International Review of Economics and Finance 80 (2022) 906–937

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Orthogonal portfolios to assess estimation risk

Luis Chavez-Bedoya, Francisco Rosales *

Graduate School of Business, Universidad ESAN, Lima, Peru

ARTICLE INFO

JEL classification:
G23

Keywords:
Portfolio optimization
Orthogonal portfolios
Estimation risk
Global minimum-variance portfolio
Zero-investment portfolio
Active portfolio management

ABSTRACT

This document presents the various advantages of using portfolio rules composed by linear combinations of the orthogonal components derived from the optimal solution to a linearly constrained mean–variance portfolio optimization problem. We argue that this practice has value in and of itself since it pushes forward the tractability of the out-of-sample performance measure, and the identification of risk sources in the portfolio. This structure is further used to propose new correction schemes based on shrinkage factors that improve out-of-sample performance, and to study its limiting behavior as both the sample size and the number of assets increase. Additionally, our results are compared with those corresponding to the theoretical and implementable three-fund rules of Kan and Zhou (2007) so the benefits of using orthogonal portfolio rules are highlighted.

1. Introduction

Orthogonality is a fundamental concept in both asset pricing and active portfolio management. In asset pricing, two orthogonal portfolios, a minimum-variance "zero-beta" portfolio and the market index "unit-beta" portfolio, are the two determinants of individual expected asset returns, see e.g. Roll (1980), MacKinlay (1995), Campbell et al. (1997) and Ferson and Siegel (2015). In active portfolio management, the optimal portfolio consists of combining optimally two orthogonal portfolios, a "unit-beta" benchmark and a "zero-beta" active portfolio (Gibbons et al., 1989 and Grinold & Kahn, 2000). Besides the unit or zero exposure to "beta", a characteristic feature of orthogonal portfolios is the absence of linear correlation between their corresponding returns.

It is well known that the optimal portfolio derived from the mean–variance framework of Markowitz (1952) suffers a decline in out-of-sample performance when its main inputs—the mean vector and the covariance matrix of asset returns—are replaced by their sample estimates into the optimal solution. Thus, an important part of the finance literature has produced strategies and methods to understand and reduce estimation risk in portfolio optimization.[1] In this context, the work of Kan and Zhou (2007), extended in Kan et al. (2016) and Kan et al. (2018), is of particular importance since *i)* the portfolios and correction schemes proposed by the authors are successful in theory and practice and thus provide an appropriate baseline; and *ii)* the out-of-sample performance measure introduced in the previous references can be directly linked to the definition of orthogonality used in this document.

* Corresponding author.
E-mail address: frosales@esan.edu.pe (F. Rosales).

[1] Estimation risk was studied, for example, in Best and Grauer (1991), Britten-Jones (1999), Brown (1976), Frost and Savarino (1986), Jobson and Korkie (1980), Jorion (1985) and Lai et al. (2011), and different approaches have been devised to increase out-of-sample performance using the mean–variance framework. Some of these approaches include: using a Bayesian setting (Bawa et al., 1979; Brown, 1976; Frost & Savarino, 1986; Harvey et al., 2010), replacing the ML estimator of the mean with a Bayes–Stein one (Jorion, 1986, 1991), using shrinkage estimators for the covariance matrix (Ledoit & Wolf, 2003, 2004), incorporating resampling techniques (Michaud & Michaud, 2008), exploiting asset pricing models (Black & Litterman, 1992; MacKinlay & Pástor, 2000; Pástor, 2000), constraining the norm of the optimal portfolio (DeMiguel, Garlappi, Nogales, et al., 2009), adding extra constraints (Jagannathan & Ma, 2003), using the global minimum-variance portfolio (Basak et al., 2009; Bodnar & Okhrin, 2008; Bodnar et al., 2018; Frahm, 2008; Jagannathan & Ma, 2003; Kempf & Memmel, 2006; Okhrin & Schmid, 2006) or a portfolio with no estimation risk (DeMiguel, Garlappi, & Uppal, 2009; Duchin & Levy, 2009) and incorporating timing strategies (Kirby & Ostdiek, 2012).

https://doi.org/10.1016/j.iref.2022.02.046
Received 14 February 2020; Received in revised form 9 July 2021; Accepted 13 February 2022
Available online 12 March 2022
1059-0560/

Kan and Zhou (2007) defined out-of-sample performance as the expectation of the usual mean–variance objective function evaluated in portfolio weights that depend on the sample estimates of the parameters of the multivariate normal distribution governing excess returns. These portfolio weight functions—also known as portfolio rules— are linear combinations of optimal portfolios derived from the mean–variance problem without estimation risk when plug-in estimators are used.[2] Moreover, the coefficients of the linear combination are selected such that the expected out-of-sample performance is maximized. In this framework, the usual two-fund rule of the plug-in tangency portfolio (S) and the risk-free asset (F) is dominated by a three-fund rule (KZ) that also incorporates the plug-in minimum-variance portfolio (G).[3] This framework was then extended in: *i)* (Zhou, 2008) to active portfolio management; and *ii)* (Tu & Zhou, 2011) to the combination of the $1/N$ portfolio with one of four portfolio rules (plug-in estimator S, KZ rule, Jorion, 1986 rule and the MacKinlay & Pástor, 2000 rule). Recently, Kan et al. (2016) and Kan et al. (2018) introduced the out-of-sample Sharpe ratio as an alternative performance measure and evaluate several portfolio rules that are fully invested in risky assets.

In this paper, we explore the role of portfolio orthogonality in estimation risk. Inspired on a constrained mean–variance portfolio optimization problem, we start by defining orthogonality in the context of portfolio rules. Second, we analyze the three-fund rule of Kan and Zhou (2007) to give an interpretation of their optimal coefficients and to build alternative portfolio rules based on orthogonal portfolios. Third, we study the out-of-sample performance of related portfolio rules based on orthogonal plug-in estimators of some optimal portfolios, and compare their performance with that of the non-orthogonal combinations of Kan and Zhou (2007). Fourth, we assess the out-of-sample performance of portfolio rules under the limiting conditions of Bodnar et al. (2018) in which both the number of assets (n) and the sample size (t) tend to infinity and $\frac{n}{t} = c \in (0, 1)$.

We start by defining orthogonality between two portfolio rules if *i)* their expected out-of-sample covariance is zero; and *ii)* the covariance of their expected out-of-sample returns is zero. Additionally, two portfolio rules are normalized under an attribute if almost surely, one portfolio rule (A) has unit exposure and the other one (B) has zero exposure to the common linear attribute. Using this definition, we study normalized orthogonal portfolio rules consisting on the plug-in estimators of the two uncorrelated components of the optimal solution to a mean–variance portfolio problem subject to a linear constraint enforcing a target exposure in an attribute; but, most importantly, the tangency portfolio can be replicated by a particular linear combination of this two "optimal'' portfolios. Moreover, this definition is handy to illustrate that the optimal weights to S and G in the KZ rule act as if aiming to (imperfectly) replicate the optimal linear combination of A and B when the chosen attribute is the position in risky assets; and, therefore, KZ rule exhibits a lower theoretical expected out-of-sample performance than the proposed orthogonal rule.

In general, if the optimal coefficients of the linear combination of orthogonal portfolio rules A and B are found à la (Kan & Zhou, 2007), then they become functions of unknown quantities.[4] It turns out that when the unknown parameters of the optimal coefficients are replaced by their estimators, expected out-of-sample performance declines dramatically specially for small data samples. This fact creates the necessity for additional adjustments, and orthogonality also helps to provide an alternative to generate extra deterministic shrinkage coefficients for the orthogonal portfolio rules. We were able to develop additional correction schemes—based only on the number of assets and data observations—such that, under certain conditions on the Sharpe ratio of B, the out-of-sample performance of our new portfolio rules is greater than the one of the implementable version of Kan and Zhou (2007). Moreover, applying the conditions of Bodnar et al. (2018), and using orthogonality properties, we are able to determine whether the introduced correction factors are optimal or not.

With the previous developments we aim to showcase the usefulness of orthogonal portfolios when dealing with estimation risk. Our results show that this approach is useful to: *i)* give more structure to identify the sources of estimation risk, *ii)* provide new portfolio rules that are theoretically superior to KZ, *iii)* propose new correction schemes to improve out-of-sample performance of our orthogonal portfolio rules, *iv)* analyze the pertinence and impact of estimating the exposure of the tangency portfolio to a particular attribute; and, *v)* compute the limiting performance of portfolio rules which, in turn, helps to determine if the proposed correction schemes lead to optimal holdings at least under the limiting conditions of Bodnar et al. (2018).

This document is organized as follows. In Section 2 we present definitions and terminology of modern portfolio theory and orthogonality that are required to present portfolio rules. In Section 3 we study the out-of-sample performance of the proposed rules—distinguishing the cases when optimal weights are known and when they have to be estimated—and compared them with KZ three-fund rule. The limiting behavior of portfolio rules is investigated in detail in Section 4. Finally, Section 5 closes the document with some conclusions and recommendations for future research.

2. Orthogonality, mean–variance optimization and portfolio rules

In this section, we introduce some definitions and concepts required to solve a portfolio optimization problem under estimation risk. We start by defining portfolio orthogonality and presenting the optimal solution to a constrained portfolio problem when the parameters of the multivariate normal distribution are known. We then introduce the concepts of portfolio rules and plug-in estimators, and extend the definition of orthogonality to portfolios that are functions of observed data.

[2] Technically, a portfolio rule is a function that transforms t sample excess returns observations of n risky assets into portfolio weights. In the case of i.i.d. normal returns it is enough to consider functions of the maximum-likelihood estimators of their mean vector and covariance matrix. This portfolio rules are random vectors by construction since the sample return observations are generated by a stochastic process.

[3] This finding was originally presented in Kan and Zhou (2007) and later discussed, extended and refined in Kan et al. (2016) and Kan et al. (2018). Also, this finding illustrates that, when estimation risk is considered, the optimal investment rule benefits from the additional diversification provided by G, which has a non-zero optimal coefficient, and the shrinkage of S is given by an optimal coefficient between zero and one.

[4] In particular, we refer to the Sharpe ratio of the portfolio estimated by B and the theoretical exposure of the tangency portfolio to the attribute.

2.1. Orthogonality and mean–variance optimization

Let $\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} \neq k\boldsymbol{\pi}$, $k \in \mathbb{R} \setminus \{0\}$ denote the excess return random vector of risky assets, where $\boldsymbol{\Sigma}$ is non-singular, and $\boldsymbol{\pi} \neq \mathbf{0}$ denotes an n-dimensional vector. If $\boldsymbol{\omega}_P \in \mathbb{R}^n$ represents the proportions of the n risky assets in portfolio P, the excess return of such portfolio $Z_P = \boldsymbol{\omega}_P^\top \boldsymbol{Z}$ follows $Z_P \sim \mathcal{N}(\mu_P, \sigma_P^2)$, where $\mu_P(\boldsymbol{\omega}_P; \boldsymbol{\mu}) := \boldsymbol{\omega}_P^\top \boldsymbol{\mu}$, and $\sigma_P^2(\boldsymbol{\omega}_P; \boldsymbol{\Sigma}) := \boldsymbol{\omega}_P^\top \boldsymbol{\Sigma} \boldsymbol{\omega}_P$. Further, let $\boldsymbol{\omega}_P \neq \mathbf{0}$, such that $\theta_P(\boldsymbol{\omega}_P; \boldsymbol{\mu}, \boldsymbol{\Sigma}) := \mu_P/\sigma_P$ is a well-defined Sharpe ratio (SR) for portfolio P. We also introduce attribute Π, and a corresponding vector of exposures of each risky asset to Π given by $\boldsymbol{\pi}$. Finally, the resulting exposure of portfolio $\boldsymbol{\omega}_P$, to this attribute is measured by $\boldsymbol{\omega}_P^\top \boldsymbol{\pi}$.

Definition 1. Two portfolios A and B, with weights $\boldsymbol{\omega}_A \neq \mathbf{0}$ and $\boldsymbol{\omega}_B \neq \mathbf{0}$, are orthogonal if the excess returns of A and B are uncorrelated, i.e., $\boldsymbol{\omega}_A^\top \boldsymbol{\Sigma} \boldsymbol{\omega}_B = 0$. Also, portfolios A and B are normalized under attribute $\boldsymbol{\pi}$ if: i) A has unit exposure to the attribute, i.e., $\boldsymbol{\omega}_A^\top \boldsymbol{\pi} = 1$ and, ii) B has zero exposure to the attribute, i.e., $\boldsymbol{\omega}_B^\top \boldsymbol{\pi} = 0$.

The normalized orthogonal structure of Definition 1 is widely used in active portfolio management because the benchmark has minimum risk with unit exposure to an attribute (beta) and the optimal active portfolio maximizes value added with zero exposure to the same attribute. For more details the reader is referred to Grinold and Kahn (2000) and the definition of optimal orthogonal portfolios in Ferson and Siegel (2015).

If we introduce a risk-aversion parameter $\gamma \in \mathbb{R}_+$, the performance of a particular portfolio $\boldsymbol{\omega}_P$ can be evaluated by a utility metric given by

$$\mathcal{U}(\boldsymbol{\omega}_P; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma) := \boldsymbol{\omega}_P^\top \boldsymbol{\mu} - \frac{\gamma}{2}\, \boldsymbol{\omega}_P^\top \boldsymbol{\Sigma} \boldsymbol{\omega}_P. \tag{1}$$

Under the previous setting, consider the investors' general portfolio optimization problem, hereafter **GPOP**, as:

$$\max_{\boldsymbol{\omega}} \left\{ \mathcal{U}(\boldsymbol{\omega}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma) : \boldsymbol{\omega}^\top \boldsymbol{\pi} = \varrho \right\}, \quad \varrho \in \mathbb{R}. \tag{2}$$

The optimal solution of **GPOP** clearly depends on both the choice of $\boldsymbol{\pi}$ and the value taken by parameter ϱ; but, most importantly, it has a very close relationship with the normalized orthogonal portfolios presented in Definition 1. Without loss of generality, the optimal solution of **GPOP** can always be expressed as a linear combination of two particular orthogonal portfolios that are also normalized under attribute $\boldsymbol{\pi}$. Proposition 1 formalizes this relationship.

Proposition 1. *Given an attribute with individual asset exposures $\boldsymbol{\pi}$ and a target exposure ϱ, the optimal solution to* **GPOP** *follows*

$$\boldsymbol{\omega}^*(\boldsymbol{\pi}, \varrho) = \varrho \boldsymbol{\omega}_A(\boldsymbol{\pi}) + \boldsymbol{\omega}_B(\boldsymbol{\pi}), \tag{3}$$

where portfolios A *and* B *are orthogonal and normalized under $\boldsymbol{\pi}$ according to Definition 1.*

Proof. Trivial. Consider the following portfolios:

$$\begin{aligned}
\boldsymbol{\omega}_A(\boldsymbol{\pi}) &= \operatorname{argmin}_{\boldsymbol{\omega}} \left\{ \boldsymbol{\omega}^\top \boldsymbol{\Sigma} \boldsymbol{\omega} : \boldsymbol{\omega}^\top \boldsymbol{\pi} = 1 \right\}, \\
&= \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\pi}}{\boldsymbol{\pi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\pi}}, \qquad (4) \\
\boldsymbol{\omega}_B(\boldsymbol{\pi}) &= \operatorname{argmax}_{\boldsymbol{\omega}} \left\{ \boldsymbol{\mu}^\top \boldsymbol{\omega} - \frac{\gamma}{2} \boldsymbol{\omega}^\top \boldsymbol{\Sigma} \boldsymbol{\omega} : \boldsymbol{\omega}^\top \boldsymbol{\pi} = 0 \right\}, \\
&= \frac{1}{\gamma} \left(\boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\pi} \boldsymbol{\pi}^\top \boldsymbol{\Sigma}^{-1}}{\boldsymbol{\pi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\pi}} \right) \boldsymbol{\mu}, \qquad (5)
\end{aligned}$$

and note that they satisfy $\boldsymbol{\omega}_A(\boldsymbol{\pi})^\top \boldsymbol{\Sigma} \boldsymbol{\omega}_B(\boldsymbol{\pi}) = 0$, hence they are orthogonal under $\boldsymbol{\pi}$. Plug in the results in (3) and verify that the resulting expression corresponds to the optimal solution to **GPOP**, obtained by the first order conditions of a strictly concave objective function. □

Proposition 1 implies that the optimal solution to **GPOP** can be expressed as a linear combination of two particular orthogonal portfolios that are normalized under $\boldsymbol{\pi}$: A, with weights given by (4), minimizes variance with unit exposition to the attribute; and, B, with weights given by (5), maximizes expected utility in (1) but with zero exposure to the same attribute. Also note that, since the excess returns of these portfolios are uncorrelated, it follows that $\mathcal{U}(\boldsymbol{\omega}^*(\boldsymbol{\pi}, \varrho)) = \mathcal{U}(\varrho \boldsymbol{\omega}_A(\boldsymbol{\pi})) + \mathcal{U}(\boldsymbol{\omega}_B(\boldsymbol{\pi}))$, and $Z^*(\boldsymbol{\pi}, \varrho) = \varrho Z_A(\boldsymbol{\pi}) + Z_B(\boldsymbol{\pi})$ with $\mathrm{Cov}(Z_A(\boldsymbol{\pi}), Z_B(\boldsymbol{\pi})) = 0$, where $Z_A(\boldsymbol{\pi}) = \boldsymbol{\omega}_A(\boldsymbol{\pi})^\top \boldsymbol{Z}$ and $Z_B(\boldsymbol{\pi}) = \boldsymbol{\omega}_B(\boldsymbol{\pi})^\top \boldsymbol{Z}$.

The solution to the unconstrained version of **GPOP** is the tangency portfolio S with weights $\boldsymbol{\omega}_S = \gamma^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. Moreover, note that portfolio S can be expressed as a particular combination of the two normalized orthogonal portfolios of Proposition 1 because

$$\boldsymbol{\omega}_S = \varrho_S \boldsymbol{\omega}_A(\boldsymbol{\pi}) + \boldsymbol{\omega}_B(\boldsymbol{\pi}) \tag{6}$$

when $\varrho_S = \gamma^{-1} \boldsymbol{\pi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, i.e., the optimal solution of **GPOP** yields the unconstrained solution when the target exposure to attribute $\boldsymbol{\pi}$, i.e. ϱ, matches the exposure of portfolio S to the same attribute.[5] Based on Eq. (6), there are infinite pairs of normalized orthogonal portfolios such that a particular linear combination of them yields the tangency portfolio.

[5] The set of the optimal portfolios to **GPOP**, written as a function of ϱ and for a given $\boldsymbol{\pi}$, corresponds to a parabola when projected in the mean–variance space. Also note that portfolio S is located in that parabola for any $\boldsymbol{\pi}$, portfolio B is always in its vertex and portfolio A is, in general, outside the parabola.

Table 1
Portfolio optimization variants for certain values of ϱ when $\boldsymbol{\pi} = \mathbf{1}$.

j	Portfolio name	ϱ	Weight $\boldsymbol{\omega}_j$	Sqr. SR. θ_j^2	Utility $\mathcal{U}(\boldsymbol{\omega}_j, \cdot)$
S	Tangency	free	$\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$	$\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$	$\frac{\theta_S^2}{2\gamma}$
G	Minimum-variance A	1	$\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$	$\frac{(\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}$	$\mu_G - \frac{\gamma}{2}\sigma_G^2$
$\tilde{\mathrm{G}}$	Minimum-variance B	$\frac{1}{\sigma_G^2}$	$\boldsymbol{\Sigma}^{-1}\mathbf{1}$	θ_G^2	$\frac{\mu_G}{\sigma_G^2} - \frac{\gamma}{2}\frac{1}{\sigma_G^2}$
H	Hedge	0	$\frac{1}{\gamma}\boldsymbol{R}\boldsymbol{\mu}$	$\boldsymbol{\mu}^\top\boldsymbol{R}\ \boldsymbol{\mu}$	$\frac{\theta_H^2}{2\gamma}$

An important attribute to consider is $\boldsymbol{\pi} = \mathbf{1}$ because it represents the position in risky assets. Under this attribute, the linear constraint of **GPOP** forces the optimal portfolio to be invested in risky assets in a proportion equal to ϱ, and ω_A in (4) represents the global minimum-variance portfolio. Moreover, the solutions of **GPOP** under different values of $\varrho \in \mathbb{R}$ when $\boldsymbol{\pi} = \mathbf{1}$ are well documented in the literature, see e.g. Meucci (2005) and Grinold and Kahn (2000). Some important portfolios related to $\boldsymbol{\pi}$ are listed in Table 1, where

$$\boldsymbol{R} := \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{7}$$

was used to ease the notation.

We stress that portfolios G and $\tilde{\mathrm{G}}$ in Table 1 are the optimal solutions to minimization problems with respect to the variance component inside the utility function (1) alone, in the same fashion as in (4), and should be considered separately. The distinction between these minimum-variance portfolios strives in that G is fully invested in risky assets, while $\tilde{\mathrm{G}}$ has a net position in risky assets equal to $1/\sigma_G^2$, where $\sigma_G^2 = (\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}$. Portfolio H of the same table is sometimes called "hedge portfolio", it has zero net investment in risky assets, and its squared Sharpe ratio, θ_H^2, matches the squared slope of the asymptote of the minimum variance frontier. This portfolio can also be written as $\boldsymbol{\omega}_H = \gamma^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mu_G\mathbf{1})$, which will be a representation of use in Section 3.2.

Remark 1. Given $\boldsymbol{\pi} = \mathbf{1}$ and exposure ϱ, the solution to the **GPOP** can be expressed as

$$\boldsymbol{\omega}^*(\mathbf{1}, \varrho) = \varrho\boldsymbol{\omega}_G + \boldsymbol{\omega}_H. \tag{8}$$

Thus, portfolio $\boldsymbol{\omega}^*$ given by (8) is a linear combination of two orthogonal portfolios normalized under $\boldsymbol{\pi} = \mathbf{1}$: G and H with corresponding weights ϱ and 1. It is easy to see that the value of ϱ that maximizes the utility in (1) is $\varrho_S = \frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}$ and that this selection yields $\boldsymbol{\omega}^*(\mathbf{1}, \varrho_S) = \boldsymbol{\omega}_S$, meaning that portfolio S can be replicated with G and H, as

$$\boldsymbol{\omega}_S = \left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right)\boldsymbol{\omega}_G + \boldsymbol{\omega}_H. \tag{9}$$

Notice that ϱ_S corresponds to the total position on risky assets of the tangency portfolio, or, equivalently, the exposure of portfolio S to attribute $\boldsymbol{\pi} = \mathbf{1}$. Additionally, it is straightforward to verify that $\mathcal{U}(\boldsymbol{\omega}_S) = \mathcal{U}(\varrho_S\boldsymbol{\omega}_G) + \mathcal{U}(\boldsymbol{\omega}_H) = \frac{\theta_G^2}{2\gamma} + \frac{\theta_H^2}{2\gamma} = \frac{\theta_S^2}{2\gamma}$. All these results follow directly from Proposition 1.

Remark 2. Consider Remark 1. If instead of portfolio G we use portfolio $\tilde{\mathrm{G}}$ of Table 1, it is also possible to replicate portfolio S as

$$\boldsymbol{\omega}_S = \left(\frac{1}{\gamma}\mu_G\right)\boldsymbol{\omega}_{\tilde{G}} + \boldsymbol{\omega}_H. \tag{10}$$

It is important to note that portfolios $\tilde{\mathrm{G}}$ and H are orthogonal and normalized under an attribute $\boldsymbol{\pi} = \sigma_G^2\mathbf{1}$, and $\mathcal{U}(\boldsymbol{\omega}_S) = \mathcal{U}(\gamma^{-1}\mu_G\boldsymbol{\omega}_{\tilde{G}}) + \mathcal{U}(\boldsymbol{\omega}_H) = \frac{\theta_G^2}{2\gamma} + \frac{\theta_H^2}{2\gamma} = \frac{\theta_S^2}{2\gamma}$. We stress that expressions (9) and (10) are only two possible linear combinations of orthogonal portfolios that reach tangency performance.

2.2. Portfolio rules and plug-in estimators

Suppose the investor has available $t > n$ periods of observed i.i.d. multivariate normal excess returns $\boldsymbol{\Phi} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t\}$ and wants to form a portfolio for period $t+1$. Then, a portfolio rule $\hat{\boldsymbol{\omega}}$ is defined as a function of the historical excess returns data. Since maximum likelihood (ML) estimates of the sample mean and covariance matrices, $\hat{\boldsymbol{\mu}} = \frac{1}{t}\sum_{i=1}^t \boldsymbol{Z}_i$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{t}\sum_{i=1}^t(\boldsymbol{Z}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{Z}_i - \hat{\boldsymbol{\mu}})^\top$, are sufficient statistics to characterize the process $\boldsymbol{\Phi}$, it will be enough to consider portfolio rules that are functions of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.

In the context of Roll (1980), portfolio B could pay the role of an orthogonal "zero-beta" portfolio (although not a minimum-variance one) for a generally inefficient "market index" given by portfolio A. Under these choices, for a portfolio j we have that the mispricing is $\alpha_j = \mu_j - \mu_B - \pi_j\mu_A$, where π_j is the exposure of portfolio j to attribute $\boldsymbol{\pi}$. Consequently, $\alpha_j = 0$ if either j belongs to the aforementioned parabola or $\boldsymbol{\pi} = \boldsymbol{\mu}$. In the last case, portfolio A will be a multiple of the tangency portfolio S and B will be the risk-free asset.

Following Kan and Zhou (2007), and conditional on the weights of portfolio rule P being chosen as $\hat{\boldsymbol{\omega}}_P$, a measure of out-of-sample performance of $\hat{\boldsymbol{\omega}}_P = \hat{\boldsymbol{\omega}}_P(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, can be defined as

$$\mathcal{U}(\hat{\boldsymbol{\omega}}_P; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma) = \hat{\boldsymbol{\omega}}_P^\top \boldsymbol{\mu} - \frac{\gamma}{2}\, \hat{\boldsymbol{\omega}}_P^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_P. \tag{11}$$

Since $\mathcal{U}(\hat{\boldsymbol{\omega}}_P, \cdot)$ is a random variable, it is natural to use its expectation, $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_P, \cdot)]$, to evaluate the performance of portfolio rules. We will refer to $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_P)]$ as the expected out-of-sample utility of (a random) portfolio rule P.[6]

Next, we extend the definition of orthogonality and normalization under an attribute in the context of portfolio rules.

Definition 2. Two portfolio rules A and B, with weights $\hat{\boldsymbol{\omega}}_A \neq \mathbf{0}$ a.s. and $\hat{\boldsymbol{\omega}}_B \neq \mathbf{0}$ a.s. and both functions of $\boldsymbol{\Phi} = \{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_t\}$, are orthogonal if: i) the expected covariance of their excess returns is zero, i.e., $\mathbb{E}[\hat{\boldsymbol{\omega}}_A^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_B] = 0$; and, ii) the covariance of their expected excess returns is zero, i.e., $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_A^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_B^\top \boldsymbol{\mu}) = 0$.[7] Also, portfolio rules A and B are normalized under attribute $\boldsymbol{\pi}$ if: iii) A has unit exposure to the attribute, i.e., $\boldsymbol{\pi}^\top \hat{\boldsymbol{\omega}}_A = 1$ a.s.; and, iv) B has zero exposure to the attribute, i.e., $\boldsymbol{\pi}^\top \hat{\boldsymbol{\omega}}_B = 0$ a.s.

Consider a set of portfolio rules of the form

$$\hat{\boldsymbol{\omega}}_{(A,B)}(a, b) := a \times \hat{\boldsymbol{\omega}}_A + b \times \hat{\boldsymbol{\omega}}_B, \quad a, b \in \mathbb{R} \tag{12}$$

where $\hat{\boldsymbol{\omega}}_A$ and $\hat{\boldsymbol{\omega}}_B$ are two different portfolio rules. In addition, we stress that if the symbol Ã denotes a portfolio rule in the left-hand side of (12), then the associated constant in the right-hand side is denoted as $\tilde{a}$. The aforementioned set of portfolio rules are, in general, invested in three-funds ($\hat{\boldsymbol{\omega}}_A$, $\hat{\boldsymbol{\omega}}_B$ and the risk-free asset) unless we impose a constraint such that $\hat{\boldsymbol{\omega}}_{(A,B)}(a, b)$ is fully invested in risky assets.

If portfolio rules $\hat{\boldsymbol{\omega}}_A$ and $\hat{\boldsymbol{\omega}}_B$ in (12) satisfy condition i) of Definition 2, i.e., $\mathbb{E}[\hat{\boldsymbol{\omega}}_A^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_B] = 0$, then the expected out-of-sample utility of $\hat{\boldsymbol{\omega}}_{(A,B)}(a, b)$ can be decomposed as

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{(A,B)}(a, b))] = \mathbb{E}[\mathcal{U}(a \times \hat{\boldsymbol{\omega}}_A)] + \mathbb{E}[\mathcal{U}(b \times \hat{\boldsymbol{\omega}}_B)]. \tag{13}$$

Hence, for the chosen performance measure and portfolio rule structure, orthogonality will be manifested basically for the possibility to have the decomposition in (13).

We stress that not all portfolio rules as defined in (12) need to be linear combinations of orthogonal portfolios. Throughout this document we avoid the non-orthogonal representation of portfolio rules, and give preference to the orthogonal decomposition as described in Definition 2. Moreover, unless stated otherwise, we will consider an attribute with $\boldsymbol{\pi} = \mathbf{1}$; and, most of our selected portfolio rules just replace $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ in the expressions of the portfolios in Table 1, i.e., they are the plug-in estimators of the corresponding portfolios.

It is important to mention that the theoretical results of the paper will hold for any attribute $\boldsymbol{\pi}$, so it will be enough to replace $\mathbf{1}$ with $\boldsymbol{\pi}$ in both the portfolio rules and in parameters representing mean, variance and squared Sharpe ratios. The analysis to determine appropriate attributes is beyond the scope of the paper.

Let $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ be the corresponding plug-in estimators of portfolios $\boldsymbol{\omega}_G$ and $\boldsymbol{\omega}_H$ of Table 1, i.e., $\hat{\boldsymbol{\omega}}_G = \frac{\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}$ and $\hat{\boldsymbol{\omega}}_H = \frac{1}{\gamma} \hat{\boldsymbol{R}} \hat{\boldsymbol{\mu}}$. If we introduce a portfolio rule that is a linear combination of the previous portfolio rules, it can be represented as $\hat{\boldsymbol{\omega}}_{(\mathrm{G,H})}(g, h) = g \times \hat{\boldsymbol{\omega}}_G + h \times \hat{\boldsymbol{\omega}}_H$. In Proposition 2, we will show in detail that portfolio rules $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ are orthogonal and normalized under $\boldsymbol{\pi} = \mathbf{1}$, i.e., they satisfy the conditions of Definition 2: $\mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_H] = 0$, $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_H^\top \boldsymbol{\mu}) = 0$, and $\mathbf{1}^\top \hat{\boldsymbol{\omega}}_G = 1$, $\mathbf{1}^\top \hat{\boldsymbol{\omega}}_H = 0$.

Proposition 2. *Let $\hat{\theta}_H^2 = \hat{\boldsymbol{\mu}}^\top \hat{\boldsymbol{R}} \hat{\boldsymbol{\mu}}$ be the plug-in estimator of the squared Sharpe ratio of portfolio* H, *and let $f : \mathbb{R} \to \mathbb{R}$ denote a function such that $\mathbb{E}[f(\hat{\theta}_H^2)] < \infty$ and $n > 3$. Then, according to Definition 2, portfolio rules $\hat{\boldsymbol{\omega}}_G$ and $f(\hat{\theta}_H^2) \times \hat{\boldsymbol{\omega}}_H$ are orthogonal and normalized under attribute $\boldsymbol{\pi} = \mathbf{1}$.*

Proof. See Appendix B. □

Note that the implications of Proposition 2 are twofold. On the one hand it shows that the plug-in rules $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ are orthogonal and normalized under $\boldsymbol{\pi}$ (recall that non-random portfolios G and H in Table 1 are orthogonal and normalized); but on the other hand, it highlights that normalized orthogonality is preserved when portfolio rule $\hat{\boldsymbol{\omega}}_H$ is scaled by a function of $\hat{\theta}_H^2$.

We close this section by presenting three additional examples of portfolio rules. First, we consider $\hat{\boldsymbol{\omega}}_{(\tilde{\mathrm{G}},\mathrm{H})}(\tilde{g}, h) = \tilde{g} \times \hat{\boldsymbol{\omega}}_{\tilde{\mathrm{G}}} + h \times \hat{\boldsymbol{\omega}}_H$ which represents the linear combination of $\hat{\boldsymbol{\omega}}_{\tilde{\mathrm{G}}} = \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$—the plug-in estimator of $\tilde{\mathrm{G}}$ of Table 1—and $\hat{\boldsymbol{\omega}}_H$. In Appendix C, we prove that, according to Definition 2, $\hat{\boldsymbol{\omega}}_{\tilde{\mathrm{G}}}$ and $\hat{\boldsymbol{\omega}}_H$ are orthogonal but not normalized under $\boldsymbol{\pi} = \mathbf{1}$ since $\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \neq 1$ in general. Second we consider $\hat{\boldsymbol{\omega}}_{(\tilde{\mathrm{G}},\tilde{\mathrm{H}})}(\tilde{g}, \tilde{h}) = \tilde{g} \times \hat{\boldsymbol{\omega}}_{\tilde{\mathrm{G}}} + \tilde{h} \times \hat{\boldsymbol{\omega}}_{\tilde{\mathrm{H}}}$, where

$$\hat{\boldsymbol{\omega}}_{\tilde{\mathrm{H}}} = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1}). \tag{14}$$

[6] Note that when $\mathcal{U}$ is applied to $\hat{\boldsymbol{\omega}}_P$, i.e., $\mathcal{U}(\hat{\boldsymbol{\omega}}_P)$, it can be interpreted as the expected utility conditional on $\hat{\boldsymbol{\omega}}_P$ as portfolio weights. Therefore, $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_P)]$ is just a performance measure of portfolio rule P and, for simplicity, we call it expected out-of-sample utility.

[7] These conditions imply that portfolio rules A and B will be orthogonal in the sense that the covariance of it out-of-sample excess returns, $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_A^\top \boldsymbol{Z}_{t+1}, \hat{\boldsymbol{\omega}}_B^\top \boldsymbol{Z}_{t+1})$, is equal to zero.

Portfolio rule $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ resembles $\hat{\boldsymbol{\omega}}_H$; however, in Appendix C we prove that portfolio rules $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ and $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ satisfy $\mathbb{E}[\hat{\boldsymbol{\omega}}_{\tilde{G}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_{\tilde{H}}] = 0$ but $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_{\tilde{G}}^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{\tilde{H}}^\top \boldsymbol{\mu}) \neq 0$ and are therefore not orthogonal. Moreover, the aforementioned rules are not normalized because $\mathbf{1}^\top \hat{\boldsymbol{\omega}}_{\tilde{G}} \neq 1$ and $\mathbf{1}^\top \hat{\boldsymbol{\omega}}_{\tilde{H}} \neq 0$ in general.[8] Finally, we introduce portfolio rule $\hat{\boldsymbol{\omega}}_{(G,\breve{H})}(g,\breve{h}) = g \times \hat{\boldsymbol{\omega}}_G + \breve{h} \times \hat{\boldsymbol{\omega}}_{\breve{H}}$, where

$$\hat{\boldsymbol{\omega}}_{\breve{H}} = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}^{-1} \left(\hat{\boldsymbol{\mu}} - \frac{\mu_G}{\sigma_G^2} \frac{t}{t-n-2} \frac{\mathbf{1}}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \right). \tag{15}$$

Portfolio rule $\hat{\boldsymbol{\omega}}_{\breve{H}}$ resembles $\hat{\boldsymbol{\omega}}_H$; but, also in Appendix C we show that portfolio rules $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_{\breve{H}}$ are not orthogonal because $\mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_{\breve{H}}] = 0$ but $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{\breve{H}}^\top \boldsymbol{\mu}) \neq 0$. Finally, these two portfolio rules are not normalized because $\mathbf{1}^\top \hat{\boldsymbol{\omega}}_{\breve{H}} \neq 0$ in general; but, interestingly, $\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\omega}}_{\breve{H}}] = 0$.[9]

3. Theoretical and implementable portfolio rules

In this section, we propose three-fund rules constructed by optimally combining orthogonal portfolio rules, and compare their structure with the three-fund rule of Kan and Zhou (2007). We extend the discussion from the theoretical case, i.e., when the coefficients are known, to the implementable case, i.e., when the coefficients are not known and need to be estimated.

3.1. Theoretical portfolio rules

When the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and need to be estimated from the data by $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ respectively, the well-known two-fund rule that combines the risk-free asset and the tangency portfolio can be improved.[10] In particular, this can be done using three-fund rule Q that optimally combines orthogonal and normalized plug-in portfolio rules $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ and has weights given by

$$\hat{\boldsymbol{\omega}}_{(G,H)}(g_Q^*, h_Q^*) = g_Q^* \times \hat{\boldsymbol{\omega}}_G + h_Q^* \times \hat{\boldsymbol{\omega}}_H, \tag{16}$$

$$g_Q^* = \left[\frac{t-n-1}{t-2} \right] \left(\frac{1}{\gamma} \frac{\mu_G}{\sigma_G^2} \right), \tag{17}$$

$$h_Q^* = \left[\frac{(t-n)(t-n-3)}{t(t-2)} \right] \left(\frac{\theta_H^2}{\theta_H^2 + \frac{n-1}{t}} \right). \tag{18}$$

Coefficients g_Q^* and h_Q^* were selected optimally (and independently from each other following Proposition 2) using expected out-of-sample utility, i.e., g_Q^* maximizes $\mathbb{E}[\mathcal{U}(g_G \times \hat{\boldsymbol{\omega}}_G)]$ and h_Q^* maximizes $\mathbb{E}[\mathcal{U}(h_H \times \hat{\boldsymbol{\omega}}_H)]$. To ease the presentation, hereafter we will use the shorthand notation $\hat{\boldsymbol{\omega}}_Q := \hat{\boldsymbol{\omega}}_{(G,H)}(g_Q^*, h_Q^*)$ to denote portfolio rule Q.

Recall that (9) shows that in the absence of uncertainty we obtain $\boldsymbol{\omega}_S = \frac{1}{\gamma} \frac{\mu_G}{\sigma_G^2} \boldsymbol{\omega}_G + \boldsymbol{\omega}_H$. From (18) we have $0 \leq h_Q^* \leq 1$, implying that $\hat{\boldsymbol{\omega}}_Q$ always has a long position in $\hat{\boldsymbol{\omega}}_H$ but it invest less aggressively in such plug-in portfolio than $\boldsymbol{\omega}_S$ does on $\boldsymbol{\omega}_H$. The decision of going long or short in $\hat{\boldsymbol{\omega}}_G$ depends on the sign of μ_G or equivalently on the sign of the total position in risky assets of portfolio S. Moreover, $|g_Q^*| \leq \frac{1}{\gamma} \frac{|\mu_G|}{\sigma_G^2}$ so $\hat{\boldsymbol{\omega}}_Q$ invests less aggressively in $\hat{\boldsymbol{\omega}}_G$ compared to the investment of $\boldsymbol{\omega}_S$ in $\boldsymbol{\omega}_G$.

Assuming the values of g_Q^* and h_Q^* in (17) and (18) are known and $t > n+3$, the expected out-of-sample utility of this rule is

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_Q)] = \left[\frac{t-n-1}{t-2} \right] \frac{\theta_G^2}{2\gamma} + \left[\frac{(t-n)(t-n-3)}{(t-2)(t-n-1)} \right] \left(\frac{\theta_H^2}{\theta_H^2 + \frac{n-1}{t}} \right) \frac{\theta_H^2}{2\gamma}. \tag{19}$$

Q based on orthogonal portfolios (G,H)

The first term of (19) corresponds to $\mathbb{E}[\mathcal{U}(g_Q^* \times \hat{\boldsymbol{\omega}}_G)]$ and the second one to $\mathbb{E}[\mathcal{U}(h_Q^* \times \hat{\boldsymbol{\omega}}_H)]$. Also, notice that as $t \to \infty$, $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_Q)]$ approaches $\theta_G^2/2\gamma + \theta_H^2/2\gamma$. The latter is equal to the expected utility of the tangency portfolio S because $\theta_S^2 = \theta_G^2 + \theta_H^2$.

As a second case, we introduce $\mathcal{M}$ as the portfolio rule investing optimally in orthogonal but not normalized portfolio rules $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ and $\hat{\boldsymbol{\omega}}_H$ (the plug-in estimators of portfolios $\tilde{G}$ and H in Table 1). This portfolio rule is given by

$$\hat{\boldsymbol{\omega}}_{\mathcal{M}} := \hat{\boldsymbol{\omega}}_{(\tilde{G},H)}(\tilde{g}_{\mathcal{M}}^*, h_{\mathcal{M}}^*) = \tilde{g}_{\mathcal{M}}^* \times \hat{\boldsymbol{\omega}}_{\tilde{G}} + h_{\mathcal{M}}^* \times \hat{\boldsymbol{\omega}}_H, \tag{20}$$

where clearly $h_{\mathcal{M}}^*$ is given by (18) and

M based in orthogonal
(G~,H), but not normalized portfolio.

$$\tilde{g}_{\mathcal{M}}^* = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)} \right] \left(\frac{1}{\gamma} \mu_G \right). \tag{21}$$

The expected out-of-sample utility of portfolio rule $\mathcal{M}$ is

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})] = \left[\frac{(t-n-1)(t-n-4)}{(t-2)(t-n-2)} \right] \frac{\theta_G^2}{2\gamma} + \left[\frac{(t-n)(t-n-3)}{(t-2)(t-n-1)} \right] \left(\frac{\theta_H^2}{\theta_H^2 + \frac{n-1}{t}} \right) \frac{\theta_H^2}{2\gamma}. \tag{22}$$

---

[8] We further discuss this case in Section 3, where it is shown that $\hat{\boldsymbol{\omega}}_{(\tilde{G},\tilde{H})}(\tilde{g},\tilde{h})$ is indeed equivalent to the rule proposed in Kan and Zhou (2007).

[9] This property is also satisfied by portfolio rule $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ in (14) as $\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\omega}}_{\tilde{H}}] = 0$ holds.

[10] The expected out-of-sample utility of some portfolio rules can be further improved using an unbiased estimator for the covariance matrix and modifying the portfolios accordingly. Consider for example $\tilde{\boldsymbol{\Sigma}} = \frac{t}{t-1} \hat{\boldsymbol{\Sigma}}$, Bayes estimators or others, see e.g. Ledoit and Wolf (2003), and the discussion in Lai et al. (2011). Since our interest is to improve and interpret Kan and Zhou's fund rules, and the ML estimator fulfill this purpose, we do not elaborate further on other estimators.

Plug-in portfolio rules $\tilde{\mathrm{G}}$ and H are orthogonal according to Definition 2. Then, the first term of (22) corresponds to $\mathbb{E}[\mathcal{U}(\tilde{g}^*_{\mathcal{M}} \times \hat{\boldsymbol{\omega}}_{\tilde{G}})]$ and the second corresponds to $\mathbb{E}[\mathcal{U}(\tilde{h}^*_{\mathcal{M}} \times \hat{\boldsymbol{\omega}}_H)]$.

In the deterministic case, Eq. (10) of Remark 2 shows that $\boldsymbol{\omega}_S = \frac{1}{\gamma}\mu_G \boldsymbol{\omega}_{\tilde{G}} + \boldsymbol{\omega}_H$, then portfolio rule $\mathcal{M}$ invests less aggressively in portfolio rule $\tilde{\mathrm{G}}$ compare to the investment of $\boldsymbol{\omega}_{\tilde{G}}$ in $\boldsymbol{\omega}_S$. Moreover, the first term in expression (22) guarantees the expected out-of-sample utility of $\tilde{g}^*_{\mathcal{M}} \times \hat{\boldsymbol{\omega}}_{\tilde{G}}$ is lower than the one of $g^*_{\mathcal{Q}} \times \hat{\boldsymbol{\omega}}_G$ and thus $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})]$. This difference in expected out-of-sample utility can be explained by normalization because $\hat{\boldsymbol{\omega}}_G$ needs to estimate $n-1$ weights (the n weights add up to one) while $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ needs n estimates. Even though the differences between rules $\mathcal{Q}$ and $\mathcal{M}$ are subtle, they relate to the distinction between normalized and not normalized orthogonal portfolio rules in the presence of estimation risk.

3.2. Kan & Zhou's statement revisited

Kan and Zhou (2007) presented a portfolio rule that preceded $\mathcal{Q}$ and $\mathcal{M}$ and combined the plug in estimators of $\tilde{\mathrm{G}}$ and S to diversify estimation error. This rule is based on non-orthogonal portfolio rules and we will explain and compare its composition and performance in terms of orthogonal rules. Namely, Kan and Zhou (2007) considered the following three-fund rule:

$$\hat{\boldsymbol{\omega}}_{(\tilde{G},S)}(\tilde{g}^*_{\mathcal{KZ}}, s^*_{\mathcal{KZ}}) = \tilde{g}^*_{\mathcal{KZ}} \times \hat{\boldsymbol{\omega}}_{\tilde{G}} + s^*_{\mathcal{KZ}} \times \hat{\boldsymbol{\omega}}_S, \tag{23}$$

$$\tilde{g}^*_{\mathcal{KZ}} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\frac{n}{t}}{\theta_H^2 + \frac{n}{t}}\right)\left(\frac{1}{\gamma}\mu_G\right), \tag{24}$$

KZ based on non orthogonal portfolios (G~, S), but reduce estimation risk

$$s^*_{\mathcal{KZ}} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t}}\right), \tag{25}$$

where $t > n+4$ and coefficients $\tilde{g}^*_{\mathcal{KZ}}$ and $s^*_{\mathcal{KZ}}$ maximize $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{(\tilde{G},S)}(\tilde{g}, s))]$. Hereafter, portfolio (23) will be called $\mathcal{KZ}$, and $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}} := \hat{\boldsymbol{\omega}}_{(\tilde{G},S)}(\tilde{g}^*_{\mathcal{KZ}}, s^*_{\mathcal{KZ}})$ will be used to ease the notation. Assuming the optimal factors are known, the expected out-of-sample utility of this rule is reported to be

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})] = \left[\frac{(t-n-1)(t-n-4)}{(t-2)(t-n-2)}\right]\left(1 - \frac{\frac{n}{t}}{\theta_S^2 + \frac{n}{t}\frac{\theta_S^2}{\theta_H^2}}\right)\frac{\theta_S^2}{2\gamma}. \tag{26}$$

We stress that $\mathbb{E}[\hat{\boldsymbol{\omega}}_{\tilde{G}}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_S] \neq 0$ in general; and that portfolio rules $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ and $\hat{\boldsymbol{\omega}}_S$ are not orthogonal according to Definition 2.

Consider the direct comparison between portfolio rules $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$ and $\hat{\boldsymbol{\omega}}_S$, that is

$$\begin{aligned}\hat{\boldsymbol{\omega}}_{\mathcal{KZ}} = {} & \underbrace{\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\frac{n}{t}}{\theta_H^2 + \frac{n}{t}}\right)\left(\frac{1}{\gamma}\mu_G\right)}_{\tilde{g}^*_{\mathcal{KZ}}} \times \hat{\boldsymbol{\omega}}_{\tilde{G}} \\ & + \underbrace{\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t}}\right)}_{s^*_{\mathcal{KZ}}} \times \hat{\boldsymbol{\omega}}_S.\end{aligned}$$

Here it is clear that $0 < s^*_{\mathcal{KZ}} < 1$, i.e. portfolio rule $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$ always goes long in $\hat{\boldsymbol{\omega}}_S$ shrinking its magnitude. The decision of going long or short in $\hat{\boldsymbol{\omega}}_{\tilde{G}} = \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}$, however, depends on the sign of μ_G. Thus, the presence of $\tilde{g}^*_{\mathcal{KZ}} \neq 0$ in $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ does not provide more intuition, other than the improvement of diversification to mitigate estimation risk. Next, we introduce portfolio $\mathcal{KZ}a$, which aims to mimic the $\mathcal{KZ}$ original portfolio rule. Note that it is possible to express $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$ in (23) as an optimal linear combination of $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ and $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ where the latter was defined in (14). In fact, it is easy to see that $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}} = \hat{\boldsymbol{\omega}}_{\mathcal{KZ}a}$ a.s., by writing the latter as

$$\hat{\boldsymbol{\omega}}_{\mathcal{KZ}a} := \hat{\boldsymbol{\omega}}_{(\tilde{G},\tilde{H})}(\tilde{g}^*_{\mathcal{KZ}a}, \tilde{h}^*_{\mathcal{KZ}a}) = \tilde{g}_{\mathcal{KZ}}a^* \times \hat{\boldsymbol{\omega}}_{\tilde{G}} + \tilde{h}_{\mathcal{KZ}}a^* \times \hat{\boldsymbol{\omega}}_{\tilde{H}}, \tag{27}$$

$$\tilde{g}_{\mathcal{KZ}}a^* = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\mu_G\right), \tag{28}$$

$$\tilde{h}_{\mathcal{KZ}}a^* = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t}}\right). \tag{29}$$

Now let us compare portfolio rules $\hat{\boldsymbol{\omega}}_{\mathcal{M}}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}a}$:

$$\hat{\boldsymbol{\omega}}_{\mathcal{M}} = \tilde{g}^*_{\mathcal{M}} \times \hat{\boldsymbol{\omega}}_{\tilde{G}} + \underbrace{\left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n-1}{t}}\right)}_{h^*_{\mathcal{M}}} \times \hat{\boldsymbol{\omega}}_H \tag{30}$$

$$\hat{\boldsymbol{\omega}}_{\mathcal{KZ}a} = \tilde{g}^*_{\mathcal{KZ}a} \times \hat{\boldsymbol{\omega}}_{\tilde{G}} + \underbrace{\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t}}\right)}_{\tilde{h}^*_{\mathcal{KZ}a}} \times \hat{\boldsymbol{\omega}}_{\tilde{H}}. \tag{31}$$

Note that $\tilde{g}^*_{\mathcal{M}} = \tilde{g}^*_{\mathcal{KZ}a}$, and that the weight $\tilde{h}^*_{\mathcal{KZ}a}$ given to $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ in $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}a}$ is similar, but less aggressive, compared to the weight $h^*_{\mathcal{M}} = h^*_{\mathcal{Q}}$ given to $\hat{\boldsymbol{\omega}}_H$ in both $\hat{\boldsymbol{\omega}}_{\mathcal{M}}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{Q}}$. Even though $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ does not have zero net position in risky assets for all realizations, i.e., it is not normalized; it can be represented as a linear combination of portfolio rules $\hat{\boldsymbol{\omega}}_{\tilde{G}}$ and $\hat{\boldsymbol{\omega}}_S$ resembling closely the zero-investment portfolio rule $\hat{\boldsymbol{\omega}}_H$. Thus, we conclude that the optimal scalars $s^*_{\mathcal{KZ}}$ and $\tilde{g}^*_{\mathcal{KZ}}$ in (23) have the effect of maximizing the expected out-of-sample utility, when attempting to replicate portfolio rule $\mathcal{M}$ in (20).

Inspired in the $\mathcal{KZ}$ portfolio rule, we present portfolio rule $\mathcal{Y}$ as

$$\hat{\boldsymbol{\omega}}_{\mathcal{Y}} := \hat{\boldsymbol{\omega}}_{(G,S)}(g^*_{\mathcal{Y}}, s^*_{\mathcal{Y}}) = g^*_{\mathcal{Y}} \times \hat{\boldsymbol{\omega}}_G + s^*_{\mathcal{Y}} \times \hat{\boldsymbol{\omega}}_S, \tag{32}$$

$$g^*_{\mathcal{Y}} = \left[\frac{t-n-1}{t-2}\right]\left(\frac{\frac{2\theta_H^2}{t-n-2} + \frac{n}{t} + \frac{2\theta_G^2}{t-n-2}}{\theta_H^2 + \frac{n}{t} + \frac{2\theta_G^2}{t-n-2}}\right)\left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right), \tag{33}$$

$$s^*_{\mathcal{Y}} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t} + \frac{2\theta_G^2}{t-n-2}}\right), \tag{34}$$

where $t > n+4$ and coefficients $g^*_{\mathcal{Y}}$ and $s^*_{\mathcal{Y}}$ maximize $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{(G,S)}(g,s))]$. Clearly, portfolio rule $\mathcal{Y}$ optimally combines the plug-in estimators of G and S, and the only difference with $\mathcal{KZ}$ is the choice of G instead of $\tilde{\text{G}}$.

The expected out-of-sample utility of $\mathcal{Y}$ is given by

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] = \left[\frac{t-n-1}{t-2}\right]\frac{\theta_G^2}{2\gamma} + \left[\frac{(t-n-1)(t-n-4)}{(t-2)(t-n-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t} + \frac{2\theta_H^2}{t-n-2}}\right)\frac{\theta_H^2}{2\gamma}. \tag{35}$$

Y based on non orthogonal portfolios (G, S), but normalized

The performance in (35) highlights the presence of a portfolio that together with G satisfies condition *i)* of Definition 2. This new portfolio is the result of the optimal combination of the plug-in estimators of G and S. In fact, it is trivial to show that $\hat{\boldsymbol{\omega}}_{\mathcal{Y}} = \hat{\boldsymbol{\omega}}_{\mathcal{Y}a}$ a.s. such that the latter is an optimal linear combination of non-orthogonal portfolios $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_{\check{H}}$ in (15). More precisely, portfolio rule $\mathcal{Y}a$ is defined as

$$\hat{\boldsymbol{\omega}}_{\mathcal{Y}a} := \hat{\boldsymbol{\omega}}_{(G,\check{H})}(g^*_{\mathcal{Y}a}, \check{h}^*_{\mathcal{Y}a}) = g^*_{\mathcal{Y}a} \times \hat{\boldsymbol{\omega}}_G + \check{h}^*_{\mathcal{Y}a} \times \hat{\boldsymbol{\omega}}_{\check{H}}, \tag{36}$$

Ya based on non orthogonal portfolios (G, H~)

$$g^*_{\mathcal{Y}a} = \left[\frac{t-n-1}{t-2}\right]\left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right), \tag{37}$$

$$\check{h}^*_{\mathcal{Y}a} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t} + \frac{2\theta_G^2}{t-n-2}}\right). \tag{38}$$

Notice that because $\mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_{\check{H}}] = 0$, we have $g^*_{\mathcal{Y}a} = g^*_{\mathcal{Q}}$, and therefore the second term of the expected out-of-sample utility in (35) corresponds to $\mathbb{E}[\mathcal{U}(\check{h}^*_{\mathcal{Y}a} \times \hat{\boldsymbol{\omega}}_{\check{H}})]$. Then, scalars $g^*_{\mathcal{Y}}$ and $s^*_{\mathcal{Y}}$ are such that portfolio rule $\mathcal{Y}$ attempts to replicate $\mathcal{Q}$. Moreover, an important aspect of portfolio rule $\mathcal{Y}a$ is that $\check{h}^*_{\mathcal{Y}a}$ in (38) depends on the squared Sharpe ratio of G, and cannot be determined independently from that portfolio.

In the next proposition we rank theoretical portfolio rules based on their expected out-of-sample utilities.

Proposition 3. *Consider portfolio rules $\hat{\boldsymbol{\omega}}_{\mathcal{Q}}$, $\hat{\boldsymbol{\omega}}_{\mathcal{M}}$, $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{Y}}$ given by (16), (20), (23) and (32), respectively. If $t > n+4$, then their expected out-of-sample utilities satisfy $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$ and $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$.*

Proof. The formula in (26) can be re-expressed as

$$\begin{aligned}\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})] &= \left[\frac{(t-n-1)(t-n-4)}{(t-2)(t-n-2)}\right]\frac{\theta_G^2}{2\gamma} \\ &+ \left[\frac{(t-n-1)(t-n-4)}{(t-2)(t-n-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2 + \frac{n}{t}}\right)\frac{\theta_H^2}{2\gamma}.\end{aligned} \tag{39}$$

When comparing (39) with (19), $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$ holds directly. Noting that the expected out-of-sample utility of $\tilde{h}^*_{\mathcal{KZ}a} \times \hat{\boldsymbol{\omega}}_{\tilde{H}}$ must be equal to the last term of (39), we have that $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}a})]$. Moreover, by direct inspection of (35) and (19), we

Table 2
Coefficients a and b for theoretical portfolio rules. Portfolios in columns A and B are associated with the ones of Table 1.

P	A	B	a	b
$\mathcal{KZ}$	$\tilde{\mathrm{G}}$	S	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\frac{n}{t}}{\theta_H^2+\frac{n}{t}}\right)\left(\frac{1}{\gamma}\mu_G\right)$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n}{t}}\right)$
$\mathcal{KZ}a$	$\tilde{\mathrm{G}}$	$\tilde{\mathrm{H}}$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\mu_G\right)$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n}{t}}\right)$
$\mathcal{Y}$	G	S	$\left[\frac{t-n-1}{t-2}\right]\left(\frac{\frac{2\theta_H^2}{t-n-2}+\frac{n}{t}+\frac{2\theta_G^2}{t-n-2}}{\theta_H^2+\frac{n}{t}+\frac{2\theta_G^2}{t-n-2}}\right)\left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right)$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n}{t}+\frac{2}{t-n-2}\theta_G^2}\right)$
$\mathcal{Y}a$	G	$\check{\mathrm{H}}$	$\left[\frac{t-n-1}{t-2}\right]\left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right)$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n}{t}+\frac{2}{t-n-2}\theta_G^2}\right)$
$\mathcal{Q}$	G	H	$\left[\frac{t-n-1}{t-2}\right]\left(\frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\right)$	$\left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n-1}{t}}\right)$
$\mathcal{M}$	$\tilde{\mathrm{G}}$	H	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\mu_G\right)$	$\left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\theta_H^2}{\theta_H^2+\frac{n-1}{t}}\right)$

have $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})]$. Finally, we have $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$ because

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] - \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})] = \frac{(t-n-1)\theta_G^2}{(t-n-2)(t-2)}\left(1 - \frac{t-n-4}{t-n-2} \times \frac{\theta_H^2}{\theta_H^2+\frac{n}{t}} \times \frac{\theta_H^2}{\theta_H^2+\frac{n}{t}+\frac{2\theta_G^2}{t-n-2}}\right). \quad \square$$

Proposition 3 shows that portfolio rule $\hat{\boldsymbol{\omega}}_{\mathcal{Q}}$ dominates $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$ in terms of expected out-of-sample utility.[11] The theoretical implication of this result is that using sample portfolios inspired on the optimal solution of **GPOP** in (2), i.e., orthogonal and normalized portfolio rules $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$, one theoretically reduces estimation risk in a greater magnitude than using two non-orthogonal portfolio rules belonging to the sample mean–variance efficient frontier, i.e. any linear combination of $\hat{\boldsymbol{\omega}}_{\tilde{\mathrm{G}}}$ and $\hat{\boldsymbol{\omega}}_S$. This conclusion also holds with the combination of $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_S$ (portfolio rule $\mathcal{Y}$), but in this case using a normalized minimum-variance portfolio has a positive effect because $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$.

Proposition 3 also shows that $\mathcal{M}$ and $\mathcal{Y}$ cannot be ranked in terms of expected out-of-sample utility. Thus, it is again clear that using G instead of $\tilde{\mathrm{G}}$ benefits $\mathcal{Y}$; but, this is not enough to outperform $\mathcal{M}$ since the optimal performance associated with H is always greater than the one associated with $\check{\mathrm{H}}$. Then, we can conclude that normalization helps when improving expected out-of-sample utility, and the optimal linear combinations of the non-orthogonal components of $\mathcal{KZ}$ and $\mathcal{Y}$ attempt to replicate $\hat{\boldsymbol{\omega}}_H$ by devising portfolio rules $\hat{\boldsymbol{\omega}}_{\tilde{\mathrm{H}}}$ and $\hat{\boldsymbol{\omega}}_{\check{\mathrm{H}}}$ satisfying condition *i)* of Definition 2. Additionally, optimal holdings in the aforementioned components yield an expected out-of-sample utility lower than that of the optimal position in $\hat{\boldsymbol{\omega}}_H$.

Even though portfolio rules $\mathcal{Q}$ and $\mathcal{KZ}$ might produce similar expected-out-of-sample utilities, their interpretation in terms of estimation risk is the main advantage of the former. Portfolio rule $\mathcal{KZ}$ is not strictly diversifying estimation risk when adding an extra portfolio to the plug-in estimator of S; instead, it tries to replicate the correct orthogonal components of the tangency portfolio when the attribute is the position on risky assets. Consequently, the estimation of portfolio S is equivalent to the estimation of normalized orthogonal portfolios under some particular attribute, and this fact is the main idea behind theoretical portfolio rule $\mathcal{Q}$.

In Table 2, we present a summary of the theoretical portfolio rules defined in (16), (20), (23), (27), (32) and (36). For notational ease, we define $\mathcal{V}^P$ as the expected out-of-sample utility of a particular portfolio rule P. For example, if P $= \mathcal{Q}$, then

$$\mathcal{V}^{\mathcal{Q}} = \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})]. \tag{40}$$

The same principle is applied for any P $\in \{\mathcal{KZ}, \mathcal{KZ}a, \mathcal{Y}, \mathcal{Y}a, \mathcal{Q}, \mathcal{M}\}$ in Table 2.

3.3. Implementable portfolio rules

The portfolio rules presented in Sections 3.1 and 3.2 are not implementable in practice because coefficients a and b of Table 2 depend on population parameters μ_G, σ_G^2 and θ_H^2, which are unknown and would need to be estimated.[12] Hence, in this subsection we define a set of "implementable" portfolio rules as

$$\hat{\boldsymbol{\omega}}_{(A,B)}(\hat{a},\hat{b}) := \hat{a} \times \hat{\boldsymbol{\omega}}_A + \hat{b} \times \hat{\boldsymbol{\omega}}_B, \quad \hat{a},\hat{b} \in \mathbb{R}, \tag{41}$$

where $\hat{a}$ and $\hat{b}$ are random coefficients (based on estimators of population parameters) and $\hat{\boldsymbol{\omega}}_A$ and $\hat{\boldsymbol{\omega}}_B$ are, for example, the plug-in estimators of portfolios $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ presented in Table 1.

We can apply (41) directly to the theoretical portfolio rules in Table 2 by considering $\hat{a}$ and $\hat{b}$ to be the random coefficients when the unknown parameters in columns a and b are replaced by $\hat{\mu}_G$, $\hat{\sigma}_G^2$ and $\hat{\theta}_H^2$. For example, the implementable version of portfolio rule $\mathcal{Q}$ of Table 2, named $\mathcal{Q}_I$, is

$$\hat{\boldsymbol{\omega}}_{\mathcal{Q}_I} := \hat{\boldsymbol{\omega}}_{(G,H)}(\hat{g}_{\mathcal{Q}}^*, \hat{h}_{\mathcal{Q}}^*) = \hat{g}_{\mathcal{Q}}^* \times \hat{\boldsymbol{\omega}}_G + \hat{h}_{\mathcal{Q}}^* \times \hat{\boldsymbol{\omega}}_H, \tag{42}$$

[11] This result is not equivalent to Second-order Stochastic Dominance (SSD) of portfolio rule $\mathcal{Q}$ with respect to $\mathcal{KZ}$. We can find counter examples in which the cumulative distribution functions of the out-of-sample excess returns of $\hat{\boldsymbol{\omega}}_{\mathcal{Q}}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}$, namely $\hat{\boldsymbol{\omega}}_{\mathcal{Q}}^\top \boldsymbol{Z}_{t+1}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}^\top \boldsymbol{Z}_{t+1}$, do not satisfy the conditions for SSD.

[12] We will employ only plug-in estimators for μ_G, σ_G^2 and θ_H^2, i.e., we replace $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with their ML estimators in the corresponding expressions.

$$\hat{g}^*_{Q} = \left[\frac{t-n-1}{t-2}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\right), \tag{43}$$

$$\hat{h}^*_{Q} = \left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\hat{\theta}^2_H}{\hat{\theta}^2_H+\frac{n-1}{t}}\right). \tag{44}$$

The expected out-of-sample utility of portfolio rule $\mathcal{Q}_I$ in (42) is $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}_I})]$; but, in general it will be different than the sum of the individual performances of $\hat{g}^*_Q \times \hat{\boldsymbol{\omega}}_G$ and $\hat{h}^*_Q \times \hat{\boldsymbol{\omega}}_H$ because $\mathbb{E}[(\hat{g}^*_Q\ \hat{h}^*_Q)\times\hat{\boldsymbol{\omega}}^\top_G\boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_H] \neq 0$, i.e., portfolio rules $\hat{g}^*_Q \times \hat{\boldsymbol{\omega}}_G$ and $\hat{h}^*_Q \times \hat{\boldsymbol{\omega}}_H$ are not orthogonal based on Definition 2.

Under the implementable setting, consider additional portfolio rule $\mathcal{QS}_I$ given by

$$\hat{\boldsymbol{\omega}}_{\mathcal{QS}_I} := \hat{\boldsymbol{\omega}}_{(G,H)}(\hat{g}^*_{QS}, \hat{h}^*_{QS}) = \hat{g}^*_{QS} \times \hat{\boldsymbol{\omega}}_G + \hat{h}^*_{QS} \times \hat{\boldsymbol{\omega}}_H, \tag{45}$$

$$\hat{g}^*_{QS} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\right), \tag{46}$$

$$\hat{h}^*_{QS} = \left[\frac{(t-n-7)(t-n-5)(t-n)}{t^2(t-2)}\right]\left(\frac{\hat{\theta}^2_H}{\hat{\theta}^2_H+\frac{n-1}{t}}\right), \tag{47}$$

when $t > n+7$. The rationale behind portfolio rule $\mathcal{QS}_I$ is that its expected out-of-sample utility is greater than the one of $\mathcal{Q}_I$ when we assume that the last factors of $\hat{h}^*_Q$ in (44) and $\hat{h}^*_{QS}$ in (46) are both equal to $\theta^2_H/(\theta^2_H + \frac{n-1}{t})$ (for the proof see Appendix D). Therefore, portfolio rule $\mathcal{QS}_I$ includes extra correction factors (lower than one and depending only on t and n) for both $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ and these extra factors aim to reduce the impact of the estimator error of random coefficients by avoiding aggressive allocations. More specifically, when we compare the coefficients of portfolio rules $\mathcal{QS}_I$ and $\mathcal{Q}_I$ we obtain $\hat{g}^*_{QS} = \frac{(t-n-4)}{t}\hat{g}^*_Q$ and $\hat{h}^*_{QS} = \frac{(t-n-5)(t-n-7)}{(t-n-3)t}\hat{h}^*_Q$. Recall that the proof in Appendix D does not show that the expected out-of-sample utility of $\mathcal{QS}_I$ is greater than that of $\mathcal{Q}_I$; however, it presents an interesting method to determine extra shrinking factors that may improve out-of-sample performance of our class of implementable portfolio rules. The comparison between the theoretical and empirical performances of $\mathcal{QS}_I$ and $\mathcal{Q}_I$ will be addressed later in the section.

Motivated by $\mathcal{QS}_I$ in (45), we define an additional implementable portfolio rule $\mathcal{QS}a_I$ as:

$$\hat{\boldsymbol{\omega}}_{\mathcal{QS}a_I} := \hat{\boldsymbol{\omega}}_{(G,H)}(\hat{g}^*_{QSa}, \hat{h}^*_{QSa}) = \hat{g}^*_{QSa} \times \hat{\boldsymbol{\omega}}_G + \hat{h}^*_{QSa} \times \hat{\boldsymbol{\omega}}_H, \tag{48}$$

$$\hat{g}^*_{QSa} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\right), \tag{49}$$

$$\hat{h}^*_{QSa} = \left[\frac{(t-n-7)(t-n-5)(t-n)}{t^2(t-2)}\right]\left(\frac{\hat{\theta}^2_H}{\hat{\theta}^2_H+\frac{n}{t}}\right). \tag{50}$$

Notice that $\hat{g}^*_{QSa} = \hat{g}^*_{QS}$ and $\hat{h}^*_{QSa}$ differs with $\hat{h}^*_{QS}$ only in the denominator of the last factor in (50). Portfolio rule $\mathcal{QS}a_I$ will be of use when analyzing the expected out-of-sample utility of the implementable version of the three-fund rule of Kan and Zhou (2007).

Table 3 summarizes the implementable portfolio rules including random coefficients $\hat{a}$ and $\hat{b}$, where, not surprisingly, all studied portfolio rules result to be linear combinations of plug-in estimators $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$. In particular, note that the implementable three-fund rule of Kan and Zhou ($\mathcal{KZ}_I$) can be expressed in terms of H because $\hat{\boldsymbol{\omega}}_{\tilde{H}}$ in (27) coincides with $\hat{\boldsymbol{\omega}}_H$ when μ_G is replaced with $\hat{\mu}_G$. Similarly, note that $\tilde{G}$ can be trivially replaced with G since $\hat{\mu}_G \times \hat{\boldsymbol{\omega}}_{\tilde{G}}$ is equal to $(\hat{\mu}_G/\hat{\sigma}^2_G)\times\hat{\boldsymbol{\omega}}_G$. However, in the case of portfolio rule $\mathcal{Y}a$ in (36), $\hat{\boldsymbol{\omega}}_{\check{H}}$ given by (15) does not coincide with $\hat{\boldsymbol{\omega}}_H$, and also the implementable version of coefficient $\check{h}^*_{\mathcal{Y}a}$ in (38) depends on both $\hat{\theta^2_G}$ and $\hat{\theta^2_H}$. For these inconveniences and the impossibility to find closed-form expressions for the expected out-of-sample utility, we will not continue working with the implementable version of $\mathcal{Y}a$. We formalize the previous observations regarding implementable portfolio rules in the following remark.

Remark 3. Implementable three-fund rules $\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{KZ}_I$, $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ in Table 3 can be expressed as

$$\hat{\boldsymbol{\omega}}_{P_I} = \left(x(n,t;P_I)\times\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\right)\times\hat{\boldsymbol{\omega}}_G + \left(y(n,t;P_I)\times f_k(\hat{\theta}^2_H;n,t)\right)\times\hat{\boldsymbol{\omega}}_H, \tag{51}$$

where $\mathrm{P}_I \in \{\mathcal{Q}_I, \mathcal{M}_I, \mathcal{KZ}_I, \mathcal{QS}_I, \mathcal{QS}a_I\}$, $x(n,t;P_I) \in (0,1)$ and $y(n,t;P_I) \in (0,1)$ depend on n and t, only; and, function f_k is such that

$$f_k(q;n,t) = \frac{q}{q+\frac{n-k}{t}}, \quad 0 \le k < n, \quad q > 0. \tag{52}$$

If $\mathrm{P}_I \in \{\mathcal{Q}, \mathcal{QS}_I, \mathcal{M}_I\}$, then $k = 1$ and otherwise $k = 0$. Furthermore, the aforementioned portfolio rules invest in $\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\times\hat{\boldsymbol{\omega}}_G$ and $f_k(\hat{\theta}^2_H;n,t)\times\hat{\boldsymbol{\omega}}_H$ which are not orthogonal.

Note that, in general, orthogonality for the three-fund rules of Remark 2 is lost when we try to estimate and replicate the total position in risky assets of the tangency portfolio by introducing the factor $\gamma^{-1}(\hat{\mu}_G/\hat{\sigma}^2_G)$ for $\hat{\boldsymbol{\omega}}_G$. If we compare $\hat{\boldsymbol{\omega}}_{P_I}$ in (51) with $\boldsymbol{\omega}_S$

Table 3
Estimated coefficients a and b for implementable rules.

P_I	A	B	$\hat{a}$	$\hat{b}$
$\mathcal{KZ}_I$	G	H	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)$	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2+\frac{n}{t}}\right)$
$\mathcal{Q}_I$	G	H	$\left[\frac{t-n-1}{(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)$	$\left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2+\frac{n-1}{t}}\right)$
$\mathcal{M}_I$	G	H	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)$	$\left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2+\frac{n-1}{t}}\right)$
$\mathcal{QS}_I$	G	H	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)$	$\left[\frac{(t-n)(t-n-5)(t-n-7)}{t^2(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2+\frac{n-1}{t}}\right)$
$\mathcal{QS}a_I$	G	H	$\left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)$	$\left[\frac{(t-n)(t-n-5)(t-n-7)}{t^2(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2+\frac{n}{t}}\right)$

in (9), it is immediate that they have the same structure; but, the presence of estimation error leads to less aggressive holdings in both G and H. Moreover, inspecting $\hat{a}$ and $\hat{b}$ in Table 3, it is clear that $\mathcal{Q}_I$ is the most aggressive and $\mathcal{QS}_I$ is the least aggressive of all the implementable three-fund rules considered. Recall that $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ are the only portfolio rules with adjusted shrinkage (correction) factors due to random coefficients; and, they could have been derived by the same procedure of Appendix D but starting from $\mathcal{M}_I$ and $\mathcal{KZ}_I$, respectively.

Similar to $\mathcal{V}^P$ in (40), we denote the expected out-of-sample utility of an implementable rule P_I in Table 3 as

$$\mathcal{V}^{P_I} := \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{P_I})] = \mathbb{E}[\mathcal{U}(\hat{g}_{P_I} \times \hat{\boldsymbol{\omega}}_G + \hat{h}_{P_I} \times \hat{\boldsymbol{\omega}}_H)], \quad \mathrm{P}_I \in \{\mathcal{KZ}_I, \mathcal{Q}_I, \mathcal{M}_I, \mathcal{QS}_I, \mathcal{QS}a_I\},$$

which can be further decomposed as $\mathcal{V}^{P_I} = \mathcal{V}_G^{P_I} + \mathcal{V}_H^{P_I} + \mathcal{V}_{G\wedge H}^{P_I}$, where $\mathcal{V}_G^{P_I} = \mathbb{E}[\mathcal{U}(\hat{g}_{P_I} \times \hat{\boldsymbol{\omega}}_G)]$, $\mathcal{V}_H^{P_I} = \mathbb{E}[\mathcal{U}(\hat{h}_{P_I} \times \hat{\boldsymbol{\omega}}_H)]$ and $\mathcal{V}_{G\wedge H}^{P_I} = -\gamma\mathbb{E}[\hat{g}_{P_I}\hat{h}_{P_I} \times \hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_H]$, i.e., the term accounting for the interaction between funds G and H. Furthermore, using the results in Kan et al. (2016), one can express these components as

$$\mathcal{V}_G^{P_I} := \kappa_{P_I,1} \times \theta_G^2 + \kappa_{P_I,2} \times \theta_H^2 + \kappa_{P_I,3}, \tag{53}$$

$$\mathcal{V}_H^{P_I} := \kappa_{P_I,4} \times \theta_H^2 \mathbb{E}[f_k(Y_1)] + \kappa_{P_I,5} \times \mathbb{E}[f_k(Y_2)^2 Y_2], \tag{54}$$

$$\mathcal{V}_{G\wedge H}^{P_I} := \kappa_{P_I,6} \times \mathbb{E}[f_k(Y_2)Y_2], \tag{55}$$

where $\kappa_{P_I,j} = \kappa_{P_I,j}(t; n, \gamma)$, $t > n+7$, $j = 1, \ldots, 6$, denote functions of t and n described in Appendix E, and f_k was defined in (52) and evaluated in $Y_1 \sim G_{n+1,t-n-1}^{t\theta_H^2}$ and $Y_2 \sim G_{n-1,t-n-1}^{t\theta_H^2}$, where $G_{p,q}^{\delta} = \frac{p}{q} \times F_{p,q}^{\delta}$, and $F_{p,q}^{\delta}$ denotes a non-central F distribution with p and q degrees of freedom and with non-centrality parameter equal to δ.

This representation is useful because it facilitates the comparison across the expected out-of-sample utilities of the portfolio rules. Moreover, note that $\kappa_{P_I,1}$, $\kappa_{P_I,2}$ and $\kappa_{P_I,3}$ are polynomials associated with the performance of G, $\kappa_{P_I,4}$ and $\kappa_{P_I,5}$ are associated with the performance of H, and $\kappa_{P_I,6}$ is related to the interaction term. In the case of the performance of H, we can make a further identification, since $\kappa_{P_I,4}$ appears only in the mean term and $\kappa_{P_I,5}$ appears only in the variance term. Regarding portfolio G we are not able to make this additional identification. From Fig. 1, for example, it is clear that the $\mathcal{QS}_I$ is the worst rule for $\kappa_{P_I,4}$, i.e., regarding the mean of H, but the better one for $\kappa_{P_I,5}$, i.e. regarding the variance of H, and the interaction term $\kappa_{P_I,6}$. Notice also that $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ share the same polynomials.

Also, from Fig. 1 alone we are unable to assess if rule $\mathcal{QS}_I$ is superior to all other rules since the answer depends on the magnitudes of $\theta_H^2\mathbb{E}[f_k(Y_1)]$, $\mathbb{E}[f_k(Y_2)^2 Y_2]$ and $\mathbb{E}[f_k(Y_2)Y_2]$. Moreover, it is possible to establish conditions under which, for example, $\mathcal{QS}_I$ is superior to $\mathcal{Q}_I$ in terms of expected out-of-sample utility. These conditions are formalized in the following proposition.

Proposition 4. *Consider implementable portfolio rules $\hat{\boldsymbol{\omega}}_{\mathcal{Q}_I}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{QS}_I}$ defined in* (42) and (45)*, respectively. Assuming the decomposition of expected out-of-sample utility given by* (53)*,* (54) and (55)*, we have the following statements:*

(1) If $t > n+4$, then $\mathcal{V}_G^{\mathcal{QS}_I} > \mathcal{V}_G^{\mathcal{Q}_I}$.
(2) If $t > n+7$, then $\mathcal{V}_{G\wedge H}^{\mathcal{QS}_I} > \mathcal{V}_{G\wedge H}^{\mathcal{Q}_I}$.
(3) Let $t > n+7$ and

$$\mathrm{UB}_{H,1} = \frac{1}{1-\beta} \times \left[(a_1 - b_1 - b_1c_1 + a_1\beta) + \sqrt{(a_1 - b_1 - b_1c_1 + a_1\beta)^2 + a_1^2(1-\beta^2)}\right], \tag{56}$$

where

$$\beta = \frac{(t-n-5)(t-n-7)}{t(t-n-3)}, \quad a_1 = \frac{n-1}{t}, \quad b_1 = \frac{n+1}{t}, \quad \textit{and} \quad c_1 = \frac{t-n-3}{t}. \tag{57}$$

If $\theta_H^2 \leq \mathrm{UB}_{H,1}$, then $\mathcal{V}_H^{\mathcal{QS}_I} \geq \mathcal{V}_H^{\mathcal{Q}_I}$ and $\mathcal{V}^{\mathcal{QS}_I} > \mathcal{V}^{\mathcal{Q}_I}$.

Proof. See Appendix F. □

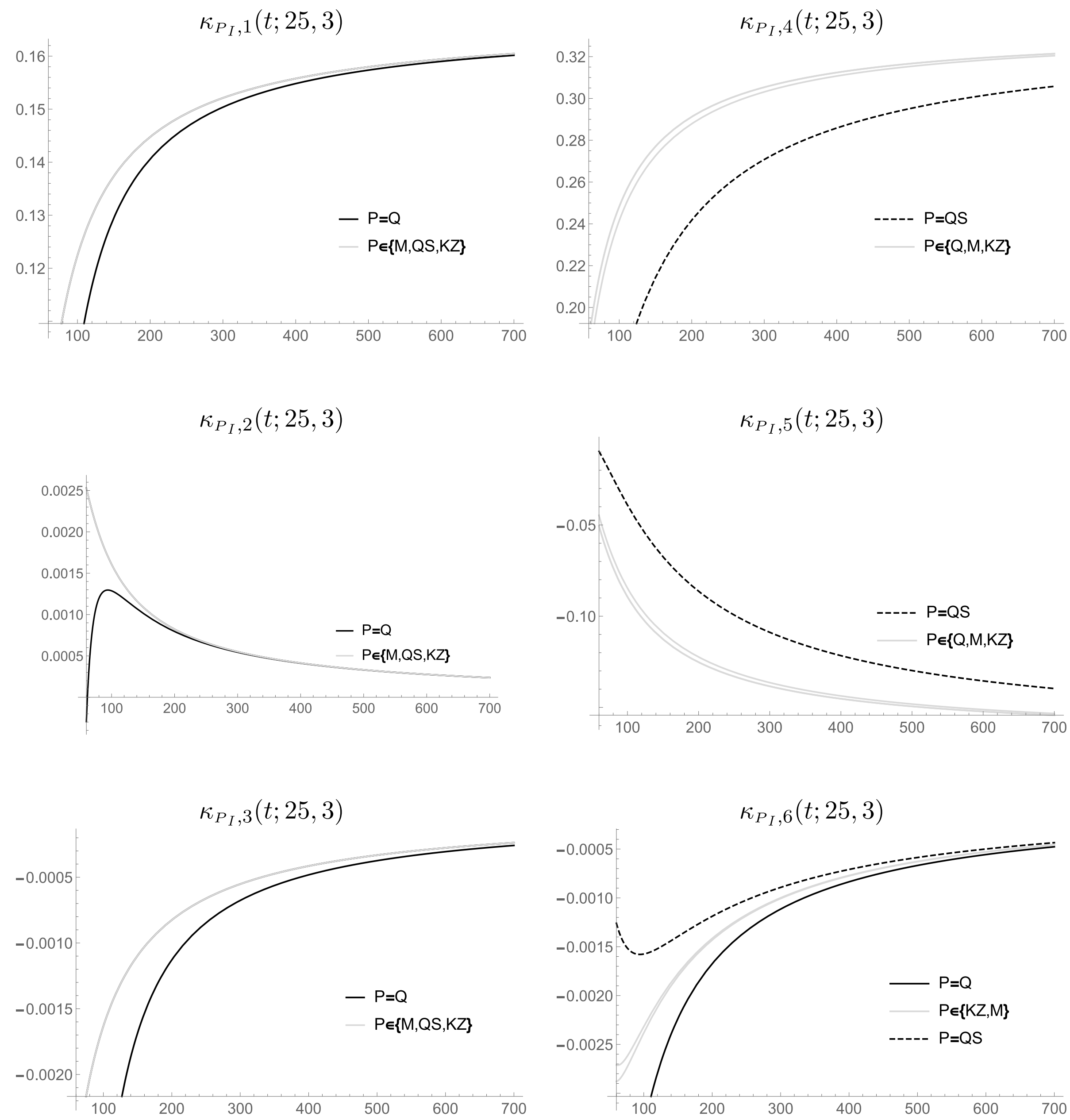

Fig. 1. Polynomials in (53)–(55) with $n = 25$, $\gamma = 3$ and $t \in [60, 700]$.

Statements *(1)* and *(2)* of Proposition 4—the former related to portfolio G and the latter to interaction between G and H—show that corrections due to random coefficients considered in $\mathcal{QS}_I$ produce improvements in expected out-of-sample utility with respect to $\mathcal{Q}_I$. Additionally, in statement *(3)* we found upper bound $\text{UB}_{H,1}$, depending on n and t only, such that if $\theta_H^2 \leq \text{UB}_{H,1}$, then $\mathcal{QS}_I$ is superior to $\mathcal{Q}_I$ in terms of the expected out-of-sample utility associated to H. With respect to the behavior of the bound $\text{UB}_{H,1}$, we can prove that is decreasing in t for a fixed n, and increasing in n for a fixed t. The numerical experiments of Appendix A illustrate that for almost all the data sets considered, θ_H^2 satisfies the constraint of statement *(3)*.

The comparison of the expected out-of-sample utilities of $\mathcal{QS}_I$ and $\mathcal{KZ}_I$ is more complicated because function f_k in (52) is different in both portfolio rules. Nonetheless, it is straightforward to verify that $\mathcal{V}_G^{\mathcal{QS}_I} = \mathcal{V}_G^{P_I}$ where $P_I = \left\{\mathcal{KZ}_I, \mathcal{M}_I, \mathcal{QS}a_I\right\}$ because from Appendix D that $x^* = \frac{(t-n-1)(t-n-4)}{t(t-2)}$ is such that $x^* = \text{argmax}_x \left\{\mathbb{E}\left[\mathcal{U}\left(x \times \gamma^{-1}(\hat{\mu}_G/\hat{\sigma}_G^2)\hat{\boldsymbol{\omega}}_G\right)\right]\right\}$. In the next proposition, we analyze the expected out-of-sample utilities of $\mathcal{KZ}_I$ and $\mathcal{QS}a_I$ because they share function $f_0(q)$.

Proposition 5. *Consider implementable portfolio rules $\hat{\boldsymbol{\omega}}_{\mathcal{QS}a_I}$ and $\hat{\boldsymbol{\omega}}_{\mathcal{KZ}_I}$ defined in (48) and Table 3, respectively. Assuming $t > n+7$ and the decomposition of expected out-of-sample utility given by (53), (54)* and *(55), we have the following statements:*

(1) $\mathcal{V}_{G\wedge H}^{\mathcal{QS}a_I} > \mathcal{V}_{G\wedge H}^{\mathcal{KZ}_I}$.
(2) Let

$$\mathrm{UB}_{H,2} = \frac{1}{2-\lambda} \times \left[(-b_2 - b_2 c_2 + a_2\lambda) + \sqrt{(-b_2 - b_2 c_2 + a_2\lambda)^2 + a_2^2\lambda(2-\lambda)}\right], \tag{58}$$

where $a_2 = a_1$, $b_2 = b_1$, $c_2 = \frac{n}{n-1}c_1$ with a_1, b_1 and c_1 given by (57), and

$$\lambda = \frac{(t-n-1)(t-n-4)}{(t-n)(t-n-3)} + \frac{(t-n-5)(t-n-7)}{t(t-n-3)}. \tag{59}$$

If $\theta_H^2 \le \mathrm{UB}_{H,2}$, then $\mathcal{V}_H^{\mathcal{QS}a_I} \ge \mathcal{V}_H^{\mathcal{KZ}_I}$ and $\mathcal{V}^{\mathcal{QS}a_I} > \mathcal{V}^{\mathcal{KZ}_I}$.

Proof. See Appendix G. □

Thus, Proposition 5 provides a condition under which implementable rule $\mathcal{QS}a_I$ has a greater expected out-of-sample utility than $\mathcal{KZ}_I$. Also, upper bound (58) is increasing in n for a fixed t; but, for n fixed is increasing until some t^* and thus is decreasing in t. Therefore, $\mathcal{QS}a_I$ will tend to improve its performance with respect to $\mathcal{KZ}_I$ for smaller values of θ_H^2 and a bigger asset universe. The fact that $\mathrm{UB}_{H,2} < \mathrm{UB}_{H,1}$ for $t > n+7$ together with $\mathcal{V}_G^{\mathcal{KZ}_I} > \mathcal{V}_G^{\mathcal{Q}_I}$ imply that $\mathcal{KZ}_I$ will be inclined to have better out-of-sample performance than $\mathcal{Q}_I$. In the numerical experiments (see Appendix A for more details) we find that $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ perform better than $\mathcal{KZ}_I$ and $\mathcal{Q}_I$ for almost all scenarios with the exception of those with large values of t. Also, in general we observe that $\mathcal{V}^{\mathcal{KZ}_I} > \mathcal{V}^{\mathcal{Q}_I}$. Moreover, $\mathcal{V}^{\mathcal{QS}_I}$ and $\mathcal{V}^{\mathcal{QS}a_I}$ have very similar expected out-of-sample utilities but the latter tend to have higher performance in small sample sizes because $\mathcal{QS}a_I$ has slightly less aggressive allocations in $\hat{\boldsymbol{\omega}}_H$ than $\mathcal{QS}_I$.

When we compare implementable portfolio rules $\mathcal{M}_I$ and $\mathcal{Q}_I$, it is straightforward to notice that $\mathcal{V}^{\mathcal{M}_I} > \mathcal{V}^{\mathcal{Q}_I}$ holds. However, the relationship between the expected out-of-sample performances of $\mathcal{KZ}_I$ and $\mathcal{M}_I$ is not clear from the information on Table 3. To analyze this relationship, we define an additional portfolio rule $\mathcal{M}a_I$ given by:

$$\hat{\boldsymbol{\omega}}_{\mathcal{M}a_I} := \hat{\boldsymbol{\omega}}_{(G,H)}(\hat{g}_{\mathcal{M}a}, \hat{h}^*_{\mathcal{M}a}) = \hat{g}_{\mathcal{M}a} \times \hat{\boldsymbol{\omega}}_G + \hat{h}_{\mathcal{M}a} \times \hat{\boldsymbol{\omega}}_H, \tag{60}$$

$$\hat{g}_{\mathcal{M}a} = \left[\frac{(t-n-1)(t-n-4)}{t(t-2)}\right]\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right), \tag{61}$$

$$\hat{h}_{\mathcal{M}a} = \left[\frac{(t-n)(t-n-3)}{t(t-2)}\right]\left(\frac{\hat{\theta}_H^2}{\hat{\theta}_H^2 + \frac{n}{t}}\right). \tag{62}$$

The only difference between implementable portfolio rules $\mathcal{M}a_I$ and $\mathcal{M}_I$ is that the former contains the term $\frac{n}{t}$ in the last factor of $\hat{h}_{\mathcal{M}a}$ in (62) while the latter has the term $\frac{n-1}{t}$ in the last factor of $\hat{h}_{\mathcal{M}}$ (see Table 3 for more details). Consequently, $\mathcal{M}a_I$ is (slightly) more conservative than $\mathcal{M}_I$ with respect to the weight given to the plug-in estimator of H. If $t > n+4$, then we can obtain results similar to those of Propositions 4 and 5 but for $\mathcal{KZ}_I$ and $\mathcal{M}a_I$. More specifically, we can show that: *i)* $\mathcal{V}_{G\wedge H}^{\mathcal{KZ}_I} > \mathcal{V}_{G\wedge H}^{\mathcal{M}a_I}$ holds; and, *ii)* if $\theta_H^2 \le \mathrm{UB}_{H,3}$, then $\mathcal{V}_H^{\mathcal{KZ}_I} \ge \mathcal{V}_H^{\mathcal{M}a_I}$ and $\mathcal{V}^{\mathcal{KZ}_I} > \mathcal{V}^{\mathcal{M}a_I}$ where upper bound $\mathrm{UB}_{H,3}$ is given by

$$\mathrm{UB}_{H,3} = \frac{1}{1-\vartheta} \times \left[(a_2 - b_2 - b_2c_2 + a_2\vartheta) + \sqrt{(a_2 - b_2 - b_2c_2 + a_2\vartheta)^2 + a_2^2(1-\vartheta^2)}\right], \tag{63}$$

with a_2, b_2, c_2 as in Proposition 5 and

$$\vartheta = \frac{(t-n-4)(t-n-1)}{(t-n-3)(t-n)}. \tag{64}$$

It is important to mention that condition $\theta_H^2 \le \mathrm{UB}_{H,3}$ was satisfied in almost all data sets and scenarios considered in Appendix A (with the exception of some scenarios of Data Set IV). Because the expected out-of-sample utilities of $\mathcal{M}_I$ and $\mathcal{M}a_I$ tend to be very similar, we expect to have $\mathcal{V}^{\mathcal{KZ}_I} > \mathcal{V}^{\mathcal{M}_I}$. The latter is indeed the case observed in our numerical experiments.

Summarizing these findings, we empirically obtained that $\mathcal{V}^{\mathcal{KZ}_I} > \mathcal{V}^{\mathcal{M}_I}$ and analytically showed that $\mathcal{V}^{\mathcal{M}_I} > \mathcal{V}^{\mathcal{Q}_I}$. However, for theoretical rules we have $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$, as stated by Proposition 3. This reversal in performance happens because the greater estimation risk of implementable portfolio rules is mitigated by having more conservative allocations on their corresponding non-orthogonal components, i.e., smaller factors $x(n,t;P_I)$ and $y(n,t;P_I)$ of Remark 3. Nonetheless, theoretical rules benefits with more aggressive allocations on their orthogonal/uncorrelated components. Notice that among the three aforementioned implementable (theoretical) portfolio rules, $\mathcal{KZ}_I$ ($\mathcal{KZ}$) is the most conservative and $\mathcal{Q}_I$ ($\mathcal{Q}$) is the most aggressive. Furthermore, the relative performance of the implementable rules of this section depends on the value of θ_H^2 which is unknown to investors. This shows that finding appropriate deterministic shrinkage factors for the non-orthogonal components of implementable portfolio rules cannot be based on a one-size-fits-all formula but rather in more complex relationships of unknown parameters like squared Sharpe ratios.

We close this section referring to the results of the numerical experiments in Appendix A. The main conclusions of these experiments are threefold: *i)* the differences in performance of the compared theoretical portfolio rules are very similar; *ii)* when we pass from theoretical to implementable portfolio rules, the extra estimation risk generated by either the unknown correction factors (or the exposure of the tangency portfolio to the attribute) produces a significant loss in out-of-sample performance and this is more severe when t is small; and *iii)* the orthogonality structure used to equip implementable portfolio rules $\mathcal{Q}S_I$ and $\mathcal{Q}Sa_I$ was indeed useful to reduce estimation risk especially for small sample sizes. Finally, conclusion *iii)* is robust under more sophisticated estimators of θ_S and θ_H as, for example, the ones proposed in Kan and Zhou (2007).

4. Limiting performance of portfolio rules

In this section, we study the performance of the theoretical and implementable portfolio rules presented in Section 3 under the limiting conditions $t \to \infty$, $n \to \infty$ and $\frac{n}{t} = c$ where $c \in (0,1)$, as in Bodnar et al. (2018). Also, this section includes the effects of correction factors $x(n,t;P_I)$ and $y(n,t;P_I)$, defined in Eq. (51) of Remark 3, in the expected out-of-sample utility and Sharpe ratio under the aforementioned limiting conditions.

Throughout this section we will assume that μ_G, σ_G^2 and θ_H^2 are functions of the number of assets, n, but as $n \to \infty$ they approach to finite quantities $\bar{\mu}_G$, $\bar{\sigma}_G^2 \neq 0$ and $\bar{\theta}_H^2$; and, we define $\bar{\theta}_G^2 := \bar{\mu}_G^2/\bar{\sigma}_G^2$ and $\bar{\theta}_S^2 := \bar{\theta}_G^2 + \bar{\theta}_H^2$.

4.1. Limiting expected out-of-sample utility of theoretical portfolio rules

Based on portfolio rule $\mathcal{Q}$, and assuming $x, y \in \mathbb{R}/\{0\}$ and f_k defined as in (52), we introduce a general theoretical portfolio rule $\mathcal{B}$ given by:

$$\hat{\boldsymbol{\omega}}_{\mathcal{B}}(x,y) := x \times \frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\hat{\boldsymbol{\omega}}_G + y \times f_1(\theta_H^2; n,t)\hat{\boldsymbol{\omega}}_H. \tag{65}$$

The expected out-of-sample utility of $\mathcal{B}$ is defined as $\mathcal{V}^{\mathcal{B}}(x,y) := \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{B}}(x,y))]$.[13] Because $\hat{\boldsymbol{\omega}}_G$ and $\hat{\boldsymbol{\omega}}_H$ are orthogonal and coefficients x and y in (65) are all deterministic, we have the following performance decomposition

$$\mathcal{V}^{\mathcal{B}}(x,y) = \mathcal{V}_G^{\mathcal{B}}(x) + \mathcal{V}_H^{\mathcal{B}}(y), \tag{66}$$

where

$$\mathcal{V}_G^{\mathcal{B}}(x) := \mathbb{E}\left[\mathcal{U}\left(x \times \frac{1}{\gamma}\frac{\mu_G}{\sigma_G^2}\hat{\boldsymbol{\omega}}_G\right)\right] \tag{67}$$

$$= \frac{x}{\gamma}\theta_G^2 - \frac{x^2}{2\gamma}\frac{(t-2)\theta_G^2}{t-n-1}, \tag{68}$$

$$\mathcal{V}_H^{\mathcal{B}}(y) := \mathbb{E}\left[\mathcal{U}\left(y \times f_1(\theta_H^2; n,t)\hat{\boldsymbol{\omega}}_H\right)\right] \tag{69}$$

$$= \frac{y}{\gamma}\frac{t\theta_H^2\ f_1(\theta_H^2)}{t-n-1} - \frac{y^2}{2\gamma}\frac{t(t-2)(t\theta_H^2 + (n-1))\ f_1(\theta_H^2)^2}{(t-n-3)(t-n-1)(t-n)}. \tag{70}$$

Assuming $n \to \infty$, $t \to \infty$ and $c = \frac{n}{t}$, we define

$$\bar{\mathcal{V}}^{\mathcal{B}}(x,y) := \lim_{\substack{n\to\infty \\ t\to\infty}} \mathbb{E}\left[\mathcal{U}\left(\hat{\boldsymbol{\omega}}_{\mathcal{B}}(x,y)\right)\right] = \bar{\mathcal{V}}_G^{\mathcal{B}}(x) + \bar{\mathcal{V}}_H^{\mathcal{B}}(y), \tag{71}$$

where

$$\bar{\mathcal{V}}_G^{\mathcal{B}}(x) := \lim_{\substack{n\to\infty \\ t\to\infty}} \mathcal{V}_G^{\mathcal{B}}(x) = \frac{x}{\gamma}\bar{\theta}_G^2 - \frac{x^2}{2\gamma}\frac{\bar{\theta}_G^2}{(1-c)}, \tag{72}$$

$$\bar{\mathcal{V}}_H^{\mathcal{B}}(y) := \lim_{\substack{n\to\infty \\ t\to\infty}} \mathcal{V}_H^{\mathcal{B}}(y) = \frac{y}{\gamma}\frac{\bar{\theta}_H^2}{1-c}\left(\frac{\bar{\theta}_H^2}{\bar{\theta}_H^2 + c}\right) - \frac{y^2}{2\gamma}\frac{\bar{\theta}_H^2 + c}{(1-c)^3}\left(\frac{\bar{\theta}_H^2}{\bar{\theta}_H^2 + c}\right)^2. \tag{73}$$

Coefficients $\bar{x}_{\mathcal{B}}$ and $\bar{y}_{\mathcal{B}}$ that maximize $\bar{\mathcal{V}}^{\mathcal{B}}(x,y)$ in (71) can be found individually from expressions (72) and (73), and they are given by

$$\bar{x}_{\mathcal{B}} = (1-c), \qquad \text{and} \qquad \bar{y}_{\mathcal{B}} = (1-c)^2. \tag{74}$$

The values in (74) yield the following limiting individual performances

$$\bar{\mathcal{V}}_G^{\mathcal{B}}(\bar{x}_{\mathcal{B}}) = (1-c)\frac{\bar{\theta}_G^2}{2\gamma}, \tag{75}$$

$$\bar{\mathcal{V}}_H^{\mathcal{B}}(\bar{y}_{\mathcal{B}}) = (1-c)\left(\frac{\bar{\theta}_H^2}{\bar{\theta}_H^2 + c}\right)\frac{\bar{\theta}_H^2}{2\gamma}. \tag{76}$$

[13] Portfolio rule $\mathcal{B}$ provides a general representation of rules combining the plug-in estimators of G and H. Thus, portfolio rule $\mathcal{Q}$ is a particular case of $\mathcal{B}$ when x and y are allowed to be functions of n and t only.

Hence, the optimal expected out-of-sample utility (under the imposed limiting assumptions) generated by portfolio rule $\hat{\boldsymbol{\omega}}_{\mathcal{B}}$ is

$$\bar{\mathcal{V}}^{\mathcal{B}}(\bar{x}_{\mathcal{B}},\bar{y}_{\mathcal{B}})=\bar{\mathcal{V}}^{\mathcal{B}}_{G}(\bar{x}_{\mathcal{B}})+\bar{\mathcal{V}}^{\mathcal{B}}_{H}(\bar{y}_{\mathcal{B}})=(1-c)\left[\frac{\bar{\theta}^2_G}{2\gamma}+\left(\frac{\bar{\theta}^2_H}{\bar{\theta}^2_H+c}\right)\frac{\bar{\theta}^2_H}{2\gamma}\right]. \tag{77}$$

We observe that $\bar{\mathcal{V}}^{\mathcal{B}}(\bar{x}_{\mathcal{B}},\bar{y}_{\mathcal{B}})$ decreases as c increases; but, as $c\to 0$ it converges to the theoretical performance of the tangency portfolio, $\bar{\theta}^2_S/2\gamma$. Moreover, $\bar{\mathcal{V}}^{\mathcal{B}}(\bar{x}_{\mathcal{B}},\bar{y}_{\mathcal{B}})$ in (77) can be derived directly from the expected out-of-sample utility of $\mathcal{Q}$, $\mathcal{V}^{\mathcal{Q}}=\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})]$, by taking the appropriate limits to expression (19). Then, we can also define:

$$\bar{\mathcal{V}}^{\mathcal{Q}}:=\lim_{\substack{n\to\infty\\ t\to\infty}}\mathcal{V}^{\mathcal{Q}}, \tag{78}$$

and clearly $\bar{\mathcal{V}}^{\mathcal{Q}}=\bar{\mathcal{V}}^{\mathcal{B}}(\bar{x}_{\mathcal{B}},\bar{y}_{\mathcal{B}})$ holds.[14] Additionally, if we relate portfolio rules $\mathcal{Q}$ and $\mathcal{B}$ using Eqs. (65) and (16), then both portfolio rules will coincide when we set $x=\frac{t-n-1}{t-2}$ and $y=\frac{(t-n)(t-n-3)}{t(t-2)}$. Under the limiting conditions ($n\to\infty$, $t\to\infty$ and $c=\frac{n}{t}$), it is clear that x and y will be equal to $\bar{x}_{\mathcal{Q}}=(1-c)$ and $\bar{y}_{\mathcal{Q}}=(1-c)^2$, respectively; and, they coincide with the expressions of $\bar{x}_{\mathcal{B}}$ and $\bar{y}_{\mathcal{B}}$ in (74). Therefore, the shrinkage factors of $\mathcal{Q}$ coincide with the optimal ones in the limit.

4.2. Limiting expected out-of-sample utility of implementable portfolio rules

By the notation introduced in Section 3.3, and assuming $x,y\in\mathbb{R}/\{0\}$ and f_k defined as in (52), we define a general implementable portfolio rule $\mathcal{I}$ based on the plug-in estimators of portfolios G and H:

$$\hat{\boldsymbol{\omega}}_{\mathcal{I}}(x,y):=x\times\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\hat{\boldsymbol{\omega}}_G+y\times f_k(\hat{\theta}^2_H;n,t)\hat{\boldsymbol{\omega}}_H. \tag{79}$$

The expected out-of-sample utility of $\hat{\boldsymbol{\omega}}_{\mathcal{I}}(x,y)$ can be found with the results given in Appendix D. Portfolio rule $\mathcal{I}$ encompasses portfolio rules $\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{QS}_I$, $\mathcal{QS}a_I$ and $\mathcal{KZ}_I$ in Table 3 when x and y are allowed to be functions of n and t. Consider:

$$\begin{aligned}\bar{\mathcal{V}}^{\mathcal{I}}(x,y)&:=\lim_{\substack{n\to\infty\\ t\to\infty}}\mathbb{E}\left[\mathcal{U}\left(\hat{\boldsymbol{\omega}}_{\mathcal{I}}(x,y)\right)\right] &(80)\\ &=\bar{\mathcal{V}}^{\mathcal{I}}_G(x)+\bar{\mathcal{V}}^{\mathcal{I}}_H(y)+\bar{\mathcal{V}}^{\mathcal{I}}_{G\wedge H}(x,y), &(81)\end{aligned}$$

where

$$\begin{aligned}\bar{\mathcal{V}}^{\mathcal{I}}_G(x)&:=\lim_{\substack{n\to\infty\\ t\to\infty}}\mathbb{E}\left[\mathcal{U}\left(x\times\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\hat{\boldsymbol{\omega}}_G\right)\right] &(82)\\ &=\frac{x}{\gamma}\frac{\bar{\theta}^2_G}{1-c}-\frac{x^2}{2\gamma}\frac{\bar{\theta}^2_G}{(1-c)^3}, &(83)\\ \bar{\mathcal{V}}^{\mathcal{I}}_H(y)&:=\lim_{\substack{n\to\infty\\ t\to\infty}}\mathbb{E}\left[\mathcal{U}\left(y\times f_k(\hat{\theta}^2_H;n,t)\hat{\boldsymbol{\omega}}_H\right)\right] &(84)\\ &=\frac{y}{\gamma}\frac{\bar{\theta}^2_H}{(1-c)}\frac{\bar{\theta}^2_H+c}{(\bar{\theta}^2_H+c+c(1-c))}-\frac{y^2}{2\gamma}\frac{1}{(1-c)^3}\frac{(\bar{\theta}^2_H+c)^3}{(\bar{\theta}^2_H+c+c(1-c))^2}, &(85)\\ \bar{\mathcal{V}}^{\mathcal{I}}_{G\wedge H}(x,y)&:=\lim_{\substack{n\to\infty\\ t\to\infty}}\mathbb{E}\left[-xy\left(\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\right)f_k(\hat{\theta}^2_H;n,t)\hat{\boldsymbol{\omega}}_G^{\top}\boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_H\right]=0. &(86)\end{aligned}$$

First, the limiting expected out-of-sample performances in (83) and (85) are independent of parameter k of function f_k and, most important, they are available in closed-form. Second, in the limit, portfolio rules $(\gamma^{-1}\hat{\mu}_G/\hat{\sigma}^2_G)\hat{\boldsymbol{\omega}}_G$ and $f_k(\hat{\theta}^2_H;n,t)\hat{\boldsymbol{\omega}}_H$ are orthogonal according to Definition 2, and this property was not observed when t and n were both finite.[15] The same expression permits coefficients $\bar{x}_{\mathcal{I}}$ and $\bar{y}_{\mathcal{I}}$—that maximize the limiting expected out-of-sample performance in (81)—to be found individually from (83) and (85); after some computations we have:

$$\bar{x}_{\mathcal{I}}=(1-c)^2,\qquad\text{and}\qquad\bar{y}_{\mathcal{I}}=(1-c)^2\left(\frac{\bar{\theta}^4_H+2\bar{\theta}^2_Hc-c^2\bar{\theta}^2_H}{\bar{\theta}^4_H+2\bar{\theta}^2_H+c^2}\right), \tag{87}$$

and they yield the following optimal limiting expected out-of-sample utilities:

$$\bar{\mathcal{V}}^{\mathcal{I}}_G(\bar{x}_{\mathcal{I}})=(1-c)\frac{\bar{\theta}^2_G}{2\gamma},\qquad\text{and}\qquad\bar{\mathcal{V}}^{\mathcal{I}}_H(\bar{y}_{\mathcal{I}})=(1-c)\left(\frac{\bar{\theta}^2_H}{\bar{\theta}^2_H+c}\right)\frac{\bar{\theta}^2_H}{2\gamma}. \tag{88}$$

[14] It is important to mention that all theoretical portfolio rules of Table 2 produce the same limiting expected out-of-sample utility as the one of $\mathcal{Q}$. Aiming for a more succinct exposition, we do not consider the aforementioned rules in the present analysis.

[15] It can be shown that

$$\lim_{\substack{n\to\infty\\ t\to\infty}}\mathrm{Cov}\left(x\times\frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}^2_G}\hat{\boldsymbol{\omega}}_G^{\top}\boldsymbol{\mu},\,y\times f_k(\hat{\theta}^2_H;n,t)\hat{\boldsymbol{\omega}}_H^{\top}\boldsymbol{\mu}\right)=0.$$

Then, the aforementioned result and expression (86) imply that the portfolio rules are orthogonal under the limiting conditions.

Consequently, the optimal limiting performance of the implementable portfolio rule in (79) is equal to the sum of the expressions in (88), and

$$\bar{\mathcal{V}}^{\mathcal{I}}(\bar{x}_{\mathcal{I}},\bar{y}_{\mathcal{I}}) = (1-c)\left[\frac{\bar{\theta}_G^2}{2\gamma} + \left(\frac{\bar{\theta}_H^2}{\bar{\theta}_H^2+c}\right)\frac{\bar{\theta}_H^2}{2\gamma}\right]. \tag{89}$$

Interestingly, $\bar{\mathcal{V}}^{\mathcal{I}}(\bar{x}_{\mathcal{I}},\bar{y}_{\mathcal{I}}) = \bar{\mathcal{V}}^{\mathcal{B}}(\bar{x}_{\mathcal{B}},\bar{y}_{\mathcal{B}})$ and individual optimal expected out-of-sample utilities in (88) are equal to the ones in (75) and (76). Therefore, under our limiting assumptions, implementable portfolio rule $\mathcal{I}$ in (79) could potentially achieve the same optimal performance as theoretical rule $\mathcal{B}$ in (65). However, $\bar{y}_{\mathcal{I}}$ in (87) is unknown which implies that optimal performance $\bar{\mathcal{V}}_H^{\mathcal{I}}(\bar{y}_{\mathcal{I}})$ in (88) cannot be obtained in practice. Moreover, $\bar{y}_{\mathcal{I}} < \bar{y}_{\mathcal{B}}$ and $\bar{x}_{\mathcal{I}} < \bar{x}_{\mathcal{B}}$ and therefore stronger optimal correction factors are obtained for portfolio rule $\mathcal{I}$ compared to those of $\mathcal{B}$.

Portfolio rule $\mathcal{I}$ in (79) can be clearly associated with implementable portfolio rules $\mathcal{KZ}_I$, $\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ of Section 3.3. In particular, coefficients x and y in (79) can be interpreted as the limits of correction factors $x(n,t;P_I)$ and $y(n,t;P_I)$ in Eq. (51) of Remark 3. For example, if we consider portfolio rule $\mathcal{QS}_I$ in (45) and Table 3 it is clear that

$$x(n,t;\mathcal{QS}_I) = \frac{(t-n-1)(t-n-4)}{t(t-2)}, \quad \text{and} \quad y(n,t;\mathcal{QS}_I) = \frac{(t-n)(t-n-5)(t-n-7)}{t^2(t-2)}. \tag{90}$$

Under limiting conditions $n\to\infty$ $t\to\infty$ and $c = n/t$, the deterministic correction factors in (90) are equal to $(1-c)^2$ and $(1-c)^3$, respectively. If we apply the same logic to the rest of implementable rules ($\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{QS}a_I$ and $\mathcal{KZ}_I$), and the deterministic parts of the coefficients of $\hat{\omega}_G$ and $\hat{\omega}_H$ satisfy

$$\lim_{\substack{n\to\infty\\ t\to\infty}} x(n,t;P_I) = \begin{cases} (1-c) & \text{if} \quad P_I = \mathcal{Q}_I, \\ (1-c)^2 & \text{if} \quad P_I \in \left\{\mathcal{KZ}_I,\mathcal{M}_I,\mathcal{QS}_I,\mathcal{QS}a_I\right\}, \end{cases} \tag{91}$$

and

$$\lim_{\substack{n\to\infty\\ t\to\infty}} y(n,t;P_I) = \begin{cases} (1-c)^2 & \text{if} \quad P_I \in \left\{\mathcal{KZ}_I,\mathcal{Q}_I,\mathcal{M}_I\right\}, \\ (1-c)^3 & \text{if} \quad P_I \in \left\{\mathcal{QS}_I,\mathcal{QS}a_I\right\}. \end{cases} \tag{92}$$

Expression (91) shows that, in the limit, the deterministic part of coefficient of $\hat{\omega}_G$ in $\mathcal{Q}_I$ converges to $(1-c)$; however, for $\mathcal{KZ}_I$, $\mathcal{M}_I$, $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ it converges to the optimal value $\bar{x}_{\mathcal{I}} = (1-c)^2$. This implies that the limiting expected out-of-sample performance attributed to $\hat{\omega}_G$ in $\mathcal{Q}_I$ is lower than the ones of the other implementable rules because $\bar{\mathcal{V}}_G^{\mathcal{I}}(1-c) < \bar{\mathcal{V}}_G^{\mathcal{I}}(\bar{x}_{\mathcal{I}})$. Moreover, (92) indicates that the deterministic scaling of $\hat{\omega}_H$ in implementable rules $\mathcal{KZ}_I$, $\mathcal{Q}_I$, $\mathcal{M}_I$ produces a coefficient equal to $(1-c)^2$ while $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ produce one equal to $(1-c)^3$ and neither of them are optimal. From (87), we have that $\bar{y}_{\mathcal{I}} < (1-c)^2$ holds and in the next proposition we want to analyze: i) conditions under which $\bar{y}_{\mathcal{I}} > (1-c)^3$ or $\bar{y}_{\mathcal{I}} < (1-c)^3$; and, ii) the relationship between limiting performances $\bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^2)$ and $\bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^3)$.

Proposition 6. *Let $\bar{y}_{\mathcal{I}}$ given by (87) and $c = \frac{n}{t}$. The following statements hold:*

(1) $\bar{y}_{\mathcal{I}} > (1-c)^3$ *if and only if* $\bar{\theta}_H^4 + \bar{\theta}_H^2 c + c^2 > c$.
(2) *If* $\bar{\theta}_H^2 \geq \frac{1}{3}$, *then* $\bar{y}_{\mathcal{I}} \geq (1-c)^3$.
(3) *If* $\bar{\theta}_H^2 < \frac{1}{3}$ *and* $c \in \left(\frac{1-\bar{\theta}_H^2}{2} - \frac{\sqrt{1-3\bar{\theta}_H^4-2\bar{\theta}_H^2}}{2}, \frac{1-\bar{\theta}_H^2}{2} + \frac{\sqrt{1-3\bar{\theta}_H^4-2\bar{\theta}_H^2}}{2}\right)$, *then* $\bar{y}_{\mathcal{I}} < (1-c)^3$ *and* $\bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^3) > \bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^2)$.
(4) $\bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^3) > \bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^2)$ *if and only if* $\bar{\theta}_H^2 < \sqrt{c(2-c)}$.
(5) *If* $\bar{\theta}_H^2 > 1$, *then* $\bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^3) < \bar{\mathcal{V}}_H^{\mathcal{I}}((1-c)^2)$.

Proof. Statements *(1)*, *(2)* and the first part of *(3)* can be derived directly from the condition $\bar{y}_{\mathcal{I}} < (1-c)^3$. The second part of *(3)* follows from the concavity of $\bar{\mathcal{V}}_H^{\mathcal{I}}(y)$. Finally, statements *(4)* and *(5)* follow from comparing the values of $\bar{\mathcal{V}}_H^{\mathcal{I}}(y)$ when $y = (1-c)^2$ and $y = (1-c)^3$. □

From the results of Proposition 6, $y = (1-c)^2$ benefits typically from higher values of $\bar{\theta}_H^2$ and smaller values of c. On the other side, deterministic scaling $(1-c)^3$ performs better than $(1-c)^2$ for lower limiting Sharpe ratios of portfolio H and higher values of c. Thus, the extra correction factor given to $\hat{\omega}_H$ in $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ is beneficial under certain conditions as it can provide a higher expected out-of-sample utility than the ones corresponding to $\mathcal{KZ}_I$, $\mathcal{Q}_I$ and $\mathcal{M}_I$. Additionally, Propositions 4 and 5 produce a relationship similar to the one of statement *(4)* of Proposition 6. It is easy to verify than under the limiting conditions of this section, upper bounds $\mathrm{UB}_{H,1}$ in (56) and $\mathrm{UB}_{H,2}$ in (58) converge to $\sqrt{c(2-c)}$.

Moreover, the results show that determining appropriate deterministic correction factors in implementable portfolio rules is a complicated issue because, even under strong limiting conditions, they depend on unknown parameters and complicated relationships between variables. Then, we can take advantage of orthogonality properties when devising new shrinkage factors, for example the one of $\hat{\omega}_H$ in $\mathcal{QS}_I$ and $\mathcal{QS}a_I$.

Table 4
Summary of data sets utilized for numerical experiments.

Label	Data description	Period	n	t	θ^2_G	θ^2_H	θ^2_S
I	25 Portfolios formed on size and Book-to-market	1926/07–2018/11	25	1109	0.0294	0.0654	0.0948
II	32 Portfolios formed on Size, Book-to-Market, and Operating Profitability	1963/07–2018/11	32	665	0.0342	0.1335	0.1677
III	10 Industry Portfolios	1926/07–2018/11	10	1109	0.0289	0.0053	0.0341
IV	10 Portfolios Formed on Momentum	1927/01–2018/11	10	1103	0.0411	0.0297	0.0708
V	25 Portfolios Formed on Book-to-Market and Investment	1963/07–2018/1	25	341	0.0380	0.0906	0.1285

5. Concluding remarks

This document explores the use of orthogonality to improve and understand estimation risk in portfolio optimization. We constructed theoretical portfolio rules based on plug-in estimators of "optimal" orthogonal portfolios derived from a linearly constrained mean–variance optimization problem. These proposed rules outperform the theoretical portfolio rule of Kan and Zhou (2007) giving a clear identification of the sources of estimation risk. In summary, we show that Kan and Zhou (2007) rule is not strictly diversifying estimation risk when adding an extra portfolio; instead, it tries to replicate the correct orthogonal components of the tangency portfolio when the attribute is the position on risky assets.

In terms of implementable portfolio rules, using orthogonality properties we were able to develop additional correction schemes such that, under certain conditions on the Sharpe ratio of the hedge portfolio, the expected out-of-sample performance of our new portfolio rules is greater than the one of the implementable version of Kan and Zhou (2007). Interestingly, these conditions depend only on the number of assets and the number of observations and are straightforward to verify.

We studied the out-of-sample performance of theoretical and implementable portfolio rules under the limiting conditions of Bodnar et al. (2018) in which both the number of assets and the sample size tend to infinity. Then, we determine that constructing appropriate deterministic shrinkage factors is a challenging problem because, even under the limiting conditions, the optimal degree of correction depends on unknown variables and complicated relationships between them. Moreover, orthogonality properties of portfolio rules were important to assess the performance of shrinking factors under the aforementioned limiting conditions.

Finally, through orthogonality we were able to decompose the sources of estimation risk which is specially helpful when analyzing the suitability of candidate rules incorporating constraints that are not satisfied by portfolio S. This opens an interesting direction for future research, to extend our results from the one constraint case $\boldsymbol{A}\boldsymbol{\varpi} = \boldsymbol{\varrho}$, with $\boldsymbol{A} = \mathbf{1}^\top$ and $\boldsymbol{\varrho} = \varrho$ as in (2), to the multiple constraint set with rectangular $\boldsymbol{A}$ and arbitrary vector of target exposures $\boldsymbol{\varrho}$, in the quest for estimation risk assessment.

CRediT authorship contribution statement

Luis Chavez-Bedoya: Conceptualization, Methodology, Writing – original draft. **Francisco Rosales:** Conceptualization, Methodology, Software, Writing – reviewing and editing.

Appendix A. Numerical experiments

To compare the expected out-of-sample performance of the theoretical and implementable portfolio rules presented in Tables 2 and 3 respectively, five data sets are evaluated independently.[16] For all our return data points we use the monthly excess returns obtained by subtracting the T-bill rate from the reported average value-weighted return. In this exercise we assume that $\gamma = 3$, and that $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, are known (and equal to their sample estimates), so they can be employed to compute the required population parameters θ^2_G, θ^2_H and θ^2_S. The data details are summarized in Table 4. where n denotes the number of assets, and t indicates the number of observations for each data set.

Here we investigate the results concerning the known coefficients case (theoretical rules), and the unknown coefficients (implementable rules). The results, measured in terms of out-of-sample utility, for each data set are presented in Tables 5–9. Each table differentiates between two types of implementable rules, one with (and another without) the Kan & Zhou correction for the Sharpe ratio estimation.[17]

[16] The data sets are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

[17] See Kan and Zhou (2007) where the bias correction for the Sharpe ratio estimation is discussed in detail.

Table 5
Data set I: book-to-market, $n = 25$. For clarity, all quantities have been multiplied by 100.

Data set I	Theoretical rules				Implementable rules w.o. KZ correction					Implementable rules w. KZ correction				
Components	$\mathcal{Q}$	$\mathcal{M}$	$\mathcal{KZ}$	$\mathcal{Y}$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QSa}_I$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QSa}_I$
$t = 100$														
Fund 1	0.371	0.360	0.1660	0.371	−0.015	0.208	0.208	0.208	0.208	−0.015	0.208	0.208	0.208	0.208
Fund 2	0.174	0.174	0.360	0.165	−0.573	−0.573	−0.488	−0.018	−0.009	0.024	0.024	0.044	0.126	0.127
Interaction	0.000	0.000	0.000	0.000	−0.092	−0.065	−0.063	−0.043	−0.043	−0.048	−0.034	−0.032	−0.022	−0.022
Total	0.544	0.534	0.526	0.536	−0.680	−0.431	−0.343	0.146	0.156	−0.039	0.198	0.220	0.311	0.313
$t = 300$														
Fund 1	0.451	0.448	0.437	0.451	0.379	0.396	0.396	0.396	0.396	0.379	0.396	0.396	0.396	0.396
Fund 2	0.449	0.449	0.448	0.437	0.319	0.319	0.332	0.384	0.389	0.389	0.389	0.391	0.397	0.396
Interaction	0.000	0.000	0.000	0.000	−0.012	−0.011	−0.011	−0.010	−0.010	−0.008	−0.008	−0.007	−0.007	−0.007
Total	0.900	0.897	0.885	0.888	0.686	0.704	0.718	0.771	0.776	0.759	0.778	0.780	0.787	0.786
$t = 500$														
Fund 1	0.467	0.465	0.585	0.467	0.428	0.434	0.434	0.434	0.434	0.428	0.434	0.434	0.434	0.434
Fund 2	0.597	0.597	0.465	0.584	0.546	0.546	0.552	0.568	0.570	0.558	0.558	0.558	0.559	0.558
Interaction	0.000	0.000	0.000	0.000	−0.006	−0.005	−0.005	−0.005	−0.005	−0.005	−0.004	−0.004	−0.004	−0.004
Total	1.064	1.062	1.050	1.052	0.969	0.975	0.980	0.997	0.999	0.981	0.987	0.987	0.989	0.988
$t = 700$														
Fund 1	0.474	0.473	0.678	0.474	0.447	0.450	0.450	0.450	0.450	0.447	0.450	0.450	0.450	0.450
Fund 2	0.689	0.689	0.473	0.678	0.662	0.662	0.665	0.672	0.674	0.663	0.663	0.663	0.663	0.662
Interaction	0.000	0.000	0.000	0.000	−0.004	−0.004	−0.004	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003
Total	1.163	1.162	1.151	1.152	1.106	1.109	1.112	1.119	1.120	1.107	1.110	1.110	1.111	1.110

Table 6
Data set II: size, book-to-market and operating profitability, $n = 32$. For clarity, all quantities have been multiplied by 100.

Data set II	Theoretical rules				Implementable rules w.o. KZ correction					Implementable rules w. KZ correction				
Components	$\mathcal{Q}$	$\mathcal{M}$	$\mathcal{KZ}$	$\mathcal{Y}$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QSa}_I$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QSa}_I$
$t = 100$														
Fund 1	0.390	0.378	0.434	0.390	−0.127	0.236	0.236	0.236	0.236	−0.127	0.236	0.236	0.236	0.236
Fund 2	0.451	0.451	0.378	0.433	−0.363	−0.363	−0.259	0.368	0.372	0.275	0.275	0.295	0.353	0.351
Interaction	0.000	0.000	0.000	0.000	−0.159	−0.101	−0.097	−0.060	−0.059	−0.086	−0.055	−0.052	−0.032	−0.032
Total	0.840	0.828	0.812	0.823	−0.648	−0.229	−0.121	0.544	0.549	0.062	0.456	0.479	0.557	0.555
$t = 300$														
Fund 1	0.511	0.507	1.099	0.511	0.433	0.459	0.459	0.459	0.459	0.433	0.459	0.459	0.459	0.459
Fund 2	1.119	1.119	0.507	1.098	1.002	1.002	1.015	1.084	1.086	1.046	1.046	1.047	1.042	1.038
Interaction	0.000	0.000	0.000	0.000	−0.022	−0.019	−0.019	−0.016	−0.016	−0.017	−0.015	−0.015	−0.013	−0.013
Total	1.630	1.626	1.606	1.609	1.414	1.442	1.455	1.527	1.529	1.463	1.490	1.491	1.488	1.484
$t = 500$														
Fund 1	0.534	0.532	1.404	0.534	0.495	0.503	0.503	0.503	0.503	0.495	0.503	0.503	0.503	0.503
Fund 2	1.421	1.421	0.532	1.403	1.379	1.379	1.383	1.404	1.404	1.384	1.384	1.384	1.379	1.377
Interaction	0.000	0.000	0.000	0.000	−0.011	−0.010	−0.010	−0.009	−0.009	−0.010	−0.009	−0.009	−0.008	−0.008
Total	1.956	1.953	1.936	1.937	1.863	1.872	1.877	1.898	1.898	1.870	1.879	1.879	1.874	1.872
$t = 700$														
Fund 1	0.545	0.543	1.579	0.545	0.518	0.522	0.522	0.522	0.522	0.518	0.522	0.522	0.522	0.522
Fund 2	1.594	1.594	0.543	1.578	1.572	1.572	1.574	1.582	1.582	1.572	1.572	1.571	1.568	1.566
Interaction	0.000	0.000	0.000	0.000	−0.007	−0.007	−0.007	−0.006	−0.006	−0.007	−0.006	−0.006	−0.006	−0.006
Total	2.138	2.137	2.122	2.122	2.083	2.087	2.090	2.098	2.098	2.083	2.088	2.087	2.084	2.083

To clarify the presentation of our results, recall that for theoretical portfolio rules $\mathcal{Q}$, $\mathcal{M}$, $\mathcal{KZ}a$ and $\mathcal{Y}a$ are associated to Table 2 and that they take the form of $\hat{\omega}_P = a_P \times \hat{\omega}_A + b_P \times \hat{\omega}_B$, where for rules $\mathcal{Q}$ and $\mathcal{M}$ we consider A = G and B = H; while for rules $\mathcal{KZ}a$ and $\mathcal{Y}a$ we used A = G and B = S. This particular set of rules were selected since all of them exhibit a null expected covariance of excess returns. It is also worthwhile noting that rules $\mathcal{KZ}a$ and $\mathcal{Y}a$ have the exact same performances as $\mathcal{KZ}$ and $\mathcal{Y}$, respectively. Similarly, portfolio rules $\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{KZ}_I$, $\mathcal{QS}_I$ and $\mathcal{QSa}_I$ are associated to Table 3 and they take the form of $\hat{\omega}_{P_I} = \hat{g}_{P_I} \times \hat{\omega}_G + \hat{h}_{P_I} \times \hat{\omega}_H$, where it is made explicit that all implementable rules are linear combinations of portfolios G and H. In addition, the rows in Tables 5–9 group the information via sample size and composition of the overall performance in terms of "Fund 1", "Fund 2", "Interaction" and "Total". The meaning of these labels varies mildly depending on the type of rule (theoretical or implementable). In particular, for the theoretical rules, we consider: for "Fund 1" the quantity $\mathcal{V}_A^P := \mathbb{E}[\mathcal{U}(a_P \times \hat{\omega}_A)]$; for "Fund 2" the quantity $\mathcal{V}_B^P := \mathbb{E}[\mathcal{U}(b_P \times \hat{\omega}_B)]$; for "interaction" the quantity $\mathcal{V}_{A \wedge B}^P := \mathcal{V}^P - \mathcal{V}_A^P - \mathcal{V}_B^P$; and for "Total" the quantity $\mathcal{V}^P = \mathbb{E}[\mathcal{U}(\hat{\omega}_P)]$. Similarly, for the implementable rules, we consider: for "Fund 1" the expression $\mathcal{V}_G^{P_I} := \mathbb{E}[\mathcal{U}(\hat{a}_{P_I} \times \hat{\omega}_G)]$; for "Fund 2" the expression $\mathcal{V}_H^{P_I} := \mathbb{E}[\mathcal{U}(\hat{b}_{P_I} \times \hat{\omega}_H)]$; for "Interaction" the expression $\mathcal{V}_{G \wedge H}^{P_I} := \mathcal{V}^{P_I} - \mathcal{V}_G^{P_I} - \mathcal{V}_H^{P_I}$; and for "Total" the expression $\mathcal{V}^{P_I} = \mathbb{E}[\mathcal{U}(\hat{\omega}_{P_I})]$. Note that all these expectations (theoretical and implementable) are computed in closed-form using the results of Section 3.2.

Table 7
Data set III: momentum, $n = 10$. For clarity, all quantities have been multiplied by 100.

Data set III	Theoretical rules				Implementable rules w.o. KZ correction					Implementable rules w. KZ correction				
Components	$\mathcal{Q}$	$\mathcal{M}$	$\mathcal{KZ}$	$\mathcal{Y}$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$
$t = 100$														
Fund 1	0.437	0.427	0.004	0.437	0.196	0.264	0.264	0.264	0.264	0.196	0.264	0.264	0.264	0.264
Fund 2	0.004	0.004	0.427	0.004	−0.399	−0.399	−0.345	−0.248	−0.226	−0.145	−0.145	−0.12	−0.087	−0.076
Interaction	0.000	0.000	0.000	0.000	−0.021	−0.018	−0.017	−0.015	−0.014	−0.013	−0.011	−0.010	−0.009	−0.008
Total	0.441	0.431	0.431	0.441	−0.224	−0.153	−0.098	0.001	0.024	0.038	0.109	0.135	0.168	0.180
$t = 300$														
Fund 1	0.466	0.463	0.012	0.466	0.402	0.408	0.408	0.408	0.408	0.402	0.408	0.408	0.408	0.408
Fund 2	0.013	0.013	0.463	0.012	−0.098	−0.098	−0.084	−0.080	−0.070	−0.027	−0.027	−0.021	−0.020	−0.016
Interaction	0.000	0.000	0.000	0.000	−0.002	−0.002	−0.002	−0.002	−0.002	−0.001	−0.001	−0.001	−0.001	−0.001
Total	0.479	0.476	0.475	0.478	0.301	0.308	0.322	0.326	0.336	0.373	0.379	0.386	0.386	0.391
$t = 500$														
Fund 1	0.472	0.470	0.018	0.472	0.435	0.437	0.437	0.437	0.437	0.435	0.437	0.437	0.437	0.437
Fund 2	0.020	0.020	0.470	0.018	−0.037	−0.037	−0.030	−0.030	−0.025	−0.002	−0.002	0.002	0.001	0.003
Interaction	0.000	0.000	0.000	0.000	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001
Total	0.492	0.490	0.488	0.490	0.397	0.400	0.407	0.406	0.412	0.433	0.435	0.438	0.437	0.440
$t = 700$														
Fund 1	0.475	0.473	0.023	0.475	0.448	0.450	0.450	0.450	0.450	0.448	0.450	0.450	0.450	0.450
Fund 2	0.025	0.025	0.473	0.023	−0.010	−0.010	−0.005	−0.006	−0.002	0.011	0.011	0.013	0.012	0.014
Interaction	0.000	0.000	0.000	0.000	−0.001	−0.001	−0.001	−0.001	0.000	0.000	0.000	0.000	0.000	0.000
Total	0.500	0.499	0.497	0.498	0.438	0.439	0.444	0.443	0.447	0.459	0.460	0.462	0.461	0.463

Table 8
Data set IV: industry, $n = 10$. For clarity, all quantities have been multiplied by 100.

Data set IV	Theoretical rules				Implementable rules w.o. KZ correction					Implementable rules w. KZ correction				
Components	$\mathcal{Q}$	$\mathcal{M}$	$\mathcal{KZ}$	$\mathcal{Y}$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$
$t = 100$														
Fund 1	0.621	0.607	0.101	0.621	0.375	0.449	0.449	0.449	0.449	0.375	0.449	0.449	0.449	0.449
Fund 2	0.110	0.110	0.607	0.100	−0.177	−0.177	−0.128	−0.033	−0.016	−0.003	−0.003	0.019	0.052	0.059
Interaction	0.000	0.000	0.000	0.000	−0.029	−0.025	−0.023	−0.020	−0.019	−0.019	−0.017	−0.015	−0.013	−0.013
Total	0.732	0.718	0.708	0.721	0.170	0.247	0.297	0.396	0.413	0.353	0.430	0.452	0.488	0.495
$t = 300$														
Fund 1	0.664	0.659	0.225	0.664	0.598	0.605	0.605	0.605	0.605	0.598	0.605	0.605	0.605	0.605
Fund 2	0.238	0.238	0.659	0.224	0.184	0.184	0.193	0.200	0.204	0.195	0.195	0.197	0.201	0.201
Interaction	0.000	0.000	0.000	0.000	−0.005	−0.005	−0.004	−0.004	−0.004	−0.004	−0.004	−0.003	−0.003	−0.003
Total	0.902	0.897	0.884	0.887	0.778	0.785	0.794	0.801	0.806	0.790	0.797	0.799	0.803	0.803
$t = 500$														
Fund 1	0.672	0.669	0.289	0.672	0.634	0.637	0.637	0.637	0.637	0.634	0.637	0.637	0.637	0.637
Fund 2	0.302	0.302	0.669	0.288	0.277	0.277	0.28	0.282	0.284	0.272	0.272	0.272	0.274	0.272
Interaction	0.000	0.000	0.000	0.000	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002
Total	0.974	0.971	0.959	0.960	0.909	0.912	0.915	0.917	0.919	0.904	0.907	0.907	0.909	0.907
$t = 700$														
Fund 1	0.675	0.673	0.329	0.675	0.649	0.650	0.650	0.650	0.650	0.649	0.650	0.650	0.650	0.650
Fund 2	0.341	0.341	0.673	0.328	0.325	0.325	0.327	0.328	0.328	0.319	0.319	0.318	0.32	0.318
Interaction	0.000	0.000	0.000	0.000	−0.002	−0.002	−0.002	−0.002	−0.001	−0.001	−0.001	−0.001	−0.001	−0.001
Total	1.016	1.014	1.003	1.004	0.973	0.974	0.976	0.977	0.977	0.967	0.968	0.967	0.969	0.967

A.1. Theoretical rules

The presentation of the results for theoretical rules (known coefficients' case) is organized by first studying Data Set I, via inspection of Table 5. Afterwards, the discussion is extended to the other five data sets in Table 4, by highlighting common features across them (see the theoretical rules in Tables 5–9).

Data Set I

Consider Data Set I in Table 5 and let $t = 100$. Comparing expected out-of-sample utilities, it is clear that all strategies are very close to each other.[18] In fact, using this performance metric, it is easy to verify that Proposition 3 holds for all data sets, in the sense that $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{M}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$ and $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Q}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$, and we are using the fact that $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}}a)] = \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{KZ}})]$ and $\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}_a})] = \mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}_{\mathcal{Y}})]$. Moreover, it is also immediate to note that the performances of all rules increase with the sample size (from $t = 100$ to $t = 700$), while the aforementioned rankings of the metrics remains the same for all sample sizes.

[18] To highlight the differences between the rules it was necessary to scale the expected utility by 100, as mentioned in the tables.

Table 9
Data set V: operating profitability and investment, $n = 25$. For clarity, all quantities have been multiplied by 100.

Data set V	Theoretical rules				Implementable rules w.o. KZ correction					Implementable rules w. KZ correction				
Components	$\mathcal{Q}$	$\mathcal{M}$	$\mathcal{KZ}$	$\mathcal{Y}$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$
$t = 100$														
Fund 1	0.478	0.465	0.295	0.478	0.075	0.316	0.316	0.316	0.316	0.075	0.316	0.316	0.316	0.316
Fund 2	0.308	0.308	0.465	0.294	−0.361	−0.361	−0.277	0.172	0.180	0.154	0.154	0.174	0.243	0.242
Interaction	0.000	0.000	0.000	0.000	−0.102	−0.073	−0.07	−0.048	−0.047	−0.056	−0.040	−0.038	−0.027	−0.026
Total	0.786	0.773	0.759	0.772	−0.388	−0.117	−0.031	0.440	0.449	0.172	0.430	0.452	0.532	0.532
$t = 300$														
Fund 1	0.582	0.578	0.717	0.582	0.508	0.527	0.527	0.527	0.527	0.508	0.527	0.527	0.527	0.527
Fund 2	0.734	0.734	0.578	0.716	0.631	0.631	0.643	0.691	0.694	0.667	0.667	0.668	0.671	0.669
Interaction	0.000	0.000	0.000	0.000	−0.015	−0.013	−0.013	−0.012	−0.012	−0.011	−0.010	−0.010	−0.009	−0.009
Total	1.316	1.312	1.295	1.298	1.125	1.145	1.157	1.206	1.209	1.164	1.185	1.186	1.190	1.187
$t = 500$														
Fund 1	0.602	0.600	0.922	0.602	0.563	0.570	0.570	0.570	0.570	0.563	0.570	0.570	0.570	0.570
Fund 2	0.937	0.937	0.600	0.920	0.898	0.898	0.902	0.916	0.917	0.900	0.900	0.899	0.899	0.897
Interaction	0.000	0.000	0.000	0.000	−0.007	−0.007	−0.007	−0.007	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006
Total	1.539	1.537	1.521	1.523	1.453	1.460	1.465	1.479	1.481	1.456	1.463	1.463	1.463	1.461
$t = 700$														
Fund 1	0.611	0.609	1.042	0.611	0.584	0.588	0.588	0.588	0.588	0.584	0.588	0.588	0.588	0.588
Fund 2	1.055	1.055	0.609	1.041	1.035	1.035	1.037	1.043	1.043	1.033	1.033	1.032	1.032	1.030
Interaction	0.000	0.000	0.000	0.000	−0.005	−0.005	−0.005	−0.004	−0.004	−0.005	−0.004	−0.004	−0.004	−0.004
Total	1.666	1.665	1.651	1.652	1.614	1.618	1.620	1.626	1.626	1.613	1.616	1.616	1.615	1.614

We can also analyze the contributions of Funds 1 and 2 in the overall performances (Total) when the sample size increases. In this case it is clear that, for all theoretical rules, the contributions of Fund 1 and Fund 2 increase as the sample size increases. However, the importance of each fund vary depending on the rule selected. In particular, for the $\mathcal{KZ}a$ rule, when the sample size is small ($t = 100$), Fund 2 is the most important, but this fact reverses as the sample size increases ($t = 700$), where Fund 1 ends up being the one contributing the most. However this exact fact is reversed for all the other theoretical rules. This observation suggests that for rules $\mathcal{Q}$, $\mathcal{M}$ and $\mathcal{Y}a$, portfolios associated to H can capitalize better than the ones associated to G the increments in sample size; but the opposite case occurs when the sample size increases in the $\mathcal{KZ}a$ rule.

Other data sets

The conclusions presented for Data Set I regarding performance ordering across rules is maintained in all data sets, for all sample sizes. The patterns observed for the performance composition in Data Set I, can be easily extended to Data Sets II and V. Data sets III and IV however show that the importance of Funds 1 and 2 do not vary with the sample size.

A.2. Implementable rules

We explore the results for implementable rules (unknown coefficients' case) in a fashion similar to the previous section. Once again, we study each portfolio rule in Table 4 and the summarized results are presented in Tables 5–9. We start by exclusively discussing the case corresponding to implementable rules without the KZ correction, and later extend the analysis to rules with the KZ correction in the last subsection.

Data Set I

Observing Table 5, it is clear that the performance of rule $\mathcal{QS}aI$ is greater or equal, in terms of expected out-of-sample utility to all other strategies for all sample sizes. This result suggests that the extra correction factors of $\mathcal{QS}a_I$ indeed improves performance over all other implementable alternatives. Its performance is mildly better than the one of $\mathcal{QS}_I$, which is the second best rule. Noticeably, portfolio rule $\mathcal{Q}_I$ (the leading strategy for the theoretical rules), is not close to $\mathcal{QS}_I$ in the performance ranking of the implementable rules, but for some values of t (see $t = 100$ for any data set), it is indeed the rule with the worst performance. This observation highlights the empirical importance of extra adjustments in $\mathcal{Q}_I$ to regulate the aggressiveness of the orthogonal portfolio components.

For implementable rules we also observe that the driving force of performance is Fund 2. It is also noticeable that the orthogonality of Fund 1 and Fund 2 in portfolio rules $\mathcal{Q}_I$, $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ is lost, and that the magnitude of the loss is inversely proportional to the sample size (see $t = 100$ and compare this result wit the case $t = 700$ where the interaction term has almost vanished). Also note that the rate of reduction of this interaction term is smaller for the $\mathcal{KZ}_I$ rule. It is also worthwhile noting that Propositions 4 and 5 (and all its numerals) hold for this data set, and thus rule $\mathcal{QS}_I$ performs better than $\mathcal{Q}_I$ and $\mathcal{QS}a_I$ performs better than $\mathcal{KZ}_I$.

Other data sets

It is clear that $\mathcal{QS}a_I$ is the best implementable rule in all data sets (see the columns corresponding to implementable rules w.o. KZ correction). However, the differences among all the rules are getting smaller as t increases, with a clear lead by $\mathcal{QS}a_I$, and a close second place given by $\mathcal{QS}_I$, with the exception of data set III, where the $\mathcal{KZ}$ rule provides a greater performance than the $\mathcal{QS}_I$ rule for $t = 500$ and $t = 700$. It is also worthwhile noting that similar from data set I, Fund 1 contributes the most to the overall performance for data sets II and V, but this does not hold for data sets III and IV, where Fund 2 has the most participation in the total. This fact is aligned with the observation that for data sets I, II and V it holds that $\theta_H^2 > \theta_G^2$, while for data sets III and IV it holds that $\theta_H^2 < \theta_G^2$ as shown in Table 4. Finally, when studying the evolution of the interaction in the remaining data sets, we observe that all rules portray a reduction in the interaction term as the sample size increases, with $\mathcal{KZ}_I$ being the rule that exhibits the slowest reduction rate.

Lastly, we stress that Propositions 4 and 5 (and all its numerals) hold for all the remaining data sets, with the exception of data set IV, where only 4 (1), 4 (2) and 5 (1) hold, being the reason that $\theta_H^2 > \mathrm{UB}_{H,1}$ and $\theta_H^2 > \mathrm{UB}_{H,2}$, where we have omitted the report of the aforementioned upper bounds.

Implementable Rules with KZ correction

Observing the results for the implementable rules with the KZ correction and comparing them with the case that does not include it, it is apparent that rule $\mathcal{QS}aI$ remains the lead, but $\mathcal{QS}I$ is no longer the clear second best (for large or small sample sizes), now sharing this spot with the $\mathcal{KZ}_I$ rule. This is a consequence of the impact of the KZ correction over the performances for all the rules, where it is apparent that the improvement is much greater for the $\mathcal{KZ}_I$ rule itself.

In general, rules $\mathcal{QS}a$ and $\mathcal{QS}$ work well for small data samples ($t = 100$), and are better than $\mathcal{KZ}$ for all data sets. It is also noticeable that for $t = 300$ and higher, all the rules have very similar performances. but the KZ correction does improve the rules that do not have extra factors such as $\mathcal{QS}a$ and $\mathcal{QS}$. Finally, we can say that the best recommendation for small data sets is to use $\mathcal{QS}a$ or $\mathcal{QS}$ with the KZ correction.

The composition of the performances is very qualitatively the same to what was presented for the implementable rules without the KZ correction, and the reasons for the differences in the importance of the funds are exactly the same.

A.3. Theoretical vs. implementable rules

We start by noticing, as expected, that the performances from the implementable rules are dominated by those of the theoretical rules. Moreover, it is worthwhile noting that the differences are larger for small samples. For example, if we use Data Set I with $t = 100$, then $\mathcal{V}^{\mathcal{Q}} = 0.5440$ but $\mathcal{V}^{\mathcal{Q}_I} = -0.6800$; but, when $t = 300$ we obtain that $\mathcal{V}^{\mathcal{Q}} = 0.9000$ and $\mathcal{V}^{\mathcal{Q}_I} = 0.6860$.

We stress that the performance ranking of the rules cannot be extrapolated from theoretical to implementable rules. Consider the case of $\mathcal{Q}$, for example. While this strategy is the best for the known coefficient case, it turns out to be a very bad strategy when the coefficients are unknown. This situation is even more dramatic when consider a small sample size scenario, where in many cases the overall expected out-of-sample utility of $\mathcal{Q}_I$ is negative.

A.4. Limiting performance

Next, the limiting performance of theoretical and implementable portfolio rules is empirically studied. In this section we assume that the limiting Sharpe ratios of G and H, $\bar{\theta}_G^2$ and $\bar{\theta}_H^2$, are equal to θ_G^2 and θ_H^2. The values of c in this section where chosen as the ratios of the number of assets of the corresponding data set (n) and $t \in \{100, 300, 500, 700\}$. For example, the third value of c for Data Set I in Table 10 is $c = 0.05$ which is equal to the ratio of $n = 25$ to $t = 500$.

First, for each data set of Table 10 we compute $\bar{\mathcal{V}}^{\mathcal{B}}$, given by expression (77), for different values of c. The column "Theoretical" in Table 10 contains limiting performance $\bar{\mathcal{V}}^{\mathcal{B}}$. Notice that theoretical portfolio rules $\mathcal{Q}$, $\mathcal{M}$, $\mathcal{Y}a$, $\mathcal{KZ}a$ have the same limiting performance, and it is similar to the theoretical ones of Tables 5–9 for the appropriate combination of n and t associated to c. Also, it is important to observe that this limiting performance is always lower than the performance of portfolio rule $\mathcal{Q}$.

For implementable rules $\mathcal{Q}_I$, $\mathcal{M}_I$, $\mathcal{KZ}_I$, $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ the limiting performance is obtained by $\bar{\mathcal{V}}^{\mathcal{I}}(x, y)$ in (81)) which, in turn, is equal the sum of $\bar{\mathcal{V}}_G^{\mathcal{I}}(x)$ in (83) and $\bar{\mathcal{V}}_H^{\mathcal{I}}(y)$ in (85) because the interaction term is equal to zero. The values of x and y for each implementable portfolio rule correspond to $\bar{x}_{\mathcal{I}}$ and $\bar{y}_{\mathcal{I}}$ given by (91) and (92), respectively. The limiting performances of the aforementioned implementable rules are also shown in Table 10 for different values of c. We observe that the limiting performances of $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ are always greater than the ones of $\mathcal{Q}_I$, $\mathcal{M}_I$ and $\mathcal{KZ}_I$. This implies that the extra factors of $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ improve their limiting performance; but, if c tends to zero, then this positive effect vanishes. Moreover, the limiting performances of $\mathcal{QS}_I$ and $\mathcal{QS}a_I$ are lower, but relatively close, to the corresponding ones of Tables 5–9 (without KZ correction) for the appropriate values of n and t. For example, for Data Set I ($n = 25$) and $t = 100$ the performance of $\mathcal{QS}a_I$ is 0.156 (see Table 5 while its limiting performance is 0.264 for $c = 0.25$ (see Table 10). Finally, the KZ correction for these two implementable rules can generate higher performance than its limiting one for small sample sizes. For example, for Data Set I and $t = 100$ the performance of $\mathcal{QS}a_I$ is 0.313 while its associated limiting performance is 0.264.

Appendix B. Proof of Proposition 2

It is clear that $\hat{\boldsymbol{\omega}}_G^\top \mathbf{1} = 1$ a.s. and $f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_H^\top \mathbf{1} = 0$ a.s. Then, following Kan et al. (2016), we define

$$\hat{\sigma}_{(G,H)} = \hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\omega}}_H \tag{93}$$

$$= \frac{1}{\gamma} \frac{\sqrt{t}}{\sqrt{v_2}} \sigma_G \hat{\theta}_H y, \tag{94}$$

Table 10
Limiting performances of theoretical and implementable portfolio rules. All quantities have been multiplied by 100.

c	Theoretical	$\mathcal{Q}_I$	$\mathcal{M}_I$	$\mathcal{KZ}_I$	$\mathcal{QS}_I$	$\mathcal{QS}a_I$
Data set 1						
0.250	0.537	−0.199	−0.158	−0.158	0.264	0.264
0.083	0.889	0.774	0.778	0.778	0.826	0.826
0.050	1.052	1.014	1.016	1.016	1.032	1.032
0.036	1.152	1.136	1.136	1.136	1.144	1.144
Data set 2						
0.320	0.833	0.000	0.085	0.085	0.693	0.693
0.107	1.614	1.515	1.522	1.522	1.589	1.589
0.064	1.941	1.913	1.916	1.916	1.936	1.936
0.046	2.126	2.114	2.116	2.116	2.124	2.124
Data set 3						
0.100	0.438	0.055	0.060	0.060	0.139	0.139
0.033	0.477	0.373	0.374	0.374	0.383	0.383
0.020	0.490	0.438	0.438	0.438	0.442	0.442
0.014	0.498	0.467	0.467	0.467	0.469	0.469
Data set 4						
0.100	0.719	0.457	0.464	0.464	0.541	0.541
0.033	0.888	0.850	0.851	0.851	0.859	0.859
0.020	0.961	0.949	0.950	0.950	0.952	0.952
0.014	1.005	1.000	1.000	1.000	1.001	1.001
Data set 5						
0.250	0.776	0.112	0.165	0.165	0.574	0.574
0.083	1.302	1.216	1.221	1.221	1.266	1.266
0.050	1.526	1.501	1.502	1.502	1.516	1.516
0.036	1.655	1.645	1.646	1.646	1.652	1.652

where

$$\hat{\theta}_H^2 = \frac{z_2^2 + u_0}{v_2} \tag{95}$$

$$y = \frac{a}{\sqrt{v_2}} + \frac{x_{21}}{\sqrt{w_2}} y_1 + \frac{\sqrt{s_2}}{\sqrt{w_2}} y_2 \tag{96}$$

$$y_1 = \frac{x_{11}}{\sqrt{w_1}} + \frac{b x_{21}}{\sqrt{w_1}\sqrt{w_2}} + \frac{a x_{21}}{\sqrt{v_2}\sqrt{w_2}}, \tag{97}$$

$$y_2 = \frac{c}{\sqrt{w_1}} + \frac{b\sqrt{s_2}}{\sqrt{w_1}\sqrt{w_2}} + \frac{a\sqrt{s_2}}{\sqrt{v_2}\sqrt{w_2}}, \tag{98}$$

and the random variables $z_2 \sim \mathcal{N}(\sqrt{t}\theta_H, 1)$, $u_0 \sim \chi^2_{n-2}$, $v_2 \sim \chi^2_{t-n+1}$, $w_1 \sim \chi^2_{t-n+3}$, $w_2 \sim \chi^2_{t-n+2}$, $s_1 \sim \chi^2_{n-4}$, $s_2 \sim \chi^2_{n-3}$, $x_{11} \sim \mathcal{N}(0,1)$, $x_{21} \sim \mathcal{N}(0,1)$, $a \sim \mathcal{N}(0,1)$, $b \sim \mathcal{N}(0,1)$, $c \sim \mathcal{N}(0,1)$ are independent of each other. Then, it straightforward to verify that $\mathbb{E}[f(\hat{\theta}_H^2)\frac{1}{\gamma}\frac{\sqrt{t}}{\sqrt{v_2}}\sigma_G\hat{\theta}_H y] = 0$ and also $\mathbb{E}[\hat{\sigma}_{(G,H)}] = 0$. Finally, expressions (A72) and (A76) of Kan et al. (2016) imply directly that $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}, f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}) = 0$.

Appendix C. Orthogonality proofs

C.1. Pairs $(\hat{\boldsymbol{\omega}}_{\tilde{G}}, \hat{\boldsymbol{\omega}}_H)$ and $(\hat{\boldsymbol{\omega}}_{\tilde{G}}, \hat{\boldsymbol{\omega}}_{\tilde{H}})$

In this part of the appendix we present the orthogonality results of the pairs $(\hat{\boldsymbol{\omega}}_{\tilde{G}}, \hat{\boldsymbol{\omega}}_H)$ and $(\hat{\boldsymbol{\omega}}_{\tilde{G}}, \hat{\boldsymbol{\omega}}_{\tilde{H}})$. First, we aim to proof the next two expectations

$$\mathbb{E}[\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \mu_G\mathbf{1})] = 0, \tag{99}$$

$$\mathbb{E}[\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{R}}\hat{\boldsymbol{\mu}}] = 0. \tag{100}$$

Using the fact that $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are independent, it is equivalent to verify the following

$$\mathbb{E}[\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] - \mu_G\mathbb{E}[\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}] = 0, \tag{101}$$

$$\mathbb{E}\left[\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\right] - \mathbb{E}\left[\frac{(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}) \times (\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = 0. \tag{102}$$

Following the proof of expression (77) of Kan and Zhou (2007), we define the orthonormal vectors $\boldsymbol{v} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{1}/(\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1})^{\frac{1}{2}}$ and $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\mu} - \mu_G\mathbf{1})/\theta_H$ and the $n \times (n-1)$ orthonormal matrix $\boldsymbol{Q}$ with its columns orthogonal to $\boldsymbol{v}$ and its first column equal to $\boldsymbol{\eta}$. Consequently, the matrix $\boldsymbol{V} = [\boldsymbol{v}\ \boldsymbol{Q}]$ is an orthonormal basis of $\mathbb{R}^n$. The matrix $\boldsymbol{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim \mathcal{W}_n(t-1, \boldsymbol{I}_n)/t$ where $\mathcal{W}_n(t-1, \boldsymbol{I}_n)$ denotes a Wishart distribution with $t-1$ degrees of freedom and covariance matrix equal to the $n \times n$ identity matrix $\boldsymbol{I}_n$.

By Theorem 3.2.11 of Muirhead (1982), the $n \times n$ matrix $\boldsymbol{A} = (\boldsymbol{V}^\top\boldsymbol{W}^{-1}\boldsymbol{V})^{-1}$ is such that $\boldsymbol{A} \sim \mathcal{W}_n(t-1, \boldsymbol{I}_n)/t$ and

$$\boldsymbol{V}^\top\boldsymbol{W}^{-1}\boldsymbol{V} = \left[\begin{array}{c|c} \boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v} & \boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} \\ \hline \boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{v} & \boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} \end{array}\right]. \tag{103}$$

We partition $\boldsymbol{A}$ in the following way

$$\boldsymbol{A} = \left[\begin{array}{c|c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \hline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array}\right], \tag{104}$$

where $\boldsymbol{A}_{11}$ is 1×1 and $\boldsymbol{A}_{22}$ is $(n-1)\times(n-1)$. By the partitioned matrix inverse formula, we have

$$\boldsymbol{A}_{11} = \frac{1}{\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v}}, \qquad \boldsymbol{A}_{12} = -\frac{\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{Q}\boldsymbol{\Theta}^{-1}}{\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v}}, \qquad \boldsymbol{A}_{21} = \boldsymbol{A}_{12}^\top, \qquad \boldsymbol{A}_{22} = \boldsymbol{\Theta}^{-1} \tag{105}$$

and

$$\boldsymbol{\Theta} = \boldsymbol{Q}^\top\left(\boldsymbol{W}^{-1} - \frac{\boldsymbol{W}^{-1}\boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{W}^{-1}}{\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v}}\right)\boldsymbol{Q}. \tag{106}$$

By Theorem 3.2.10 of Muirhead (1982), we have

$$u = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} \sim \chi^2_{t-n}/t, \tag{107}$$

$$\boldsymbol{z} = -\boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{A}_{21} \sim \mathcal{N}\left(\mathbf{0}_{n-1}, \boldsymbol{I}_{n-1}/t\right), \tag{108}$$

$$\boldsymbol{A}_{22} \sim \mathcal{W}_{n-1}(t-1, \boldsymbol{I}_{n-1})/t, \tag{109}$$

and they are independent of each other. Let $\boldsymbol{e}_1$ be the $(n-1)$-vector such that $\boldsymbol{e}_1 = [1\ 0\ \cdots\ 0]^\top$. Kan and Zhou (2007) proved the following expressions

$$\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1} = \frac{1}{\sigma_G^2}\times\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v} = \frac{1}{\sigma_G^2}\times\frac{1}{u}, \tag{110}$$

$$\begin{aligned}\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} &= \frac{\mu_G}{\sigma_G^2}\times\boldsymbol{v}^\top\boldsymbol{W}^{-1}\boldsymbol{v} + \frac{\theta_H}{\sigma_G}\times\boldsymbol{\eta}^\top\boldsymbol{W}^{-1}\boldsymbol{v} \\ &= \frac{\mu_G}{\sigma_G^2}\times\frac{1}{u} + \frac{\theta_H}{\sigma_G}\times\frac{\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z}}{u},\end{aligned} \tag{111}$$

$$\begin{aligned}\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1} &= \frac{1}{\sigma_G^2}\times\boldsymbol{v}^\top\boldsymbol{W}^{-2}\boldsymbol{v} = \frac{1}{\sigma_G^2}\times\boldsymbol{v}^\top\boldsymbol{W}^{-1}(\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{Q}\boldsymbol{Q}^\top)\boldsymbol{W}^{-1}\boldsymbol{v} \\ &= \frac{1}{\sigma_G^2}\times\frac{(1+\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})}{u^2}.\end{aligned} \tag{112}$$

Using these results, we can express the second expectation of (102) as

$$S_2 = \mathbb{E}\left[\frac{(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})\times(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] \tag{113}$$

$$= \mathbb{E}\left[\frac{\left(\frac{1}{\sigma_G^2}\times\frac{(1+\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})}{u^2}\right)\times\left(\frac{\mu_G}{\sigma_G^2}\times\frac{1}{u}+\frac{\theta_H}{\sigma_G}\times\frac{\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}}{u}\right)}{\left(\frac{1}{\sigma_G^2}\times\frac{1}{u}\right)}\right] \tag{114}$$

$$= \mathbb{E}\left[\frac{1}{u^2}\right]\times\mathbb{E}\left[\frac{\mu_G}{\sigma_G^2}(1+\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}) + \frac{\theta_H}{\sigma_G}(1+\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})(\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})\right]. \tag{115}$$

By Theorem 3.2.12 of Muirhead (1982), we have $\frac{\boldsymbol{z}^\top\boldsymbol{z}}{\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}} \sim \chi^2_{t-1-(n-1)+1}$ and it is independent of $\boldsymbol{z}$. This fact implies that $\mathbb{E}[\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}] = \frac{n-1}{t-n-1}$. Using (107) and (108), we have that $\mathbb{E}[\frac{1}{u^2}] = \frac{t^2}{(t-n-2)(t-n-4)}$ and $\mathbb{E}[(\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})(\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})|\boldsymbol{A}_{22}] = 0$. Therefore, S_2 is given by

$$S_2 = \frac{\mu_G}{\sigma_G^2}\times\frac{t^2(t-2)}{(t-n-1)(t-n-2)(t-n-4)}. \tag{116}$$

Next, we will find the first expectation of (102). The main goal is to express the numerator of the expectation as a function of u, z and $\boldsymbol{A}_{22}$. It can be verified that

$$\frac{\theta_H}{\sigma_G} \times \boldsymbol{v}^\top \boldsymbol{W}^{-2} \boldsymbol{\eta} = \mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu} - \mu_G \mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}. \tag{117}$$

this implies that

$$S_1 = \mathbb{E}\left[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}\right] \tag{118}$$

$$= \frac{\theta_H}{\sigma_G} \times \mathbb{E}\left[\boldsymbol{v}^\top \boldsymbol{W}^{-2} \boldsymbol{\eta}\right] + \mu_G \times \mathbb{E}\left[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right]. \tag{119}$$

Using (110) and (112), the second expectation of S_1 in (119) is

$$S_{12} = \mu_G \times \mathbb{E}\left[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}\right] = \frac{\mu_G}{\sigma_G^2} \times \mathbb{E}\left[\frac{(1 + \boldsymbol{z}^\top \boldsymbol{A}_{22}^{-1} \boldsymbol{z})}{u^2}\right] \tag{120}$$

$$= \frac{\mu_G}{\sigma_G^2} \times \frac{t^2(t-2)}{(t-n-1)(t-n-2)(t-n-4)}. \tag{121}$$

Due to the fact that S_{12} is equal to S_2 in (116), we only need to demonstrate that

$$S_{11} = \frac{\theta_H}{\sigma_G} \times \mathbb{E}\left[\boldsymbol{v}^\top \boldsymbol{W}^{-2} \boldsymbol{\eta}\right] = 0 \tag{122}$$

in order to prove both (101) and (102). However, a previous step is to prove that

$$\zeta = \boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta} = \frac{\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-\frac{3}{2}} \boldsymbol{e}_1}{u} + u \times \frac{(\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-1} \boldsymbol{z})}{u^2} \times \frac{(\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}} \boldsymbol{z})}{u}. \tag{123}$$

We start with the left-hand side of (123)

$$\zeta = -\frac{\boldsymbol{A}_{21}^\top \boldsymbol{A}_{22}^{-2} \boldsymbol{e}_1}{u} + u \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{v}) \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta}) \tag{124}$$

$$= \boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{\Theta} \boldsymbol{e}_1 + u \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{v}) \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta}) \tag{125}$$

$$= \boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q}\left[\boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta} - u \times (\boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{v}) \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta})\right]$$
$$+ u \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{v}) \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta}) \tag{126}$$

$$= \boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta}. \tag{127}$$

Consequently, the numerator inside the expectation of (122) is

$$\boldsymbol{v}^\top \boldsymbol{W}^{-2} \boldsymbol{\eta} = (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{v}) \times (\boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta}) + \boldsymbol{v}^\top \boldsymbol{W}^{-1} \boldsymbol{Q} \boldsymbol{Q}^\top \boldsymbol{W}^{-1} \boldsymbol{\eta} \tag{128}$$

$$= \frac{1}{u} \times \frac{\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}} \boldsymbol{z}}{u} + \frac{\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-\frac{3}{2}} \boldsymbol{e}_1}{u} + \frac{(\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-1} \boldsymbol{z}) \times (\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}} \boldsymbol{z})}{u^2} \tag{129}$$

which in turn implies that S_{11} in (122) is

$$S_{11} = \frac{\theta_H}{\sigma_G} \times \mathbb{E}\left[\frac{\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}} \boldsymbol{z}}{u^2} + \frac{\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-\frac{3}{2}} \boldsymbol{e}_1}{u} + \frac{(\boldsymbol{z}^\top \boldsymbol{A}_{22}^{-1} \boldsymbol{z}) \times (\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}} \boldsymbol{z})}{u^2}\right] = 0 \tag{130}$$

and the proof is complete for expressions (99) and (100).

Next we will prove the following:

$$\mathrm{Cov}(\hat{\boldsymbol{\omega}}_{\mathrm{G}}^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{\mathrm{H}}^\top \boldsymbol{\mu}) \neq 0, \tag{131}$$

$$\mathrm{Cov}(\hat{\boldsymbol{\omega}}_{\mathrm{G}}^\top \boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{H}^\top \boldsymbol{\mu}) = 0. \tag{132}$$

To prove (131) we have to show that

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})] \neq \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}]\ \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})]. \tag{133}$$

Working with the right-hand side of (133), we have

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}] = \frac{t}{t-n-2} \frac{\mu_G}{\sigma_G^2}, \tag{134}$$

$$\mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})] = \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\mu}] - \mu_G \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}] \tag{135}$$

$$= \frac{t\theta_S^2}{t-n-2} - \frac{t\theta_G^2}{t-n-2} \tag{136}$$

$$= \frac{t\theta_H^2}{t-n-2}. \tag{137}$$

Consequently, (134) and (137) yield

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}]\ \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})] = \frac{t^2\theta_H^2}{(t-n-2)^2}\frac{\mu_G}{\sigma_G^2}. \tag{138}$$

Working with the left-hand side of (133) and by the independence of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, we have

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})] = \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] - \mu_G \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}]. \tag{139}$$

By Corollary 14 of Styan (1989), the expectation of $\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}$ is

$$\mathbb{E}[\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}] = \frac{t^2\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}}{(t-n-1)(t-n-4)} + \frac{t^2\theta_S^2\ \boldsymbol{\Sigma}^{-1}}{(t-n-1)(t-n-2)(t-n-4)}, \tag{140}$$

which implies that

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] = \frac{t^2\theta_S^2}{(t-n-2)(t-n-4)}\frac{\mu_G}{\sigma_G^2}, \tag{141}$$

$$\begin{aligned}\mu_G \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}] &= \frac{t^2\theta_G^2}{(t-n-1)(t-n-4)}\frac{\mu_G}{\sigma_G^2} + \frac{t^2\theta_S^2}{(t-n-1)(t-n-2)(t-n-4)}\frac{\mu_G}{\sigma_G^2} \\ &= \frac{t^2}{(t-n-2)(t-n-4)}\left[\frac{(t-n-2)\theta_G^2}{t-n-1} + \frac{\theta_S^2}{t-n-1}\right]\frac{\mu_G}{\sigma_G^2}.\end{aligned} \tag{142}$$

Using expressions (141) and (142), the expectation in (139) reduces to:

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \mu_G \mathbf{1})] = \frac{t^2\theta_H^2}{(t-n-1)(t-n-4)}\frac{\mu_G}{\sigma_G^2}, \tag{143}$$

which is different from (138). Then, expression (131) holds.

To prove (132) we need to show that

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{R}}\boldsymbol{\mu}] = \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}]\ \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{R}}\boldsymbol{\mu}]. \tag{144}$$

Working with the right-hand side of (144), we have

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] = \frac{t}{t-n-2}\frac{\mu_G}{\sigma_G^2}, \tag{145}$$

$$\mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{R}}\boldsymbol{\mu}] = \frac{t\ \theta_H^2}{t-n-1}, \tag{146}$$

and

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}]\ \mathbb{E}[\boldsymbol{\mu}^\top \hat{\boldsymbol{R}}\boldsymbol{\mu}] = \frac{t^2\theta_H^2}{(t-n-2)(t-n-1)}\frac{\mu_G}{\sigma_G^2}. \tag{147}$$

The left-hand side of (144) can be expressed as

$$\mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{R}}\boldsymbol{\mu}] = \mathbb{E}[\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}] - \mathbb{E}\left[\frac{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right]. \tag{148}$$

The first term of the right-hand side of (148) is given in (141). Using (147), to prove (132) we only need to show that:

$$\mathbb{E}\left[\frac{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \frac{t^2}{(t-n-2)(t-n-4)}\left(\theta_G^2 + \frac{3\theta_H^2}{t-n-1}\right)\frac{\mu_G}{\sigma_G^2}. \tag{149}$$

Replacing (110) and (111) in $\mathbb{E}\left[(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3/\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right]$ and using the independence properties, we have:

$$\mathbb{E}\left[\frac{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3}{\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \sigma_G^2\ \mathbb{E}\left[\frac{1}{u^2}\right]\ \mathbb{E}\left[\left(\frac{\mu_G}{\sigma_G^2} + \frac{\theta_H}{\sigma_G} \times \boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z}\right)^3\right] \tag{150}$$

$$= \sigma_G^2\ \mathbb{E}\left[\frac{1}{u^2}\right]\ \mathbb{E}\left[\left(\frac{\mu_G}{\sigma_G^2}\right)^3 + \frac{3\theta_H^2}{\sigma_G^2}\frac{\mu_G}{\sigma_G^2}(\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z})^2\right] \tag{151}$$

$$= \mathbb{E}\left[\frac{1}{u^2}\right]\left(\frac{\mu_G}{\sigma_G^2}\theta_G^2 + 3\frac{\mu_G}{\sigma_G^2}\theta_H^2\mathbb{E}\left[(\boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z})^2\right]\right) \tag{152}$$

$$= \frac{t^2}{(t-n-2)(t-n-4)}\left(\theta_G^2 + \frac{3\theta_H^2}{t-n-1}\right)\frac{\mu_G}{\sigma_G^2}, \tag{153}$$

where the last equation holds because $\boldsymbol{e}_1^\top t^{\frac{1}{2}}(t\boldsymbol{A}_{22})^{-\frac{1}{2}}\boldsymbol{z}|(t\boldsymbol{A}_{22}) = \boldsymbol{X} \sim \mathcal{N}(0, \boldsymbol{e}_1^\top \boldsymbol{X}^{-1}\boldsymbol{e}_1)$ and $\mathbb{E}[(t\boldsymbol{A}_{22})^{-1}] = \frac{1}{t-n-1}\boldsymbol{I}_{n-1}$.

C.2. Pair $(\hat{\boldsymbol{\omega}}_G, \hat{\boldsymbol{\omega}}_{\breve{H}})$

In this part we will prove the following results:

$$\mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} - \frac{\mu_G}{\sigma_G^2}\frac{t}{t-n-2}\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})^2}\right] = 0, \tag{154}$$

$$\mathrm{Cov}(\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{\breve{H}}^\top\boldsymbol{\mu}) \neq 0. \tag{155}$$

We begin with the second term of the expectation in (154). Using the formula for the expectation of a quadratic form, we have

$$\mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\boldsymbol{\Sigma}\frac{\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \boldsymbol{\omega}_G^\top\boldsymbol{\Sigma}\boldsymbol{\omega}_G + \mathrm{tr}\left(\frac{\sigma_G^2}{t-n-1}\boldsymbol{\Sigma}\boldsymbol{R}\right) \tag{156}$$

$$= \sigma_G^2 + \frac{\sigma_G^2}{t-n-1}(n-1). \tag{157}$$

Then, the following equation holds

$$\mathbb{E}\left[\frac{\mu_G}{\sigma_G^2}\frac{t}{t-n-2}\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})^2}\right] = \mu_G \times \frac{t(t-2)}{(t-n-1)(t-n-2)}. \tag{158}$$

The first term of the expectation in (154) clearly satisfies

$$\mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right]. \tag{159}$$

Using the decompositions given by (119) and (129), we have

$$\mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \mathbb{E}\left[\frac{\frac{\theta_H}{\sigma_G}\times(\boldsymbol{v}^\top\boldsymbol{W}^{-2}\boldsymbol{\eta}) + \mu_G\times(\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})}{\frac{1}{\sigma_G^2}\times\frac{1}{u}}\right] \tag{160}$$

$$= \theta_H\sigma_G \times \mathbb{E}\left[\frac{\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z}}{u} + \boldsymbol{z}^\top\boldsymbol{A}_{22}^{-\frac{3}{2}}\boldsymbol{e}_1 + \frac{(\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z})\times(\boldsymbol{e}_1^\top\boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z})}{u}\right]$$

$$+ \mu_G \times \mathbb{E}\left[\frac{1+\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}}{u}\right]. \tag{161}$$

The first expectation in (161) is equal to zero. Using the fact that u is independent of $\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}$, together with $\mathbb{E}[\frac{1}{u}] = \frac{t}{t-n-2}$ and $\mathbb{E}[\boldsymbol{z}^\top\boldsymbol{A}_{22}^{-1}\boldsymbol{z}] = \frac{n-1}{t-n-1}$, we have

$$\mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right] = \mu_G \times \frac{t(t-2)}{(t-n-1)(t-n-2)}. \tag{162}$$

Moreover, (162) is equal to (158) and this implies that (154) holds and then $\mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\omega}}_{\breve{H}}^\top] = 0$.

Next, we define B_1 and B_2 such that $\mathrm{Cov}(\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}, \hat{\boldsymbol{\omega}}_{\breve{H}}^\top\boldsymbol{\mu}) = B_2 - B_1$, and $B_2 = \mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}\boldsymbol{\mu}^\top\hat{\boldsymbol{\omega}}_{\breve{H}}]$ and $B_1 = \mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}] \times \mathbb{E}[\hat{\boldsymbol{\omega}}_{\breve{H}}^\top\boldsymbol{\mu}]$. It is easy to verify that $\mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}] = \mu_G$ and $\mathbb{E}[\hat{\boldsymbol{\omega}}_{\breve{H}}^\top\boldsymbol{\mu}] = \frac{1}{\gamma}\frac{t\theta_H^2}{t-n-2}$. Using the two aforementioned expressions, we obtain

$$B_1 = \frac{\mu_G}{\gamma} \times \frac{t\theta_H^2}{t-n-2}. \tag{163}$$

We can decompose B_2 such that

$$B_2 = \frac{1}{\gamma}\left(B_{22} - \frac{\mu_G}{\sigma_G^2}\frac{t}{t-n-2}B_{21}\right), \tag{164}$$

where

$$B_{21} = \mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top\boldsymbol{\mu}\boldsymbol{\mu}^\top\hat{\boldsymbol{\omega}}_G], \tag{165}$$

$$B_{22} = \mathbb{E}\left[\frac{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right]. \tag{166}$$

Using the formula for the expectation of a quadratic form, B_{21} is given by:

$$B_{21} = \mu_G^2 + \mathrm{tr}\left(\boldsymbol{\mu}\boldsymbol{\mu}^\top \times \frac{\sigma_G^2}{t-n-1} \times \boldsymbol{R}\right) = \mu_G^2 + \frac{\sigma_G^2}{t-n-1}\theta_H^2. \tag{167}$$

By arguments similar to those of Proposition 2, we have the following equation

$$\gamma \times \mathbb{E}[\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\mu}\boldsymbol{\mu}^\top \hat{\boldsymbol{\omega}}_H] = \mu_G \times \frac{t\theta_H^2}{t-n-1} = B_{22} - \mathbb{E}\left[\frac{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3}{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})^2}\right]. \tag{168}$$

Moreover, (150) and (153) imply that

$$\mathbb{E}\left[\frac{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu})^3}{(\mathbf{1}^\top \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1})^2}\right] = \mathbb{E}\left[(\sigma_G^2)^2 \times u^2 \times \frac{1}{u^3} \times \left(\frac{\mu_G}{\sigma_G^2} + \frac{\theta_H}{\sigma_G} \times \boldsymbol{e}_1^\top \boldsymbol{A}_{22}^{-\frac{1}{2}}\boldsymbol{z}\right)^3\right] \tag{169}$$

$$= \mu_G \times \frac{t}{t-n-2}\left(\theta_G^2 + \frac{3\theta_H^2}{t-n-1}\right), \tag{170}$$

and by Eq. (168), B_{22} is given by

$$B_{22} = \mu_G \times \frac{t\theta_H^2}{t-n-1} + \mu_G \times \frac{t}{t-n-2}\left(\theta_G^2 + \frac{3\theta_H^2}{t-n-1}\right). \tag{171}$$

Replacing (167) and (171) in (164), and after some simplifications we have

$$B_2 = \frac{\mu_G}{\gamma} \times \frac{t-n}{t-n-1} \times \frac{t\theta_H^2}{t-n-2}. \tag{172}$$

Notice that B_2 in (172) and B_1 in (163) are different and therefore (155) holds.

Appendix D. Corrections due to random coefficients

Consider the following portfolio rule with weights $x, y \in \mathbb{R}/\{0\}$ and f a function of $\hat{\theta}_H^2$

$$\hat{\boldsymbol{\omega}}(x,y) = x \times \frac{1}{\gamma}\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\hat{\boldsymbol{\omega}}_G + y \times f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_H, \tag{173}$$

with expected out-of-sample utility

$$\begin{aligned}\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}(x,y))] = {} & \frac{x}{\gamma}\boldsymbol{\mu}^\top \mathbb{E}\left[\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\hat{\boldsymbol{\omega}}_G\right] + y\boldsymbol{\mu}^\top \mathbb{E}\left[f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_H\right] \\ & -\frac{\gamma}{2}\frac{x^2}{\gamma^2}\mathbb{E}\left[\left(\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right)^2 \hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_G\right] - \frac{\gamma}{2}y^2\mathbb{E}\left[f(\hat{\theta}_H^2)^2\hat{\boldsymbol{\omega}}_H^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_H\right] \\ & -\gamma\frac{xy}{\gamma}\mathbb{E}\left[\left(\frac{\hat{\mu}_G}{\hat{\sigma}_G^2}\right) f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_G^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\omega}}_H\right].\end{aligned} \tag{174}$$

Renaming the terms in (174), we have

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol{\omega}}(x,y))] = \frac{x}{\gamma}\mathrm{ev}_{uG} + y\mathrm{ev}_{gH} - \frac{x^2}{2\gamma}\mathrm{var}_{uG} - \frac{\gamma}{2}y^2\mathrm{var}_{gH} - xy\mathrm{cov}_{uG,gH}, \tag{175}$$

where for portfolio rules I and J, ev_I is its expected out-of-sample mean of I, var_I is the expected out-of-sample variance of I; and, $\mathrm{cov}_{I,J}$ is the expected out-of-sample covariance between I and J. Clearly, portfolio rules uG and gH in (175) are $\frac{\hat{\mu}_G}{\hat{\sigma}_G^2} \times \hat{\boldsymbol{\omega}}_G$ and $f(\hat{\theta}_H^2)\hat{\boldsymbol{\omega}}_H$, respectively.

Using the results of Proposition 3 of Kan et al. (2016), we have

$$\mathrm{ev}_{uG} = \frac{t\theta_G^2}{t-n-2} + \frac{t\theta_H^2}{(t-n-2)(t-n-1)}, \tag{176}$$

$$\mathrm{ev}_{gH} = \frac{1}{\gamma}\frac{t\theta_H^2}{t-n-1}\mathbb{E}[f(Y_3)], \tag{177}$$

$$\begin{aligned}\mathrm{var}_{uG} = {} & \frac{t^2(t-2)\theta_G^2}{(t-n-1)(t-n-2)(t-n-4)} + \frac{t^2(t-2)\theta_H^2}{(t-n-1)(t-n-2)(t-n-3)(t-n-4)} \\ & + \frac{t(t-2)(t-4)}{(t-n-1)(t-n-2)(t-n-3)(t-n-4)},\end{aligned} \tag{178}$$

$$\mathrm{var}_{gH} = \frac{1}{\gamma^2}\frac{t(t-2)}{(t-n-1)(t-n)}\mathbb{E}\left[f(Y_4)^2 Y_4\right], \tag{179}$$

$$\mathrm{cov}_{uG,gH} = \frac{1}{\gamma}\frac{t(t-2)}{(t-n-2)(t-n-1)(t-n)}\mathbb{E}\left[f(Y_4)Y_4\right], \tag{180}$$

where $t > n+4$, $Y_3 \sim G^{t\theta^2_H}_{n+1,t-n-1}$ and $Y_4 \sim G^{t\theta^2_H}_{n-1,t-n-1}$, where $G^{\delta}_{p,q} = \frac{p}{q}\times F^{\delta}_{p,q}$, and $F^{\delta}_{p,q}$ denotes a non-central F distribution with p and q degrees of freedom and with non-centrality parameter equal to δ.

Given some initial weights $x^0_{uG} = \frac{t-n-1}{t-2}$, $y^0_{gH} = \frac{(t-n-3)(t-n)}{t(t-2)}$ motivated by portfolio rule Q in Table 3 and assuming function f to be defined as

$$f(\hat\theta^2_H) = \frac{\hat\theta^2_H}{\theta^2_H + \frac{n-1}{t}}, \tag{181}$$

we want to find scalars α^*_{uG} and α^*_{gH}, only depending on t and n, such that

$$\mathbb{E}[\mathcal{U}(\hat{\boldsymbol\omega}(\alpha^*_{uG}\times x^0_{uG}, \alpha^*_{gH}\times y^0_{gH}))] > \mathbb{E}[\mathcal{U}(\hat{\boldsymbol\omega}(x^0_{uG}, y^0_{gH}))]. \tag{182}$$

Noticing that the expectations in (177) and (179) can be found in closed form using the raw moments of the non-central F distribution, we can find

$$\alpha^{**}_{uG} = \operatorname*{argmax}_{\alpha_{uG}}\left\{\left(\mathrm{ev}_{uG}\times\frac{x^0_{uG}}{\gamma}\right)\times\alpha_{uG} - \frac{\gamma}{2}\left[\mathrm{var}_{uG}\times\left(\frac{x^0_{uG}}{\gamma}\right)^2\right]\times(\alpha_{uG})^2\right\}, \tag{183}$$

$$\alpha^{**}_{gH} = \operatorname*{argmax}_{\alpha_{gH}}\left\{(\mathrm{ev}_{gH}\ y^0_{gH})\times\alpha_{gH} - \frac{\gamma}{2}\left[\mathrm{var}_{gH}\times(y^0_{gH})^2\right]\times(\alpha_{gH})^2\right\}, \tag{184}$$

and they are given by

$$\alpha^{**}_{uG} = \frac{t-n-4}{t}\times\frac{\theta^2_G + \frac{\theta^2_H}{t-n-1}}{\theta^2_G + \frac{\theta^2_H}{t-n-3} + \frac{t-4}{t(t-n-3)}},$$

$$\alpha^{**}_{gH} = c_{gH}\times\frac{(t\theta^2_H)^3 + 2n(t\theta^2_H)^2 + (n-1)(n+1)(t\theta^2_H)}{(t\theta^2_H)^3 + 4(n+3)(t\theta^2_H)^2 + 3(n+1)(n+3)(t\theta^2_H) + (n-1)(n+1)(n+3)},$$

where

$$c_{gH} = \frac{(t-n-7)(t-n-5)}{(t-n-3)t}.$$

Clearly $0 < \alpha^{**}_{uG} < 1$ and $0 < \alpha^{**}_{gH} < 1$, then the expected out-of-sample performance improves by taking

$$\alpha^*_{uG} = \frac{t-n-4}{t}, \tag{185}$$

$$\alpha^*_{gH} = \frac{(t-n-7)(t-n-5)}{(t-n-3)t}. \tag{186}$$

It is important to recall that coefficients α^*_{uG} and α^*_{gH} in (185) and (186) generate portfolio rule $\hat{\boldsymbol\omega}(\alpha^*_{uG}\times x^0_{uG}, \alpha^*_H\times y^0_{uG})$ which is slightly different from $\mathcal{QS}_I$ in (45). More specifically, these portfolio rules differ because the former considers factor $\hat\theta^2_H/(\theta^2_H + \frac{n-1}{t})$ for $\hat{\boldsymbol\omega}_H$ while the latter considers factor $\hat\theta^2_H/(\hat\theta^2_H + \frac{n-1}{t})$ (see Table 3 for more details). Similarly, portfolio rule $\hat{\boldsymbol\omega}(x^0_{uG}, y^0_{gH})$ differs from $\mathcal{Q}_I$ in exactly the same factors. Therefore, we have not shown here that $\mathcal{QS}_I$ has greater expected out-of-sample utility than $\mathcal{Q}_I$.

Appendix E. Functions $\kappa_{P_I}(t;n,\gamma)$

Functions for the $\mathcal{KZ}_I$, rule:

$$\kappa_{\mathcal{KZ}_I,1}(t;n,\gamma) = \frac{(t-n-1)(t-n-4)}{2\gamma(t-2)(t-n-2)}, \tag{187}$$

$$\kappa_{\mathcal{KZ}_I,2}(t;n,\gamma) = \frac{(t-n-4)(t-n-5)}{2\gamma(t-2)(t-n-2)(t-n-3)}, \tag{188}$$

$$\kappa_{\mathcal{KZ}_I,3}(t;n,\gamma) = -\frac{(t-4)(t-n-1)(t-n-4)}{2\gamma t(t-2)(t-n-2)(t-n-3)}, \tag{189}$$

$$\kappa_{\mathcal{KZ}_I,4}(t;n,\gamma) = \frac{(t-n-4)}{\gamma(t-2)}, \tag{190}$$

$$\kappa_{\mathcal{KZ}_I,5}(t;n,\gamma) = -\frac{(t-n-1)(t-n-4)^2}{2\gamma t(t-2)(t-n)}, \tag{191}$$

$$\kappa_{\mathcal{KZ}_I,6}(t;n,\gamma) = -\frac{(t-n-1)(t-n-4)^2}{\gamma t(t-2)(t-n)(t-n-2)}. \tag{192}$$

Functions for the Q_I rule:

$$\kappa_{Q_I,1}(t;n,\gamma) = \frac{t(t-n-1)(t-2n-8)}{2\gamma(t-2)(t-n-2)(t-n-4)}, \tag{193}$$

$$\kappa_{Q_I,2}(t;n,\gamma) = \frac{t(t^2-3nt+2n(n+7)-13t+24)}{2\gamma(t-2)(t-n-2)(t-n-3)(t-n-4)}, \tag{194}$$

$$\kappa_{Q_I,3}(t;n,\gamma) = -\frac{t(t-4)(t-n-1)}{2\gamma(t-2)(t-n-2)(t-n-3)(t-n-4)}, \tag{195}$$

$$\kappa_{Q_I,4}(t;n,\gamma) = \frac{(t-n)(t-n-3)}{\gamma(t-2)(t-n-1)}, \tag{196}$$

$$\kappa_{Q_I,5}(t;n,\gamma) = -\frac{(t-n)(t-n-3)^2}{2\gamma t(t-2)(t-n-1)}, \tag{197}$$

$$\kappa_{Q_I,6}(t;n,\gamma) = -\frac{(t-n-3)}{\gamma(t-2)(t-n-2)}. \tag{198}$$

Functions for the $\mathcal{M}_I$ rule:

$$\kappa_{\mathcal{M}_I,1}(t;n,\gamma) = \frac{(t-n-1)(t-n-4)}{2\gamma(t-2)(t-n-2)}, \tag{199}$$

$$\kappa_{\mathcal{M}_I,2}(t;n,\gamma) = \frac{(t-n-4)(t-n-5)}{2\gamma(t-2)(t-n-2)(t-n-3)}, \tag{200}$$

$$\kappa_{\mathcal{M}_I,3}(t;n,\gamma) = -\frac{(t-4)(t-n-1)(t-n-4)}{2\gamma t(t-2)(t-n-2)(t-n-3)}, \tag{201}$$

$$\kappa_{\mathcal{M}_I,4}(t;n,\gamma) = \frac{(t-n)(t-n-3)}{\gamma(t-2)(t-n-1)}, \tag{202}$$

$$\kappa_{\mathcal{M}_I,5}(t;n,\gamma) = -\frac{(t-n)(t-n-3)^2}{2\gamma t(t-2)(t-n-1)}, \tag{203}$$

$$\kappa_{\mathcal{M}_I,6}(t;n,\gamma) = -\frac{(t-n-3)(t-n-4)}{\gamma t(t-2)(t-n-2)}. \tag{204}$$

Functions for the QS_I rule:

$$\kappa_{QS_I,1}(t;n,\gamma) = \frac{(t-n-1)(t-n-4)}{2\gamma(t-2)(t-n-2)}, \tag{205}$$

$$\kappa_{QS_I,2}(t;n,\gamma) = \frac{(t-n-4)(t-n-5)}{2\gamma(t-2)(t-n-2)(t-n-3)}, \tag{206}$$

$$\kappa_{QS_I,3}(t;n,\gamma) = -\frac{(t-4)(t-n-1)(t-n-4)}{2\gamma t(t-2)(t-n-2)(t-n-3)}, \tag{207}$$

$$\kappa_{QS_I,4}(t;n,\gamma) = \frac{(t-n)(t-n-5)(t-n-7)}{\gamma t(t-2)(t-n-1)}, \tag{208}$$

$$\kappa_{QS_I,5}(t;n,\gamma) = -\frac{(t-n)(t-n-5)^2(t-n-7)^2}{2\gamma t^3(t-2)(t-n-1)}, \tag{209}$$

$$\kappa_{QS_I,6}(t;n,\gamma) = -\frac{(t-n-4)(t-n-5)(t-n-7)}{\gamma t^2(t-2)(t-n-2)}. \tag{210}$$

Finally, the coefficients for QSa_I are identical to the ones of QS_I.

Appendix F. Proof of Proposition 4

First, we will prove that $\mathcal{V}_G^{QS_I} > \mathcal{V}_G^{Q_I}$ holds for $t > n+4$. It is straight forward to verify from (53) and the corresponding factors of Appendix E that

$$\mathcal{V}_G^{QS_I} - \mathcal{V}_G^{Q_I} = (\kappa_{QS_I,1} - \kappa_{Q_I,1}) \times \theta_G^2 + (\kappa_{QS_I,2} - \kappa_{Q_I,2}) \times \theta_H^2 + (\kappa_{QS_I,3} - \kappa_{Q_I,3}) > 0, \tag{211}$$

because

$$\kappa_{QS_I,1} - \kappa_{Q_I,1} = \frac{(t-n-1)(n+4)^2}{2\gamma(t-2)(t-n-2)(t-n-4)} > 0, \tag{212}$$

$$\kappa_{QS_I,2} - \kappa_{Q_I,2} = \frac{((n+8)t-(n+4)(n+5))(n+4)}{2\gamma(t-2)(t-n-2)(t-n-3)(t-n-4)} > 0, \tag{213}$$

$$\kappa_{QS_I,3} - \kappa_{Q_I,3} = \frac{(t-4)(t-n-1)(n+4)(t+(t-n-4))}{2\gamma t(t-2)(t-n-2)(t-n-3)(t-n-4)} > 0. \tag{214}$$

Second, we will show that $\mathcal{V}_{G\wedge H}^{QS_I} - \mathcal{V}_{G\wedge H}^{Q_I} > 0$ for $t > n+7$. By (55) and the corresponding factors of Appendix E we have

$$\mathcal{V}_{G\wedge H}^{QS_I} - \mathcal{V}_{G\wedge H}^{Q_I} = (\kappa_{QS_I,6} - \kappa_{Q_I,6}) \times \mathbb{E}[f_1(Y_2)Y_2] > 0, \tag{215}$$

because $\mathbb{E}[f_1(Y_2)Y_2] > 0$ and

$$\kappa_{QS_I,6} - \kappa_{Q_I,6} = \frac{(t-n-3)}{\gamma(t-2)(t-n-2)}\left[1 - \frac{(t-n-4)(t-n-5)(t-n-7)}{t^2(t-n-3)}\right] > 0. \tag{216}$$

Third, we need to determine conditions such that $\mathcal{V}_H^{QS_I} \geq \mathcal{V}_H^{Q_I}$ is satisfied. Expression (54) and the corresponding factors of Appendix E make possible to express $\mathcal{V}_H^{QS_I} - \mathcal{V}_H^{Q_I}$ as

$$\begin{aligned}\mathcal{V}_H^{QS_I} - \mathcal{V}_H^{Q_I} &= (\kappa_{QS_I,4} - \kappa_{Q_I,4}) \times \theta_H^2 \mathbb{E}[f_1(Y_1)] + (\kappa_{QS_I,5} - \kappa_{Q_I,5}) \times \mathbb{E}[f_1(Y_2)^2 Y_2] \\ &= \frac{(t-n)}{2\gamma(t-2)(t-n-1)}\left[(t-n-3) - \frac{(t-n-5)(t-n-7)}{t}\right] \times \xi_B,\end{aligned} \tag{217}$$

where

$$\xi_B = \left(\frac{t-n-3}{t} + \frac{(t-n-5)(t-n-7)}{t^2}\right) \times \mathbb{E}[f_1(Y_2)^2 Y_2] - 2\theta_H^2 \mathbb{E}[f_1(Y_1)]. \tag{218}$$

Clearly, (217) implies that $\mathcal{V}_H^{QS_I} \geq \mathcal{V}_H^{Q_I}$ if and only if $\xi_B \geq 0$. Moreover function $f_1(q)$ is concave and $g(q) = f_1(q)^2 \times q$ is convex, then by Jensen's inequality $\mathbb{E}[f_1(Y_1)] \leq f_1(\mathbb{E}[Y_1])$ and $\mathbb{E}[g(Y_2)] \geq g(\mathbb{E}[Y_2])$ hold. Under the distributional assumptions of Y_1 and Y_2, we have

$$\mathbb{E}[Y_1] = \frac{t\theta_H^2 + n + 1}{t-n-3}, \quad \text{and} \quad \mathbb{E}[Y_2] = \frac{t\theta_H^2 + n - 1}{t-n-3}. \tag{219}$$

A direct implication of the expectations in (219) is that

$$f_1(\mathbb{E}[Y_1]) = \frac{\theta_H^2 + \frac{n+1}{t}}{\theta_H^2 + \frac{n+1}{t} + \frac{n-1}{t} \times \frac{t-n-3}{t}}, \tag{220}$$

$$f_1(\mathbb{E}[Y_2])^2 \mathbb{E}[Y_2] = \left(\frac{\theta_H^2 + \frac{n-1}{t}}{\theta_H^2 + \frac{n-1}{t} + \frac{n-1}{t} \times \frac{t-n-3}{t}}\right)^2 \times \frac{t}{t-n-3} \times \left(\theta_H^2 + \frac{n-1}{t}\right). \tag{221}$$

Using the previous results and β in (57), the following condition implies that $\xi_B \geq 0$:

$$\frac{t-n-3}{t} \times \frac{1+\beta}{2} \times f_1(\mathbb{E}[Y_2])^2 \mathbb{E}[Y_2] \geq \theta_H^2 f_1(\mathbb{E}[Y_1]). \tag{222}$$

The function $r(z) = \frac{x+z}{x+z+kz}$, with $k > 0$ and $x > 0$ is strictly decreasing in z. This previous fact and some algebraic manipulations generate the following inequality which implies the one in (222):

$$\frac{1+\beta}{2} \times \frac{\left(\theta_H^2 + \frac{n-1}{t}\right)^2}{\theta_H^2 + \frac{n+1}{t} + \frac{n+1}{t} \times \frac{t-n-3}{t}} - \theta_H^2 \geq 0. \tag{223}$$

Using the definitions in (57) and considering $h = \theta_H^2$, the left hand side of (223) can be expressed as a function $F_1(h)$ given by

$$F_1(h) = \frac{1+\beta}{2} \times \frac{(h+a_1)^2}{h+b_1+b_1c_1} - h. \tag{224}$$

Moreover, $F_1(0) > 0$, F_1 is strictly decreasing in h and exists a unique $h_1^* > 0$ such that $F_1(h_1^*) = 0$. This positive root is $h_1^* = \mathrm{UB}_{H,1}$. Finally, condition (223) is equivalent to $\theta_H^2 \leq \mathrm{UB}_{H,1}$ and the proof is complete.

Appendix G. Proof of Proposition 5

First, we will show that $\mathcal{V}_{G\wedge H}^{QSa_I} - \mathcal{V}_{G\wedge H}^{\mathcal{KZ}_I} > 0$ for $t > n+7$. By (55) and the corresponding factors of Appendix E we have

$$\mathcal{V}_{G\wedge H}^{QSa_I} - \mathcal{V}_{G\wedge H}^{\mathcal{KZ}_I} = (\kappa_{QSa_I,6} - \kappa_{\mathcal{KZ}_I,6}) \times \mathbb{E}[f_0(Y_2)Y_2] > 0, \tag{225}$$

because $\mathbb{E}[f_0(Y_2)Y_2] > 0$ and

$$\kappa_{QSa_I,6} - \kappa_{\mathcal{KZ}_I,6} = \frac{(t-n-4)}{\gamma t(t-2)(t-n-2)}\left[\frac{(t-n-1)(t-n-4)}{t-n} - \frac{(t-n-5)(t-n-7)}{t}\right] > 0.$$

Second, we need to determine conditions such that $\mathcal{V}_H^{QSa_I} \geq \mathcal{V}_H^{\mathcal{KZ}_I}$ is satisfied. Expression (54) and the corresponding factors of Appendix E make possible to express $\mathcal{V}_H^{QSa_I} - \mathcal{V}_H^{\mathcal{KZ}_I}$ as

$$\begin{aligned}\mathcal{V}_H^{QSa_I} - \mathcal{V}_H^{\mathcal{KZ}_I} &= (\kappa_{QSa_I,4} - \kappa_{\mathcal{KZ}_I,4}) \times \theta_H^2 \mathbb{E}[f_0(Y_1)] + (\kappa_{QSa_I,5} - \kappa_{\mathcal{KZ}_I,5}) \times \mathbb{E}[f_0(Y_2)^2 Y_2] \\ &= \frac{(t-n)}{2\gamma(t-2)(t-n-1)}\left[\frac{(t-n-1)(t-n-4)}{t-n} - \frac{(t-n-5)(t-n-7)}{t}\right] \times \xi_{B,2},\end{aligned} \tag{226}$$

where

$$\xi_{B,2} = \left(\frac{(t-n-1)(t-n-4)}{t(t-n)} + \frac{(t-n-5)(t-n-7)}{t^2}\right) \times \mathbb{E}[f_0(Y_2)^2 Y_2] - 2\theta_H^2 \mathbb{E}[f_0(Y_1)]. \tag{227}$$

Clearly, (226) implies that $\mathcal{V}_H^{QSa_I} \geq \mathcal{V}_H^{\mathcal{K}Z_I}$ if and only if $\xi_{B,2} \geq 0$. Moreover function $f_0(q)$ is concave and $g_0(q) = f_0(q)^2 \times q$ is convex, then by Jensen's inequality both $\mathbb{E}[f_0(Y_1)] \leq f_0(\mathbb{E}[Y_1])$ and $\mathbb{E}[g_0(Y_2)] \geq g_0(\mathbb{E}[Y_2])$ hold. A direct implication of the expectations in (219) is that

$$f_0(\mathbb{E}[Y_1]) = \frac{\theta_H^2 + \frac{n+1}{t}}{\theta_H^2 + \frac{n+1}{t} + \frac{n}{t} \times \frac{t-n-3}{t}}, \tag{228}$$

$$f_0(\mathbb{E}[Y_2])^2 \mathbb{E}[Y_2] = \left(\frac{\theta_H^2 + \frac{n-1}{t}}{\theta_H^2 + \frac{n-1}{t} + \frac{n}{t} \times \frac{t-n-3}{t}}\right)^2 \times \frac{t}{t-n-3} \times \left(\theta_H^2 + \frac{n-1}{t}\right). \tag{229}$$

Using the previous results and λ in (59), the following condition implies that $\xi_{B,2} \geq 0$:

$$\frac{t-n-3}{t} \times \frac{\lambda}{2} \times f_0(\mathbb{E}[Y_2])^2 \mathbb{E}[Y_2] \geq \theta_H^2 f_0(\mathbb{E}[Y_1]). \tag{230}$$

The function $r(z) = \frac{x+z}{x+z+kz}$, with $k > 0$ and $x > 0$ is strictly decreasing in z. This previous fact and some algebraic manipulations generate the following inequality which implies the one in (230):

$$\frac{\lambda}{2} \times \frac{\left(\theta_H^2 + \frac{n-1}{t}\right)^2}{\theta_H^2 + \frac{n+1}{t} + \frac{n+1}{t} \times \left(\frac{n}{n-1} \times \frac{t-n-3}{t}\right)} - \theta_H^2 \geq 0. \tag{231}$$

Using the definitions in (59) and letting $h = \theta_H^2$, the left hand side of (231) can be expressed as a function $F_2(h)$ given by

$$F_2(h) = \frac{\lambda}{2} \times \frac{(h+a_2)^2}{h+b_2+b_2c_2} - h. \tag{232}$$

Moreover, $F_2(0) > 0$, F_2 is strictly decreasing in h and exists a unique $h_2^* > 0$ such that $F_2(h_2^*) = 0$. This positive root is $h_2^* = \mathrm{UB}_{H,2}$. Finally, condition (231) is equivalent to $\theta_H^2 \leq \mathrm{UB}_{H,2}$ and the proof is complete.

Appendix H. Expected out-of-sample return and variance of portfolio rule $\hat{\omega}_{P_I}$

If function f_k is defined as in (52), $t > n+4$; and, as a shorthand notation, both $x(n,t;P_I)$ and $y(n,t;P_I)$ are replaced with x and y, then the expected out-of-sample return and variance of portfolio rule $\hat{\omega}_{P_I}$ in (51) are given by

$$\mathbb{E}[r_{t+1}^{P_I}] = \frac{x}{\gamma} \times \frac{t}{t-n-2}\left(\theta_G^2 + \frac{\theta_H^2}{t-n-1}\right) + \frac{y}{\gamma} \times \frac{t\theta_H^2}{t-n-1} \times \mathbb{E}[f_k(Y_1)], \tag{233}$$

$$\begin{aligned}
\mathrm{Var}(r_{t+1}^{P_I}) = {} & \frac{x^2}{\gamma^2} \times c_4 \times \left(\theta_G^2 + \frac{\theta_H^2 + \frac{t-4}{t}}{t-n-3}\right) + \frac{y^2}{\gamma^2} \times \frac{t(t-2)}{(t-n-1)(t-n)} \times \mathbb{E}[f_k(Y_2)^2 Y_2] \\
& + \frac{2xy}{\gamma^2} \times c_5 \times \mathbb{E}[f_k(Y_2) Y_2] \\
& + \frac{x^2}{\gamma^2} \times c_6 \times \left(\theta_G^4 + \frac{6\theta_G^2\theta_H^2}{t-n-1} + \frac{3\theta_H^4}{(t-n-3)(t-n-1)}\right) \\
& + \frac{x^2}{\gamma^2} \times c_7 \times \left(\theta_G^2 + \frac{\theta_H^2}{t-n-3}\right) + \frac{y^2}{\gamma^2} \times \frac{t\theta_H^2}{(t-n-1)(t-n)} \times \mathbb{E}[f_k(Y_2)^2 Y_2] \\
& + \frac{y^2}{\gamma^2} \times \frac{t\theta_H^2}{(t-n-3)(t-n)} \times \mathbb{E}[f_k(Y_3)^2] + \frac{y^2}{\gamma^2} \times \frac{(t\theta_H^2)^2}{(t-n-3)(t-n)} \times \mathbb{E}[f_k(Y_4)^2] \\
& + \frac{2xy}{\gamma^2} \times \theta_H^2 \times \left(c_8 \times \mathbb{E}[f_k(Y_2)Y_2] + c_9 \times \mathbb{E}[f_k(Y_3)] + c_{10} \times \theta_G^2 \times \mathbb{E}[f_k(Y_1)]\right) \\
& + \frac{2xy}{\gamma^2} \times c_9 \times t\theta_H^4 \times \mathbb{E}[f_k(Y_4)] \\
& - \frac{x^2}{\gamma^2} \times \frac{t^2}{(t-n-2)^2}\left(\theta_G^2 + \frac{\theta_H^2}{t-n-1}\right)^2 - \frac{y^2}{\gamma^2} \times \frac{(t\theta_H^2)^2}{(t-n-1)^2} \times \mathbb{E}[f_k(Y_1)]^2 \\
& - \frac{2xy}{\gamma^2} \times c_{10} \times \theta_H^2 \left(\theta_G^2 + \frac{\theta_H^2}{t-n-1}\right) \times \mathbb{E}[f_k(Y_1)],
\end{aligned} \tag{234}$$

where $Y_1 \sim G^{t\theta_H^2}_{n+1,t-n-1}$, $Y_2 \sim G^{t\theta_H^2}_{n-1,t-n-1}$, $Y_3 \sim G^{t\theta_H^2}_{n+1,t-n-3}$, $Y_4 \sim G^{t\theta_H^2}_{n+3,t-n-3}$, and

$$c_4 = \frac{t^2(t-2)}{(t-n-4)(t-n-2)(t-n-1)}, \tag{235}$$

$$c_5 = \frac{t(t-2)}{(t-n-2)(t-n-1)(t-n)}, \tag{236}$$

$$c_6 = \frac{t^2}{(t-n-4)(t-n-2)}, \tag{237}$$

$$c_7 = \frac{t(t-2)}{(t-n-4)(t-n-2)(t-n-1)}, \tag{238}$$

$$c_8 = \frac{t}{(t-n-2)(t-n-1)(t-n)}, \tag{239}$$

$$c_9 = \frac{t}{(t-n-3)(t-n-2)(t-n)}, \tag{240}$$

$$c_{10} = \frac{t^2}{(t-n-2)(t-n-1)}. \tag{241}$$

References

Basak, G. K., Jagannathan, R., & Ma, T. (2009). Jackknife estimator for tracking error variance of optimal portfolios. *Management Science*, *55*(6), 990–1002.

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

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*(5), 28–43.

Bodnar, T., & Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned wishart distributions. *Journal of Multivariate Analysis*, *99*, 2389–2405.

Bodnar, T., Parolya, N., & Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, *266*(1), 371–390.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, *54*, 655–671.

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

Campbell, J., Lo, A., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.

DeMiguel, V., Garlappi, L., Nogales, F., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*, 798–812.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*(5), 1915–1953.

Duchin, R., & Levy, H. (2009). Markowitz versus the Talmudic portfolio diversification strategies. *Journal of Portfolio Management*, *35*(2), 71.

Ferson, W., & Siegel, A. F. (2015). Optimal orthogonal portfolios with conditioning information. In *Handbook of financial econometrics and statistics* (pp. 977–1002). Springer.

Frahm, G. (2008). Linear statistical inference for global and local minimum variance portfolios. *Statistical Papers*, *51*, 789–812.

Frost, P., & Savarino, J. (1986). Portfolio size and estimation risk. *The Journal of Portfolio Management*, *12*(4), 60–64.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 1121–1152.

Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management*. New York: McGraw-Hill.

Harvey, C. R., Liechty, J. C., Liechty, M. W., & Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, *10*(5), 469–485.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*, 1651–1683.

Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*(371), 544–554.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, *58*(3), 259–278.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, *21*(3), 279–292.

Jorion, P. (1991). BayesIan and CAPM estimators of the means: Implications for portfolio selection. *Journal of Banking & Finance*, *15*(3), 717–727.

Kan, R., Wang, X., & Zhou, G. (2016). *On the value of portfolio optimization in the presence of estimation risk: the case with and without risk-free asset*: *Rotman School of management working paper no. 2819254*.

Kan, R., Wang, X., & Zhou, G. (2018). *Optimal portfolio choice with estimation risk: No risk-free asset case (October 27, 2018)*: *Rotman school of management working paper no. 2819254; 29th Australasian finance and banking conference 2016*, Available at SSRN: https://Ssrn.Com/Abstract=2819254.

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*, 621–656.

Kempf, A., & Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review*, *58*, 332–348.

Kirby, C., & Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, *47*(2), 437–467.

Lai, Z., Xing, H., & Chen, Z. (2011). Mean–variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, *5*(2A), 798–823.

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.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.

MacKinlay, A. C. (1995). Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, *38*(1), 3–28.

MacKinlay, A. C., & Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies*, *13*(4), 883–916.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Meucci, A. (2005). *Risk and asset allocation*. Springer Verlag.

Michaud, R. O., & Michaud, R. O. (2008). *Efficient asset management: A practical guide to stock portfolio optimization and asset allocation*. Oxford University Press.

Okhrin, Y., & Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, *134*, 235–256.

Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, *55*(1), 179–223.

Roll, R. (1980). Orthogonal portfolios. *Journal of Financial and Quantitative Analysis*, *15*(5), 1005–1023.

Styan, G. P. H. (1989). Three usefull expressions for expectations involving a wishart matrix and its inverse. *Statistical Data Analysis and Inference*, 283–296.

Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, *99*, 204–2015.

Zhou, G. (2008). On the fundamental law of active portfolio management: What happens if our estimates are wrong? *Journal of Portfolio Management*, *34*(4), 26.