The investor solves the maxmin problem

$$\max_{\boldsymbol{x}} \min_{\boldsymbol{\delta}} \left\{ \frac{\lambda}{2}\boldsymbol{\delta}^\top \Sigma^{-1}\boldsymbol{\delta} + (\hat{\boldsymbol{\mu}} - \boldsymbol{\delta})\boldsymbol{x} - \frac{\gamma}{2}\boldsymbol{x}^\top \Sigma \boldsymbol{x} \right\},$$

where λ is the coefficient of the penalty function determined by a given upper bound on $I(\boldsymbol{\delta})$. The solution to the maxmin problem is the same as

$$\max_{\boldsymbol{x}} \left\{ \hat{\boldsymbol{\mu}}^\top \boldsymbol{x} - \frac{1}{2\lambda}\boldsymbol{x}^\top \Sigma \boldsymbol{x} - \frac{\gamma}{2}\boldsymbol{x}^\top \Sigma \boldsymbol{x} \right\}.$$

This says that introducing model uncertainty is observationally equivalent to increasing risk aversion.

There is a large and growing application of the Hansen and Sargent approach to a number of financial issues. For example, Uppal and Wang (2003) use it to show model uncertainty is an important explanation of the underdiversification puzzle, Maenhout (2004) applies it to explain the risk premium puzzle, and Kleshchelskiy and Vincentz (2007) adapt it to derive an equilibrium model of both bond and equity markets and show that the model fits well the key empirical regularities of the data. Theoretically, Maccheroni et al. (2006) provide a variational representation of preferences that unifies both the earlier multi-prior approach and the Hansen and Sargent one into one coherent decision-making framework. Strzalecki (2008) introduces the axiomatic foundation for choosing entropy in penalizing deviations between models.

4 Optimal estimation

The operations research literature often takes the parameter estimation as given, and then explore robustness to various uncertainties. In contrast, there is a vast literature in finance that focuses on developing the best estimators of the unknown mean and variance for the purpose of portfolio selection. While most of them are developed under normality assumption for the asset returns, the resulting estimators may be robust to certain distributional assumptions, but not so to others. From a statistical point of view, there are various estimation procedures that are robust to outliers and general distributional assumptions (see Huber 2003; Maronna et al. 2006). In contrast to the robust control and the statistics literature, the methods developed in finance that are described below do not treat estimation and optimization separately. The estimation is designed to maximize the objective function at hand.

Recall that, in practice, the optimal portfolio $\boldsymbol{x}^*$ is not computable because μ and Σ are unknown. To implement the mean-variance framework of Markowitz (1952), the optimal portfolio weights are usually chosen using a two-step procedure. Suppose an investor has T periods of observed return data $R_T = \{\boldsymbol{r}^1, \boldsymbol{r}^2, \ldots, \boldsymbol{r}^T\}$ and would like to form a portfolio for period $T+1$. First, the mean and covariance matrix of the asset returns are estimated based on the observed data. Under the assumption that $\boldsymbol{r}^t$ is i.i.d. normal, the standard estimates are

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}^t, \qquad \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}^t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}^t - \hat{\boldsymbol{\mu}})^\top.$$

Second, these sample estimates are then treated as if they were the true parameters, and then used to compute the optimal portfolio weights. We call such a portfolio rule the *plug-in rule*, whose optimal portfolio weights are

$$\hat{\boldsymbol{x}} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}.$$

Statistically, the estimator is efficient, and is usually regarded as very good. However, as will be clear, the estimator is far from being good in terms of maximizing the expected utility.

In general, an estimated portfolio rule can be defined as a function of the historical returns data R_T,

$$\hat{\boldsymbol{x}} = f(\boldsymbol{r}^1, \boldsymbol{r}^2, \dots, \boldsymbol{r}^T).$$

Intuitively, an investor who uses $\hat{\boldsymbol{x}}$ should be worse off than another investor who knows the true parameters. How do we assess the utility loss from using a given portfolio rule? Based on standard statistical decision theory, we define the *loss function* of using $\hat{\boldsymbol{x}}$ as

$$L(\boldsymbol{x}^*, \hat{\boldsymbol{x}}) = u(\boldsymbol{x}^*) - u(\hat{\boldsymbol{x}}),$$

where $u(\boldsymbol{x}) = \boldsymbol{\mu}^\top\boldsymbol{x} - \frac{\gamma}{2}\boldsymbol{x}^\top\Sigma\boldsymbol{x}$. Because $\hat{\boldsymbol{x}}$ is not equal to $\boldsymbol{x}^*$ in general, the utility loss is strictly positive. However, $\hat{\boldsymbol{x}}$ is a function of R_T, so the loss depends on the realizations of the historical returns data. It is important for decision purposes to consider the average losses involving actions taken under all possible outcomes of R_T. This calls for the use of the expected loss function, also known as the *risk function*, which is defined as

$$\rho(\boldsymbol{x}^*, \hat{\boldsymbol{x}}) = \mathbb{E}[L(\boldsymbol{x}^*, \hat{\boldsymbol{x}})] = u(\boldsymbol{x}^*) - \mathbb{E}[u(\hat{\boldsymbol{x}})],$$

where the expectation is taken with respect to the true distribution of R_T. Thus, for a given $\boldsymbol{\mu}$ and Σ (or a given $\boldsymbol{x}^*$), $\rho(\boldsymbol{x}^*, \hat{\boldsymbol{x}})$ represents the expected loss over R_T that is incurred in using the portfolio rule $\hat{\boldsymbol{x}}$.

The risk function provides a criterion for ranking various portfolio rules and the rule that has the lowest risk is the most preferred. Instead, we can equivalently rank portfolios by their expected utility $\mathbb{E}[u(\hat{\boldsymbol{x}})]$. In general, one portfolio rule may generate higher expected utility than another over some parameter values of $(\boldsymbol{\mu}, \Sigma)$, but lower over some other values. In this case, the two portfolio rules do not uniformly dominate each other, and which one of them is preferable depends on the actual value of $\boldsymbol{\mu}$ and Σ. However, there are portfolio rules that are *inadmissible* in the sense that there exists another portfolio rule that generates higher expected utility for every possible choice of $(\boldsymbol{\mu}, \Sigma)$. Clearly, inadmissible portfolio rules should be eliminated from consideration.

As a major alternative to the standard maximum likelihood estimator, the plug-in rule, Jorion (1986, 1991) develops an estimator of $\boldsymbol{\mu}$ motivated by both a shrinkage consideration and a Bayesian analysis. Jorion's Bayes-Stein shrinkage portfolio rule, known as shrinkage estimators pioneered by Stein (1956), of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}}^{\mathrm{BS}} = (1 - v)\hat{\boldsymbol{\mu}} + v\hat{\mu}_g\mathbf{1}_n,$$

where v is the weight given to the shrinkage target $\hat{\mu}_g$, and $1-v$ is the weight on the sample mean. The target used by Jorion (1986) is the average excess return of the sample global minimum-variance portfolio,

$$\hat{\mu}_g = \frac{\mathbf{1}_n^\top \tilde{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_n^\top \tilde{\Sigma}^{-1}\mathbf{1}_n} = \frac{\mathbf{1}_n^\top \hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_n^\top \hat{\Sigma}^{-1}\mathbf{1}_n},$$

and the weight is

$$v = \frac{n+2}{(n+2) + T(\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_n)^\top \tilde{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_n)},$$

where $\tilde{\Sigma}$ is defined as $\frac{T}{T-n-2}\hat{\Sigma}$. From a shrinkage point of view, combining $\hat{\mu}^{\mathrm{BS}}$ with $\tilde{\Sigma}$ gives an estimator of the optimal portfolio weights. The new estimator is in general much better than the plug-in rule.

Jobson et al. (1979) and Jobson and Korkie (1980) also analyze shrinkage estimators for the mean-variance frontier. Ledoit and Wolf (2003) develop shrinkage covariance matrix estimators that are useful for a large number of assets. Lutgens (2004) summarizes some of the literature. Mori (2004), Ter Horst et al. (2006), and Kan and Zhou (2007) analyze both analytically and numerically the performance of various estimators. In addition, Kan and Zhou (2007) propose a new three-fund strategy, with investments in the risk-free asset, sample tangent portfolio, and sample global minimum variance portfolio,

$$\hat{\boldsymbol{x}}^{\mathrm{KZ}} = \frac{T-n-2}{\gamma c_1 T}\left[\hat{\eta}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}} + (1-\hat{\eta})\hat{\mu}_g\hat{\Sigma}^{-1}\mathbf{1}_n\right],$$

where $\hat{\eta} = \hat{\psi}^2/(\hat{\psi}^2 + n/T)$, $\hat{\psi}^2 = (\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_n)^\top \hat{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \hat{\mu}_g\mathbf{1}_n)$, $\hat{\mu}_g = \hat{\boldsymbol{\mu}}^\top \hat{\Sigma}^{-1}\mathbf{1}_n/\mathbf{1}_n^\top \hat{\Sigma}^{-1}\mathbf{1}_n$ and $c_1 = (T-2)(T-n-2)/((T-n-1)(T-n-4))$. Economically, the rule adds the sample global minimum variance portfolio as an investable fund to diversify the estimation risk. Mathematically, the rule is analytically obtained by maximizing the expected utility function in the choice set of convex combinations of the sample tangent and sample global minimum variance portfolios. Zhou (2008) applies the result to analyze the fundamental law of active management in the presence of estimation errors.

Despite the innovations in new rules, DeMiguel et al. (2008) and Duchin and Levy (2009) show that, due to estimation errors, existing theory-based portfolio strategies are not as good as we once thought, and the estimation window needed for them to beat the naive $1/n$ rule (that invests equally across n risky assets) is too long to be of practical use. Tu and Zhou (2008) explain that the better performance of the $1/n$ rule is because the capital asset pricing model (CAPM) is a good model, and because it holds the market portfolio when the sum of the betas being equal to one and when the CAPM holds. Although the $1/n$ rule has zero variance, it is biased. Tu and Zhou (2008) further combine it with the Kan and Zhou three-fund rule,

$$\hat{\boldsymbol{x}}_s = (1-v)\frac{1}{n}\mathbf{1}_n + v\hat{\boldsymbol{x}}^{\mathrm{KZ}}.$$

This rule makes an optimal tradeoff between bias and variance. As it turns out, this seems to be the best rule to date, and is the only theory-based investment strategy that works well across models and real data sets. In contrast, existing rules, as reported in Tu and Zhou (2008), can lose money on a risk-adjusted basis when applied to real data sets.

5 Bayesian approach

Zellner and Chetty (1965) provide a general framework for handling the parameter uncertainty problem in the Bayesian set-up. Brown (1976, 1978), Klein and Bawa (1976), and Jorion (1986) are earlier Bayesian studies on the portfolio selection problem that account for parameter uncertainty. Bawa et al. (1979) provide an extensive survey of the related research in the 1970s. In the Bayesian framework, the unknown parameters, such as $\boldsymbol{x}^*$, are regarded as random variables. Given a prior on the distribution of $\boldsymbol{\mu}$ and Σ, uncertainty about the parameters is summarized by the posterior distribution of the parameters conditional on the observed data. Integrating out the parameters over the posterior distribution gives the so-called predictive distribution for future asset returns. The optimal portfolio weights are then obtained by maximizing the expected utility under the predictive distribution, i.e.,

$$\begin{aligned}\hat{\boldsymbol{x}}^{\text{Bayes}} &= \operatorname{argmax}_{\boldsymbol{x}} \int_{\boldsymbol{r}^{T+1}} u(\boldsymbol{x}) p(\boldsymbol{r}^{T+1}|R_T)\, d\boldsymbol{r}^{T+1} \\ &= \operatorname{argmax}_{\boldsymbol{x}} \int_{\boldsymbol{r}^{T+1}} \int_{\mu} \int_{\Sigma} u(\boldsymbol{x}) p(\boldsymbol{r}^{T+1}, \boldsymbol{\mu}, \Sigma|R_T)\, d\boldsymbol{\mu}\, d\Sigma\, d\boldsymbol{r}^{T+1},\end{aligned}$$

where $u(\boldsymbol{x})$ is the utility of holding a portfolio $\boldsymbol{x}$ at time $T+1$, $p(\boldsymbol{r}^{T+1}|R_T)$ is the predictive density, and

$$p(\boldsymbol{r}^{T+1}, \boldsymbol{\mu}, \Sigma|R_T) = p(\boldsymbol{r}^{T+1}|\boldsymbol{\mu}, \Sigma, R_T) p(\mu, \Sigma|R_T),$$

where $p(\boldsymbol{\mu}, \Sigma|R_T)$ is the posterior density of $\boldsymbol{\mu}$ and Σ. Thus the Bayesian approach maximizes the average expected utility over the distribution of the parameters. It is worth mentioning that the Bayesian framework handles non-quadratic objectives, of which Harvey et al. (2003) provide an excellent example of a utility with higher moments. For a complete understanding of Bayesian decision making, Berger (1985) and Rachev et al. (2008b) provide an introduction.

Brown (1976), Klein and Bawa (1976), and Stambaugh (1997) show that, under the diffuse prior on $\boldsymbol{\mu}$ and Σ,

$$p_0(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-\frac{n+1}{2}},$$

the predictive distribution of the asset return follows a multivariate t-distribution if the returns are normal. It then follows that the Bayesian solution to the optimal portfolio weights has the same formula as for $\boldsymbol{x}^*$, except that now the parameters have to be replaced by their predictive moments, yielding

$$\hat{\boldsymbol{x}}^{\text{Bayes}} = \frac{1}{\gamma}\left(\frac{T-n-2}{T+1}\right)\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}.$$

The Bayesian weights differ from the unbiased estimator of the weights only by a small factor of $T/(T+1)$.

The real advantage of the Bayesian approach is its ability to utilize an investor's prior. Kandel and Stambaugh (1996) show that incorporating return predictability plays an important role in portfolio decisions. In particular, they show that a statistically insignificant predictability can be economically significant. To see how the idea works, consider the case in which one predictive variable, say the dividend yield denoted as DY, is used in the predictive regression such that

$$\boldsymbol{r}^t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 DY_{t-1} + \boldsymbol{v}_t.$$

To reflect a certain degree of predictability, we use a simple normal prior for $\boldsymbol{\mu}_1$,

$$p_0(\boldsymbol{\mu}_1) \propto \mathcal{N}\left[\boldsymbol{\mu}_1^p, \sigma_P^2\left(\frac{1}{s_{RR}^2}\Sigma_{RR}\right)\right],$$

where $\boldsymbol{\mu}_1^p$ is the prior mean on $\boldsymbol{\mu}_1$, σ_P^2 measures the uncertainty about predictability, and s_{RR}^2 is the average of the diagonal elements of Σ_{RR} (the asset covariance matrix). Assuming a suitable prior on all other parameters, we have a complete prior that is informative on predictability. Xia (2001) adapts this framework in continuous time, while Barberis (2000) extends this framework for long-run investment. Avramov (2004) generalizes it further to account for both parameter and predictability uncertainties, in addition to the model mispricing below.

Pástor (2000) and Pástor and Stambaugh (2000) introduce a class of interesting priors that reflect an investor's degree of belief in an asset pricing model. To see how this class of priors is formed, consider the market model regression

$$\boldsymbol{r}^t = \boldsymbol{\alpha} + B\boldsymbol{r}_{mt} + \boldsymbol{u}_t,$$

where $\boldsymbol{u}_t$ are the residuals with zero means and a non-singular covariance matrix Σ. To allow for mispricing uncertainty, Pástor (2000) and Pástor and Stambaugh (2000) specify the prior distribution of $\boldsymbol{\alpha}$ as a normal distribution conditional on Σ,

$$\boldsymbol{\alpha}|\Sigma \sim N\left[0, \sigma_{\boldsymbol{\alpha}}^2\left(\frac{1}{s_\Sigma^2}\Sigma\right)\right],$$

where s_Σ^2 is a suitable prior estimate for the average diagonal elements of Σ. The above alpha-sigma link is also explored by MacKinlay and Pástor (2000) in the classical framework. The magnitude of $\sigma_{\boldsymbol{\alpha}}$ represents an investor's level of uncertainty about the pricing ability of a given model. When $\sigma_{\boldsymbol{\alpha}} = 0$, the investor believes dogmatically in the model and there is no mispricing uncertainty. On the other hand, when $\sigma_{\boldsymbol{\alpha}} = \infty$, the investor believes that the pricing model is entirely useless. Therefore, the framework provides a useful tool to assess the value of an asset pricing model. They find that asset pricing models can serve as useful priors for making much better investment decisions.

The well-known Black and Litterman (1992) model is an ad hoc Bayesian approach that is based on an equilibrium asset pricing relation. While this model has many applications (e.g., Fabozzi et al. (2006) incorporate various trading strategies into the model), it differs from the usual Bayesian analysis by ignoring the data-generating process. However, Zhou (2009) shows that this problem can be fixed, so that allowing Bayesian learning to exploit all available information—the market views, the investor's proprietary views as well as the data. In this way, practitioners can combine insights from the Black-Litterman model with the data to generate potentially more reliable trading strategies and more robust portfolios.

Tu and Zhou (2009) explore yet another class of priors based on economic objectives. Intuitively, an investor should have some idea about the range of the optimal solution (though it can be wide). Combining this with the first-order conditions of the utility maximization problem, the associated prior on the parameters can be formed, which is objective-based. As it turns out, this type of priors perform very well as compared with other priors.

Finally, it should be pointed out that, although the above studies assume a true data-generating process, the Bayesian framework is well suited to address portfolio strategies robust to models and data-generating processes. Wang (2004) analyzes Bayesian shrinkage estimators. Tu and Zhou (2004) and Kacperczyk (2007) demonstrate how uncertainties about data-generating processes can be incorporated to make robust Bayesian investment decisions. The Bayesian model averaging is a quite general decision tool under various uncertainties, of which the parameter and model uncertainties are special cases. Cremers (2002) and Avramov (2004) apply it to aggregate predictive information across non-nested models.

6 Open issues

We conclude the paper by asking some of the important open questions. While the literature on robust portfolios from operations research is abundant and insightful, there is in general a lack of empirical studies on how the methods work with real world applications. Under practical constraints on positions, leverage, factor exposures and the like, the well understood mean-variance portfolio theory is widely applied as part of the quantitative strategies by portfolio managers (see quantitative investment books, Grinold and Kahn 2000; Chincarini and Kim 2006; Qian et al. 2007; and Fabozzi et al. 2007). As more studies focus on the empirical aspects, it will be a matter of time before the robust portfolios reviewed here will become as indispensable as the mean-variance framework.

There should also be more interactions between the operations research and finance literatures. Most of the financial studies focus on the underlying data-generating processes and their estimations, and often ignore the important ideas and methods from operations research from which they can benefit greatly. On the other hand, much of the operations research literature takes as given the estimates in applications, without modeling or taking into account of the economic forces that drive the data-generating process. In addition, it rarely studies simultaneously both the optimal estimation and the associated robust strategies. That is, rather than estimate the parameters first and then compute the optimal portfolio, we can realize this in one step by estimating the portfolio weights directly (see Aït-Sahalia and Brandt 2001 and DeMiguel and Nogales 2008 for example). Moreover, the literature seldom makes full use of available asset pricing models. By reviewing both strands of literature together, we hope there will be more future studies that can combine insights from both areas.

A final issue is that existing studies from both operations research and finance are primarily focused on the one-period model. In practice, the multi-period decision has to be considered for time-varying investment opportunities and business cycles. There are virtually no robust portfolio studies in the multi-period set-up with a large number of assets. This is another important direction waiting for future research.

Acknowledgements The authors are grateful to two anonymous referees, as well as Isaac Kleshchelski and Tan Wang for very helpful comments and suggestions.

References

Ahn, D. H., Boudoukh, J., Richardson, M., & Whitelaw, R. F. (1999). Optimal risk management using options. *Journal of Finance*, *54*, 359–375.

Aït-Sahalia, Y., & Brandt, M. W. (2001). Variable selection for portfolio choice. *Journal of Finance*, *56*, 1297–1351.

Alexander, G. J., & Baptista, A. M. (2002). Economic implications of using a mean-VaR model for portfolio selection: a comparison with mean-variance analysis. *Journal of Economic Dynamics and Control*, *26*, 1159–1193.

Alexander, G. J., & Baptista, A. M. (2004). A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, *50*, 1261–1273.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley.

Ang, A., Bekaert, G., & Liu, J. (2005). Why stocks may disappoint. *Journal of Financial Economics*, *76*, 471–508.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*, 203–228.

Avramov, D. (2004). Stock predictability and asset pricing models. *Review of Financial Studies*, *17*(3), 699–738.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, *55*, 225–264.

Basak, S., & Shapiro, A. (2001). Value-at-risk management: optimal policies and asset prices. *Review of Financial Studies*, *14*, 371–405.

Bawa, V. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, *2*, 95–121.

Bawa, V. S., Brown, S. J., & Klein, R. W. (1979). *Estimation risk and optimal portfolio choice*. Amsterdam: North-Holland.

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, *23*, 769–805.

Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, *25*, 1–13.

Ben-Tal, A., Margalit, T., & Nemirovski, A. N. (2002). Robust modeling of multi-stage portfolio problems. In H. Frenk, K. Roos, T. Terlaky, & S. Zhang (Eds.), *High performance optimization* (pp. 303–328). Dordrecht: Kluwer.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies*, *4*(2), 315–342.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*, 28–43.

Boyle, P. P., Siu, T. K., & Yang, H. (2002). Risk and probability measures. *Risk*, *15*(7), 53–57.

Brandt, M. W. (2004, forthcoming). Portfolio choice problems. *Handbook of financial econometrics*, Ait-Sahalia, Y. & L. P. Hansen (Eds.).

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, *45*, 21–58.

Brown, S. J. (1976). *Optimal portfolio choice under uncertainty*. Ph.D. dissertation, University of Chicago.

Brown, S. J. (1978). The portfolio choice problem: comparison of certainty equivalence and optimal Bayes portfolios. *Communications in Statistics—Simulation and Computation*, *7*, 321–334.

Brown, D. B., & Sim, M. (2008, forthcoming). Satisfying measures for analysis of risky positions. *Management Science*.

Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, *18*(3), 853–877.

Ceria, S., & Stubbs, R. A. (2006). Incorporating estimation errors into portfolio selection: robust portfolio construction. *Journal of Asset Management*, *7*(2), 109–127.

Chincarini, L. B., & Kim, D. (2006). *Quantitative equity portfolio management*. New York: McGraw-Hill.

Chopra, V. K., & Ziemba, W. T. (1993). The effects of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, *19*(2), 6–11.

Costa, O. L. V., & Paiva, A. C. (2002). Robust portfolio selection using linear-matrix inequalities. *Journal of Economic Dynamics and Control*, *26*(6), 889–909.

Cremers, K. J. M. (2002). Stock return predictability: a Bayesian model selection perspective. *Review of Financial Studies*, *15*(4), 1223–1249.

Delage, E., & Ye, Y. (2008, forthcoming). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*.

Dembo, R., & Rosen, D. (2000). The practice of portfolio replication. *Algo Research Quarterly*, *3*(2), 11–22.

DeMiguel, V., & Nogales, F. J. (2008, forthcoming). Portfolio selection with robust estimation. *Operations Research*.

DeMiguel, V., Garlappi, L., & Uppal, R. (2008, forthcoming). Optimal versus naive diversification: how inefficient is the $1/N$ portfolio strategy? *Review of Financial Studies*.

Duchin, D., & Levy, H. (2009, forthcoming). Markowitz versus the Talmudic portfolio diversification strategies. *Journal of Portfolio Management*.

El Ghaoui, L., & Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, *18*, 1035–1064.

El Ghaoui, L., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, *51*, 543–556.

Elliott, R. J., & Siu, T. K. (2008, forthcoming). On risk minimizing portfolios under a Markovian regime-switching Black-Scholes economy. *Annals of Operations Research*.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer.
Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). The legacy of modern portfolio theory. *Journal of Investing*, *11*(3), 7–22.
Fabozzi, F., Focardi, S. M., & Kolm, P. N. (2006). Incorporating trading strategies in the Black-Litterman framework. *Journal of Trading*, *Spring*, 28–36.
Fabozzi, F. J., Kolm, P. N., Pachamanova, D., & Focardi, S. M. (2007). *Robust portfolio optimization and management*. Hoboken: Wiley.
Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, *6*(4), 429–447.
Föllmer, H., Schied, A., & Weber, S. (2008). A robust preference and robust portfolio choice, Preprint.
Frittelli, M., & Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, *26*(7), 1473–1486.
Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: a multi-prior approach. *Review of Financial Studies*, *20*(1), 41–81.
Goldfarb, D., & Iyengar, G. (2003a). Robust convex quadratically constrained programs. *Mathematical Programming*, *97*, 495–515.
Goldfarb, D., & Iyengar, G. (2003b). Robust portfolio selection problems. *Mathematics of Operations Research*, *28*, 1–38.
Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management* (2nd ed.). New York: McGraw-Hill.
Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew and kurtosis preferences. *Review of Financial Studies*, *21*(2), 889–935.
Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, *59*(3), 667–686.
Hall, J. A., Brorsen, B. W., & Irwin, S. H. (1989). The distribution of futures prices: a test of the stable Paretian and mixture of normals hypothesis. *Journal of Financial and Quantitative Analysis*, *24*, 105–116.
Hansen, L. P., & Sargent, T. J. (2008). *Robustness*. Princeton: Princeton University Press.
Harvey, C. R., Liechty, J., Liechty, M., & Müller, P. (2003). Portfolio selection with higher moments. Working Paper, Duke University.
Holton, G. (1997). Subjective value-at-risk. *Financial Engineering News*, August.
Hong, Y., Tu, J., & Zhou, G. (2007). Asymmetries in stock returns: statistical tests and economic evaluation. *Review of Financial Studies*, *20*, 1547–1581.
Huang, D., Zhu, S. S., Fabozzi, F. J., & Fukushima, M. (2008). Portfolio selection with uncertain exit time: a robust CVaR approach. *Journal of Economic Dynamics and Control*, *32*(2), 594–623.
Huber, R. J. (2003). *Robust statistics*. New York: Wiley.
Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*, 544–554.
Jobson, J. D., Korkie, B., & Ratti, V. (1979). Improved estimation for Markowitz portfolios using James-Stein type estimators. *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, *41*, 279–284.
Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, *21*, 279–292.
Jorion, P. (1991). Bayesian and CAPM estimators of the means: implications for portfolio selection. *Journal of Banking and Finance*, *15*, 717–727.
Jorion, P. (2006). *Value-at-risk: the new benchmark for managing financial risk* (3rd ed.). New York: McGraw-Hill.
Kacperczyk, M. (2007). Asset allocation under distribution uncertainty. Working paper, University of British Columbia.
Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*, 621–656.
Kandel, S., & Stambaugh, R. F. (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance*, *51*, 385–424.
Klein, R. W., & Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, *3*, 215–231.
Kleshchelskiy, I., & Vincentz, N. (2007). Robust equilibrium yield curves. Working paper, Washington University.
Knight, F. (1921). *Risk, uncertainty and profit*. Boston: Houghton Mifflin.
Konno, H., & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, *37*, 519–531.
Konno, H., & Koshizuka, T. (2005). Mean-absolute deviation model. *IIE Transactions*, *37*, 893–900.
Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*, 603–621.

Litterman, B. (2003). *Modern investment management: an equilibrium approach*. New York: Wiley.
Lutgens, F. (2004). *Robust portfolio optimization*. Ph.D. dissertation, University of Maastricht.
Lutgens, F., & Schotman, P. (2006). Robust portfolio optimization with multiple experts. Working paper, University of Maastricht.
Maccheroni, F., Marinacci, M., & Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, *74*(6), 1447–1498.
MacKinlay, A. C., & Pástor, Ľ. (2000). Asset pricing models: implications for expected returns and portfolio selection. *Review of Financial Studies*, *13*, 883–916.
Maenhout, P. J. (2004). Robust portfolio rules and asset pricing. *Review of Financial Studies*, *17*, 951–983.
Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*, 77–91.
Markowitz, H. M. (1959). *Portfolio selection: efficient diversification of investments*. New Jersey: Wiley.
Maronna, R. A., Martin, D. R., & Yohai, V. J. (2006). *Robust statistics: theory and methods*. New York: Wiley.
Merton, R. (2003). Thoughts on the future: theory and practice of investment management. *Financial Analyst Journal*, *59*(1), 17–23.
Meucci, A. (2005). *Risk and asset allocation*. Berlin: Springer.
Mori, H. (2004). Finite sample properties of estimators for the optimal portfolio weight. *Journal of Japan Statistical Society*, *34*, 27–46.
Natarajan, K., Pachamanova, D., & Sim, M. (2008a). Incorporating asymmetric distribution information in robust value-at-risk optimization. *Management Science*, *54*(3), 573–585.
Natarajan, K., Sim, M., & Uichanco, J. (2008b). Tractable robust expected utility and risk models for portfolio optimization, Working paper, National University of Singapore.
Natarajan, K., Pachamanova, D., & Sim, M. (2009, forthcoming). Constructing risk measures from uncertainty sets. *Operations Research*.
Ortobelli, S., Rachev, S. T., Stoyanov, S., Fabozzi, F. J., & Biglova, A. (2005). The proper use of risk measures in portfolio theory. *International Journal of Theoretical and Applied Finance*, *8*(8), 1107–1133.
Pástor, Ľ. (2000). Portfolio selection and asset pricing models. *Journal of Finance*, *55*, 179–223.
Pástor, Ľ., & Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, *56*, 335–381.
Peel, D., & McLachlan, G. J. (2000). Robust mixture modeling using the t distribution. *Statistics and Computing*, *10*, 339–348.
Pinar, M. C. (2007). Robust scenario optimization based on downside-risk measure for multi-period portfolio selection. *OR Spectrum*, *29*(2), 295–309.
Pinar, M. C., & Tütüncü, R. (2005). Robust profit opportunities in risk financial portfolios. *Operations Research Letters*, *33*(4), 331–340.
Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, *55*(1), 98–112.
Qian, E., Hua, R., & Sorensen, E. (2007). *Quantitative equity portfolio management: modern techniques and applications*. New York: Chapman & Hall.
Rachev, S. T., Ortobelli, S., Stoyanov, S., Fabozzi, F. J., & Biglova, A. (2008a). Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and Applied Finance*, *11*(1), 19–54.
Rachev, S. T., Hsu, J. S. J., Bagasheva, B. S., & Fabozzi, F. J. (2008b). *Bayesian methods in finance*. New Jersey: Wiley.
Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*, 21–41.
Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, *26*(7), 1443–1471.
Rockafellar, R. T., Uryasev, S., & Zabarankin, M. (2006). Optimality conditions in portfolio analysis with generalized deviation measures. *Mathematical Programming*, *108*, 515–540.
Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica*, *20*(3), 431–449.
Rustem, B., Becker, R. G., & Marty, W. (2000). Robust min-max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control*, *24*, 1591–1621.
Shen, R., & Zhang, S. (2008). Robust portfolio selection based on a multi-stage scenario tree. *European Journal of Operational Research*, *191*(3), 864–887.
Siu, T. K., Tong, H., & Yang, H. (2001). Bayesian risk measures for derivatives via random Esscher transform. *North American Actuarial Journal*, *5*(3), 78–91.
Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics*, *45*, 285–331.
Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (vol. *1*, pp. 197–206). Berkeley: University of California Press.

Steinbach, M. C. (2001). Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Review*, *43*(1), 31–85.

Stoyanov, S. V., Rachev, S. T., & Fabozzi, F. J. (2007). Optimal financial portfolios. *Applied Mathematical Finance*, *14*(5), 403–438.

Strzalecki, T. (2008). Axiomatic foundations of multiplier preferences. Working paper, Harvard University.

Ter Horst, J., De Roon, F., & Werker, B. (2006), Incorporating estimation risk in portfolio choice. In P. Duffhues & L. Renneboog (Eds.), *Advances in corporate finance and asset pricing*, Emerald Group Publishing.

Tu, J., & Zhou, G. (2004). Data-generating process uncertainty: what difference does it make in portfolio decisions? *Journal of Financial Economics*, *72*, 385–421.

Tu, J., & Zhou, G. (2008). Being naive about naive diversification: can investment theory be consistently useful? Working paper, Washington University.

Tu, J., & Zhou, G. (2009, forthcoming). Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*.

Tütüncü, R., & Koenig, M. (2004). Robust asset allocation. *Annals of Operations Research*, *132*, 157–187.

Uppal, R., & Wang, T. (2003). Model misspecification and underdiversification. *Journal of Finance*, *58*(6), 2465–2486.

Wang, Z. (2004). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, *18*(2), 673–705.

Whittle, P. (1990). *Risk-sensitive optimal control*. New York: Wiley.

Whittle, P. (1996). *Optimal control: basics and beyond*. New York: Wiley.

Xia, Y. (2001). Learning about predictability: the effect of parameter uncertainty on dynamic asset allocation. *Journal of Finance*, *56*, 205–246.

Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science*, *44*, 673–683.

Zellner, A., & Chetty, V. K. (1965). Prediction and decision problems in regression models from the Bayesian point of view. *Journal of the American Statistical Association*, *60*, 608–616.

Zhou, G. (2008). On the fundamental law of active portfolio management: what happens if our estimates are wrong? *Journal of Portfolio Management*, *34*(3), 26–33.

Zhou, G. (2009, forthcoming) Beyond Black-Litterman: letting the data speak. *Journal of Portfolio Management*.

Zhu, S. S., & Fukushima, M. (2008, forthcoming). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*.