Portfolio Selection with Estimation Risk: a Test Based Approach*

Bertille Antoine†

Simon Fraser University

*I would like to thank Adlai Fisher, Eric Renault and Bas Werker for valuable comments and discussions. I also thank seminar and conference participants at ESEM, CEA, University of North Carolina, University of Victoria, University of Waterloo and Texas A&M University. The remaining errors are mine. Financial support from the Social Sciences and Humanities Research Council (SSHRC) is gratefully acknowledged.

†*Address: Department of Economics, 8888 University Drive, Burnaby, BC V5A1S6, Canada. Email: bertille_antoine@sfu.ca. Phone: (1)-778-782-4514*

Abstract

An important challenge of portfolio allocation arises when the (true) characteristics of returns' distribution are replaced by sample estimates. Such substitutions introduce estimation risk, which adds to traditional financial risk. I develop a new framework to provide a feasible optimal investment rule that accounts for estimation risk. In borrowing from practitioners, I evaluate funds' allocations through their probability of defeating a chosen benchmark. More precisely, the P-value investment rule maximizes the p-value of a one-sided test, ensuring that the portfolio performance is above the given threshold. When the portfolio performance is measured by the Markowitz' mean-variance criterion and when the estimation risk of the variance is ignored, the optimal investment rule is known in closed-form. The P-value investment rule is a two-fund rule when the benchmark is fixed and a three-fund rule when the benchmark is estimated. In addition, ten investment strategies are compared on simulated and empirical data.

Key words: Portfolio theory; Estimation risk; Benchmark performance; MV efficiency; Test.

JEL Classification: C4, D8, G0.

The optimal portfolio is the *best* allocation of funds across available assets, according to a well-chosen performance measure. Markowitz (1959) offers the classic definition of portfolio efficiency: a portfolio is efficient if it has the largest expected return for a given risk measured by the variance. This mean-variance efficiency provides a single-period framework[1] that still remains among the most important benchmark models used by practitioners (Michaud (1998); Meucci (2005)). In practice, however, Markowitz' optimal investment rule depends on unknown parameters, the mean and the variance of returns' distribution. To get a feasible version of this optimal rule, the unknown parameters are simply replaced by sample estimates. Such substitutions, also known as plug-in methods, give rise to several issues. First, the estimation risk is overlooked: in practice, samples are finite; consequently, estimates are different from their respective true (unknown) values. This new source of risk even appears in well-specified parametric models and adds to traditional financial risk.[2] Second, by performing these two steps, is this feasible rule optimal? Markowitz' approach can only be motivated when one believes that the estimated rule is *not too far* from the true optimal one.

In response to such limitations, one must develop an alternate framework to provide a feasible optimal investment rule that accounts for estimation risk. My framework relies on a somewhat more conservative definition of optimality. In borrowing from practitioners, I evaluate funds' allocations through their probability of defeating a chosen benchmark. Several industries are actually interested in such a goal. For instance, institutional money managers, defined benefits pension plans and endowment plans, among others, are devoted to guarantee a (chosen) minimal performance. For the chosen benchmark, the associated optimal investment rule naturally incorporates the estimation risk. In addition, it is directly applicable without requiring any additional (suboptimal) substitution step. More precisely, the P-value investment rule maximizes the probability-value (hereafter p-value) of a one-sided test, ensuring that the portfolio performance is above the given threshold. The P-value investment rule is *optimal* because it is associated with the highest probability of defeating the chosen benchmark. It also offers two main advantages. First, testing is an appropriate statistical tool for making decisions in random environments, while directly accounting for the uncertainty of such environments. Testing is actually crucial to get a feasible (true) optimal investment rule that does not require any additional substitution step. Second, maximizing the p-value increases the likelihood of defeating the chosen benchmark. Clearly, defeating the target (set by her boss) can be seen as the main challenge faced by the investor who wants to keep her job.

The P-value investment rule is quite general and flexible, allowing for any performance measure and any reference benchmark. The main contribution of this paper is to derive a closed-form formula for the P-value investment rule under two simplifying assumptions: 1) the performance is measured by the Markowitz' mean-variance criterion; and 2) the estimation risk of the variance is ignored.[3] In fact, this optimization problem amounts to maximizing an information ratio[4], which is a well-known performance measure. With a fixed benchmark, the P-value investment rule belongs to the class of two-fund investment rules. Two-fund

rules invest in the sample tangency portfolio and in the riskless asset. Only the share of wealth invested in the risky assets (vs in the riskless asset) varies among two-fund rules, and not the repartition of wealth between risky assets, which is controlled by the sample tangency portfolio. The (feasible) Markowitz' optimal mean-variance rule is the two-fund rule where the share of wealth invested in the (sample) tangency portfolio is controlled by the risk aversion parameter. The optimal P-value investment rule can be reinterpreted as a (feasible) mean-variance optimal rule associated with a *corrected* risk aversion parameter.[5] Its corrected risk aversion parameter is sample-dependent. It tends to be higher in profitable financial environments where the portfolio performance is expected to be high, and lower in less favorable financial environments. This result can be linked to the conventional wisdom suggesting that (prior to the calendar year-end) high performing investors moderate risk to "lock in" their leads, while poor performing ones gamble to "catch up" (see the empirical studies of Brown, Harlow and Starks (1996) and Goriaev, Palomino and Prat (2003)).
With an estimated (or random) benchmark, the P-value investment rule is a three-fund rule investing in the riskless asset, in the sample tangency portfolio, and in a portfolio that accounts for the correlations between the benchmark portfolio and the excess returns. The relative weights invested in the three funds depend on the correlation between these two portfolios and on their risk-return trade-offs.

The issue of estimation risk in portfolio allocation is not new.[6] One of the earliest solutions proposed in the literature is Bayesian. Since the parameters are treated as random variables, it provides a general framework where estimation risk is naturally accounted for. The study by Bawa, Brown and Klein (1979) surveys the early literature. Many others followed, including Jorion (1986), Black and Litterman (1992), Pastor and Stambaugh (2000). Because the optimal investment rule depends on unknown parameters, interest has grown in developing more feasible procedures that focus directly on the expected financial loss. These more recent approaches are very appealing as the emphasis is set on the financial cost of implementing infeasible optimal investment rules. However, in order to tackle the associated optimization problem, simplifying assumptions are required. For instance, ter Horst, de

Roon, and Werker (2006) and Kan and Zhou (2007) restrict their attentions to the class of two-fund investment rules. While ter Horst *et al.* (2006) ignore the estimation risk of the variance, Kan and Zhou (2007) (under the normality assumption of the returns) provide a closed-form optimal investment rule. However, both rules depend on nuisance parameters. So, in order to implement them, an additional suboptimal plug-in step is required.[7] More generally, when one maximizes some expected quantity, the associated optimal rule always depends on some of the (unknown) characteristics of the underlying distribution of the returns. Instead of maximizing the expected performance, I maximize the likelihood of defeating the target.[8]

Previous studies have also focused on defeating a benchmark (see Stutzer (2003) and references therein). However, to my knowledge, it has not been related to estimation risk. Moreover, these studies consider a continuous time framework instead.

Finally, ten investment strategies are compared on simulated and empirical data with respect to their out-of-sample performances, as measured by the Sharpe ratio and the certainty equivalent. Their stabilities over time, as measured by the turnover, are compared as well. Overall, the P-value investment rule performs very well, especially for smaller sample sizes. More specifically, both simulated and empirical data seem to favor P-value rules associated with smaller benchmarks. These strategies perform consistently well and are both stable and affordable.

The rest of the paper is structured as follows. Section 2 defines the framework and introduces the P-value investment rule with fixed benchmark. In section 3, competing two-fund investment rules are discussed and I compare them according to their corrected risk-aversion parameters. Section 4 presents the results of the comparative study conducted both on simulated and empirical data. In section 5, several extensions, including the introduction of random benchmarks and more elaborate performance measures, are examined. Section 6 concludes. Proofs, tables and graphs are provided in the Appendix.

1 P-value investment rule

I will first introduce the framework and then address the classical mean-variance problem. Next, I will define the P-value investment rule for a given fixed benchmark and discuss its choice.

1.1 Framework and classical Mean-Variance problem

Consider an investor who chooses a portfolio among N financial risky assets and the riskless asset. At time t, R_t denotes the vector of rates of excess returns on the N risky assets with respect to the riskless asset R_{ft}. The portfolio is built after investing the vector of weights θ into the risky assets and $(1 - \theta'\iota)$ in the riskless asset where ι is the conformable vector of ones. Each vector θ defines a different investment rule and the associated portfolio excess return, at time t, is $r_t^P(\theta) \equiv \theta' R_t$. Markowitz' optimal investment rule maximizes the following mean-variance objective function

$$\max_{\theta \in \mathbb{R}^N} \left\{ \mathrm{E}\left[r_t^P(\theta)\right] - \frac{\eta}{2} \mathrm{Var}\left[r_t^P(\theta)\right] \right\} ,$$

where η is the coefficient of relative risk aversion. Assuming stationarity, the optimal vector of weights and associated maximal performance are respectively

$$\theta_{MV}^0 = \frac{1}{\eta} \Sigma_0^{-1} \mu_0 \quad \text{and} \quad Q_{MV}^0 = \frac{1}{2\eta} \mu_0' \Sigma_0^{-1} \mu_0 \quad \text{where } \mu_0 \equiv \mathrm{E}(R_t) \text{ and } \Sigma_0 \equiv \mathrm{Var}(R_t) . \tag{1}$$

In practice, parameters μ_0 and Σ_0 are unknown: the optimal mean-variance investment rule θ_{MV}^0 is therefore *infeasible*. Markowitz (1959) provides a convenient *feasible* version of the above optimal rule by simply replacing the unknown parameters by sample estimates. Consider an investor who has observed returns over T periods and chooses a portfolio for period $(T+1)$. For the sample estimates $\hat{\mu}$ and $\hat{\Sigma}$ of the (unknown) parameters μ_0 and Σ_0, the feasible (random) investment rule and its associated (random) performance are respectively

$$\theta_{MV} = \frac{1}{\eta} \hat{\Sigma}^{-1} \hat{\mu} \quad \text{and} \quad Q_{MV} = \frac{1}{2\eta} \hat{\mu}' \hat{\Sigma}^{-1} \hat{\mu} , \tag{2}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are, for instance, the maximum likelihood estimators,

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t \quad \text{and} \quad \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})' \,. \tag{3}$$

Applying this *plug-in* method comes at a price. First, estimation risk is overlooked. In practice, the sample size T is always finite; consequently, $\hat{\mu}$ and $\hat{\Sigma}$ are different from their respective true values. Second, because the feasible rule θ_{MV} is numerically different from the true optimal one, its optimality cannot be guaranteed.

Since the performance Q_{MV} given in (2) is a random variable, it is natural to consider its expected value, which represents the out-of-sample performance an investor can achieve on average under parameter uncertainty. Under the assumption that the excess returns $\{R_t\}_t$ are identically distributed over time according to a multivariate normal distribution with mean μ_0 and covariance matrix Σ_0, Kan and Zhou (2007) show that the expected performance is always smaller than the true optimum Q^0_{MV} given in (1) (see the details in their section IIB). This also means that the investor who uses the above plug-in investment rule θ_{MV} expects to perform less well, on average, than if she were using the true rule θ^0_{MV}.

In response to the above issues, the P-value investment rule is introduced next; it naturally incorporates estimation risk and does not require any plug-in (suboptimal) step.

1.2 P-value investment rule

The P-value investment rule maximizes the p-value of a one-sided test, ensuring that the portfolio performance is above the given threshold. As pointed out earlier, Markowitz' mean-variance efficiency is a convenient framework used by practitioners. Accordingly, the following mean-variance measure of portfolio performance[9] is considered

$$Q(\mu_P, \sigma_P^2) = \mu_P - \frac{\eta}{2}\sigma_P^2 \,, \tag{4}$$

where (μ_P, σ_P^2) are respectively the first two moments of the probability distribution of the portfolio excess returns. Formally, the null hypothesis is stated as

$$H_0: \; Q(\mu_P, \sigma_P^2) > c \,, \tag{5}$$

where c is a given target. c will be considered, for the time being, as deterministic or predetermined at the time of portfolio formation.[10] To construct the associated test statistic, additional assumptions are needed on the probability distribution of the returns.

Assumption 1. *The vectors R_t of the N financial excess returns at time t, for t=1 to T, are stationary and weakly dependent such that a Central Limit Theorem applies. More formally,*
(i) $R_t \sim \mathcal{F}(\mu_0, \Sigma_0)$ for any $t = 1, \cdots, T$ where $\mathcal{F}$ is some smooth distribution function whose first two moments exist.
(ii) $\left[\sum_{t=1}^{T} R_t\right] / \sqrt{T}$ is asymptotically normally distributed with mean μ_0 and variance S_0.

The portfolio excess return at time t is $r_t^P(\theta) \equiv \theta' R_t$. The measure of portfolio performance and its estimator are respectively

$$Q_P(\theta) \;=\; \mathrm{E} r_t^P(\theta) - \frac{\eta}{2}\mathrm{Var}(r_t^P(\theta)) \quad \text{and} \quad \hat{Q}_P(\theta) = \theta'\hat{\mu} - \frac{\eta}{2}\theta'\hat{\Sigma}\theta\,, \tag{6}$$

with $\hat{\mu}$ and $\hat{\Sigma}$ given in (3).[11] The application of the vectorial central limit theorem provides the asymptotic distribution of the estimated performance: $\sqrt{T}\left[\hat{Q}_P(\theta) - Q_P(\theta)\right]$ is asymptotically normally distributed with mean 0 and variance s_Q^2. Then, for an estimator $\hat{s}_Q$ of its standard deviation s_Q, the test statistic and associated p-value are defined as follows[12]

$$\kappa(\theta) = \frac{\hat{Q}_P(\theta) - c}{\hat{s}_Q/\sqrt{T}} \quad \text{and} \quad \text{P-value}(\theta) = \int_{-\infty}^{\kappa(\theta)} f_{\mathcal{T}}(u)du\,,$$

with $f_{\mathcal{T}}$ the density function of a student random variable with $(T-1)$ degrees of freedom. Since $f_{\mathcal{T}}$ does not depend on θ, the maximization problem is conveniently rewritten as

$$\max_{\theta \in \mathbb{R}^N} [\text{P-value}(\theta)] \iff \max_{\theta \in \mathbb{R}^N} [\kappa(\theta)]\ . \tag{7}$$

Obviously, estimation risk is related to the estimation of both the mean and the variance of the portfolio. It is commonly accepted that the estimation error on the sample mean is much larger than on the sample variance[13] (for a fixed time span). Ignoring the estimation risk of the variance allows us to obtain a closed-form investment rule with a meaningful interpretation. The simplified maximization problem is now

$$\theta_p(c) = \arg\max_{\theta \in \mathbb{R}^N} \left[\frac{\theta'\hat{\mu} - \eta/2\theta'\hat{\Sigma}\theta - c}{(\theta'\hat{\Sigma}\theta)^{1/2}/\sqrt{T}}\right]. \tag{8}$$

Proposition 1.1. *Let $\hat{\mu}$ and $\hat{\Sigma}$ respectively be consistent estimators of the first two moments of the distribution of the excess returns as in (3). Under Assumption 1, for a given (deterministic) target c, the optimal P-value investment rule is defined as*

$$\theta_p(c) = \sqrt{\frac{2\eta c}{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}}}\frac{1}{\eta}\hat{\Sigma}^{-1}\hat{\mu}\,. \tag{9}$$

Several comments are worth mentioning.

First, fund managers (referred to as investors in this paper) usually have a benchmark to defeat, which is set by the financial institution they represent. In the literature (see Starks (1987), Brown, Harlow and Starks (1996) and references therein), the relationship between a fund manager and her institution has been modeled as a principal-agent relationship where the investor's (agent) main concern is to defeat a benchmark, as in securing one's job for example. In such cases, the one-sided test (5) makes sense because it focuses on defeating a predetermined target. In addition, my investment rule maximizes its p-value, or, in other words, the likelihood of the relevant goal of this investor. Maximizing this p-value actually amounts to maximizing an information ratio with a predetermined benchmark.[14] As a result, the P-value investment rule is directly rooted to the main concern of the investor, while naturally accounting for estimation risk.

Second, when the variance estimation risk is ignored, the P-value investment rule (9) is known in closed-form. It is random because it depends on the (chosen) estimates of the mean and variance of the excess returns. However, this rule is the true solution and genuinely solves the optimization problem (8). In other words, it does not come from an additional (suboptimal) plug-in step set to provide a feasible rule. Such exactness is a direct consequence of the design of the above selection method, which only makes sense because the performance measure (4) is random (due to the estimation risk). Without uncertainty, there would be no reason to run a test and therefore no p-value maximization.

Third, both the investment rule (9) and the (feasible) mean-variance rule (2) are two-fund rules that yield to the same repartition of wealth among the different financial risky assets. Only the shares of wealth invested in risky assets relative to the riskless asset differ. These shares will be discussed further in the next section when competing rules and *corrected risk* *aversion parameters* are introduced. Note that the share of the P-value rule depends on the predetermined benchmark c, which represents the minimal level of (portfolio) performance the investor wishes to guarantee with the highest possible level of confidence. In this sense, the benchmark is not really a choice variable, but rather a variable that reflects the degree of conservatism of the investor (or the financial institution she represents). Therefore, providing guidance in setting the benchmark is not obvious. In the Monte-Carlo study in section 3, the maximal (Markowitz) performance Q^0_{MV} is known and used to define three targets c of interest, respectively 10%, 50% and 90% of Q^0_{MV}. In the empirical study, the targets are based on historical performance. Nonetheless, it is still informative to derive the *optimal* target c^*, which is defined as the maximizer of the expected performance of the portfolio

$$c^* = \arg\max_{c \geq 0} \mathrm{E}\left[Q_P(\theta_p(c))\right] = \frac{1}{2\eta} \times \frac{\left[\mathrm{E}\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\mu_0}{\sqrt{\hat{\gamma}^2}}\right)\right]^2}{\left[\mathrm{E}\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}\hat{\mu}}{\hat{\gamma}^2}\right)\right]^2} \quad \text{where } \hat{\gamma}^2 \equiv \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}\,. \tag{10}$$

The optimal target c^* is infeasible since it depends on the unknown parameters μ_0 and Σ_0.[15] Interestingly enough, without estimation risk, it is easy to check that the associated investment rule is numerically equal to the true mean-variance rule, which is genuinely optimal in this case.

2 Competing rules

Section 2 discusses competing two-fund investment strategies and compares them according to their associated *corrected* risk aversion parameters. Some additional investment rules are also introduced.

2.1 Two-fund rules and corrected risk aversion

Every investment rule within the class of two-fund rules recommends the same repartition of wealth among the different risky financial assets. Only the (global) share of wealth invested in risky assets relative to the riskless asset differs. For example, the (feasible) mean-variance rule writes

$$\theta_{MV} = \frac{1}{\eta}\left[\hat{\Sigma}^{-1}\hat{\mu}\right] . \tag{11}$$

$\left[\hat{\Sigma}^{-1}\hat{\mu}\right]$ defines how wealth is allocated among risky assets, while η weights the share of wealth assigned to the risky assets: the greater η, the lower the (global) share to the risky assets.

I consider the following two-fund investment strategies.

(i) The mean-variance (Markowitz (1959)) investment rule selects the portfolio with the maximal mean-variance performance (see Section 1). The optimal allocation θ^0_{MV} given in (1) is infeasible and the feasible rule θ_{MV} (11) is obtained when the mean and the variance are replaced by the estimators (3). The estimation risk is completely ignored.

(ii) ter Horst, de Roon, and Werker (2006) propose a two-fund investment rule that incorporates the estimation risk of the mean only. Given the estimator of the mean (3), and under the assumption that the variance is known, the expected out-of-sample mean-variance performance is maximized within the class of two-fund rules. The associated optimal investment rule is

$$\theta^0_{HRW} = \frac{1}{\eta}\left(\frac{\gamma^2}{\gamma^2 + N/T}\right)\hat{\Sigma}^{-1}\hat{\mu} \qquad \text{with } \gamma^2 = \mu_0'\Sigma_0^{-1}\mu_0 \, .$$

This rule is infeasible and the associated feasible rule, θ_{HRW}, is obtained when γ^2 is replaced by its sample counterpart, $\hat{\gamma}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$.

(iii) Kan and Zhou (2007) extend the previous selection method and propose a two-fund investment rule that incorporates the estimation risk of both the mean and the variance. Given the estimators (3) of the mean and variance, the expected out-of-sample mean-variance performance is maximized within the class of two-fund rules. The associated optimal investment rule is

$$\theta^0_{KZ} = \frac{1}{\eta}\left(\frac{(T-N-4)(T-N-1)}{T(T-2)} \times \frac{\gamma^2}{\gamma^2 + N/T}\right)\hat{\Sigma}^{-1}\hat{\mu} \, .$$

This rule is infeasible and the associated feasible rule, θ_{KZ}, is given in (B.1) in Appendix B.

(iv) The Bayesian approach, Bawa, Brown and Klein (1979), maximizes the expected performance of the portfolio according to the predictive distribution of the market. This predictive distribution is built from a combination of historical observations and the chosen prior. Estimation risk is made explicit by considering the unknown parameters as random variables, described by the posterior distribution. Under the standard assumption of diffuse priors on both the mean and the variance of the excess returns, the associated optimal portfolio investment rule is

$$\theta_B = \frac{1}{\eta}\left(\frac{T-N-2}{T+1}\right)\hat{\Sigma}^{-1}\hat{\mu}\,.$$

Under the assumption that the excess returns $\{R_t\}_t$ are identically distributed over time according to a multivariate normal distribution with mean μ_0 and covariance matrix Σ_0, Kan and Zhou (2007) derive the following theoretical rankings:

$$MV^0 >> KZ^0 >> HRW^0\ , \quad \text{and} \quad MV^0 >> KZ^0 >> B >> MV$$

where $>>$ stands for "outperforms in terms of mean-variance performance". However, investment strategies MV^0, HRW^0, and KZ^0 are infeasible and have to be replaced by their feasible counterparts. Therefore, there is no guarantee that the above rankings will prevail in practice, even in simple simulations where the returns are normally distributed, as shown in section 3. However, there is another intuitive way to compare the various two-fund rules. Any two-fund rule θ_r can actually be written as follows

$$\theta_r = \frac{1}{\tilde{\eta}_r}\left[\hat{\Sigma}^{-1}\hat{\mu}\right] \quad \text{for } \tilde{\eta}_r \text{ some positive real number.}$$

As a result, the behavior of an investor that uses the above rule θ_r can be interpreted as a *modified* mean-variance behavior, or a mean-variance behavior associated with the *corrected risk aversion parameter* $\tilde{\eta}_r$. When $\tilde{\eta}_r > \eta$, θ_r invests a smaller share than the mean-variance rule θ_{MV} in the risky assets. In that sense, investor θ_r is more risk-averse than the mean-variance investor. The following corrected risk aversion parameters are associated with the above two-fund rules:

- ter Horst, de Roon and Werker (2006): $\tilde{\eta}_{HRW} = \eta \times \frac{\gamma^2 + N/T}{\gamma^2}$.
- Kan and Zhou (2007): $\tilde{\eta}_{KZ} = \eta \times \frac{\gamma^2 + N/T}{\gamma^2} \times \frac{T(T-2)}{(T-N-4)(T-N-1)}$.
- Bayesian: $\tilde{\eta}_B = \eta \times \frac{(T+1)}{(T-N-2)}$.
- P-value: $\tilde{\eta}_p(c) = \eta \times \sqrt{\frac{Q_{MV}}{c}}$ where Q_{MV} is given in (2).

It is easy to check that $\tilde{\eta}_{KZ} > \tilde{\eta}_{HRW} > \eta$ and $\tilde{\eta}_B > \eta$, as long as $T-N-4 > 0$. Note that this remains true for the corrected risk-aversion parameters of the associated feasible rules, θ_{HRW} and θ_{KZ}. Investors associated with θ^0_{HRW}, θ^0_{KZ}, or θ_B can always be reinterpreted as more risk-averse than the mean-variance investor. Recall now that θ_{MV} completely ignores estimation risk, while θ_{HRW} accounts for estimation risk of the mean only, and θ_{KZ} and θ_B for estimation risk of both the mean and variance. It is tempting to conclude that increasing the risk aversion parameter is a sensible way to account for estimation risk.

The corrected risk-aversion parameter for the P-value investment rule depends on the realized performance of the mean-variance investor, Q_{MV}, relative to the benchmark c: (i) if $c < Q_{MV}$ then $\tilde{\eta}_p > \eta$; (ii) if $c = Q_{MV}$ then $\tilde{\eta}_p = \eta$; (iii) if $c > Q_{MV}$ then $\tilde{\eta}_p < \eta$. To better understand these cases, consider an investor, with the moderate benchmark c, who faces a profitable financial environment (by chance) in the sense that the realized sample leads to high performance: likely $c < Q_{MV}$ and $\tilde{\eta}_p > \eta$. As a result, the part invested in the risky assets is smaller than suggested by the mean-variance rule. In this case at least, the P-value investment rule should lead to less extreme positions than the mean-variance. Reciprocally, with a high benchmark, more extreme positions tend to emerge with a high risk of bankruptcy (see also the empirical results and discussions in section 3).

2.2 Additional rules

To be thorough, I also consider the following investment rules in the simulation and empirical studies.

(v) The naive or equi-weighted portfolio rule, DeMiguel, Garlappi and Uppal (2008), does not involve any optimization or estimation and completely ignores the data and the estimation risk. It involves holding a portfolio with weights equal to $1/N$ in each of the risky assets.

(vi) Kan and Zhou (2007) also explore the class of three-fund investment rules when considering the sample tangency portfolio, the riskless asset, and the sample global mean-variance portfolio. The optimal rule θ^0_{KZ3} and the associated feasible rule θ_{MV3} are given in

(B.2) in Appendix B.

(vii) Garlappi, Uppal, and Wang (2006) consider a model that allows for multiple priors where the investor is averse to ambiguity. The standard mean-variance framework is modified by adding a preliminary minimization step. A constraint restricts the expected return to fall into a confidence interval around its estimated value and recognizes the existence of estimation risk. The associated minimization over the possible expected returns, subject to this constraint, reflects the investor's aversion to ambiguity. While this approach has a solid axiomatic foundation, its sequentiality cannot be directly linked to an optimality criterion. The associated rule is denoted θ_{GUW}.

Under the assumption that the excess returns $\{R_t\}_t$ are identically distributed over time according to a multivariate normal distribution with mean μ_0 and covariance matrix Σ_0, Kan and Zhou (2007) derive the following theoretical ranking

$$MV^0 >> KZ3^0 >> KZ^0 >> GUW \,,$$

where $>>$ stands for "outperforms in terms of mean-variance performance". Again, this ranking may not be guaranteed in practice when strategies KZ^0 and $KZ3^0$ are replaced by their feasible counterparts.

3 Comparative study

3.1 Comparison procedure

This comparative study relies on a *rolling-window* approach. Specifically, for a given dataset of T months of observed asset returns, I choose an estimation window of length T_w months. In each month t, starting from $t = T_w$, I use the data in the previous T_w months to estimate the parameters needed to implement a particular strategy k. These estimated parameters are then used to determine the portfolio weights θ_k. In turn, these weights are used to compute the return in month $(t+1)$. Such a procedure is continued by adding the return for the next period in the dataset and dropping the earliest return, until the end of the dataset is reached.

The rolling-window approach provides series of $(T - T_w)$ monthly *out-of-sample* returns for each portfolio strategy k, which are denoted $\hat{r}_{k,t}$ for $t = T_w + 1, \cdots, T$. Given such series of monthly out-of-sample returns, the portfolio strategies are compared according to:

(1) The out-of-sample Sharpe ratio (SR hereafter). For a strategy k, its out-of-sample SR, denoted $\widehat{SR}_k$, is defined as the ratio of the mean of (out-of-sample) excess returns, $\hat{\mu}_k$, and the standard deviation of the (out-of-sample) excess returns, $\hat{\sigma}_k$,

$$\widehat{SR}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k} \quad \text{with} \quad \hat{\mu}_k = \frac{1}{T - T_w} \sum_{t=T_w+1}^{T} \hat{r}_{k,t} \quad \text{and} \quad \hat{\sigma}_k^2 = \frac{1}{T - T_w} \sum_{t=T_w}^{T} \left(\hat{r}_{k,t} - \hat{\mu}_k\right) \left(\hat{r}_{k,t} - \hat{\mu}_k\right)' \, .$$

The Sharpe ratio is invariant to the relative weights on the riskless asset and the risky portfolio. This is the reason why the investment rules are also compared according to their out-of-sample certainty equivalents.

(2) The out-of-sample certainty equivalent (CE hereafter). For a strategy k, its out-of-sample CE, denoted $\widehat{CE}_k$, is defined as the risk-free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy. Following the common practice, the CE is calculated as the level of expected utility of a mean-variance investor

$$\widehat{CE}_k = \hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2 \, .$$

It can be shown that this corresponds to the CE of an investor with quadratic utility.

(3) The portfolio turnover. For a strategy k, its portfolio turnover is defined as the average sum of the absolute value of the trades across the N available assets. It provides information about the stability of a specific portfolio strategy, as it measures the transaction costs incurred to reallocate the portfolio at each period. I arbitrarily assume that the cost is the same for each risky asset.[16]

$$\text{Turnover}_k = \frac{1}{T - T_w - 1} \sum_{t=T_w+1}^{T-1} \left(|\theta_{k,t+1} - \theta_{k,t}|\right)' \iota \, ,$$

where ι is the column vector of ones of size N.

The results discussed in the following subsections and presented in the Appendix have been obtained with rolling windows of size $T_w = 60$ or 120, respectively 5 or 10 years of monthly data. The risk aversion parameter η is set equal to 1.[17] All the portfolio rules considered in the simulated and empirical studies are listed in Table 1. They have been introduced in Section 2. In addition, I also consider three P-value rules with respective targets $c_1 < c_2 < c_3$. For the simulated datasets, these benchmarks are selected with respect to the maximal mean-variance performance Q^0_{MV} given in (1): $c_1 = .1Q^0_{MV}$, $c_2 = .5Q^0_{MV}$ and $c_3 = .9Q^0_{MV}$. For the empirical datasets, the targets are chosen based on historical performance.

3.2 Monte-Carlo results

I will first describe the simulation experiments and then discuss the performance and stability of the different portfolio strategies listed in Table 1.

3.2.1 Simulated dataset

Asset returns are generated according to a distribution F that deviates slightly from the normal distribution. More precisely, F is a mixture of a joint normal distribution N and a deviation distribution D. The proportion of data that deviate from N is captured by the parameter h such that $F = (1-h)N + hD$. Two different values for the proportion h are considered: $h = 0$, no deviation, or $h = 0.05$, 5% of the data deviate from the joint normal distribution.[18] To generate data according to the joint normal distribution N, I consider a factor model[19] with 4 risky assets including one factor and one riskless asset. The excess returns of the factor $R_t^{(F)}$ follow a normal distribution and the excess returns of the remaining risky assets $R_{r,t}$ are generated according to: $R_{r,t} = \alpha + BR_t^{(F)} + e_t$ where the vector of mispricing coefficients α is set to 0. The coefficients of the matrix of factor loadings B are drawn from a uniform distribution between 0.5 and 1.5. The excess returns of the factor follow a normal distribution with mean 8% and standard deviation 16%. The error process e_t (uncorrelated with $R_t^{(F)}$) follows a multivariate normal distribution with mean 0 and diagonal covariance matrix with coefficients that are drawn from a uniform distribution between 0.15 and 0.25. These parameters are set according to the experiment run by DeMiguel and Nogales (2008). Finally, I consider three different deviations D from the joint normal distribution. Here again, I follow the set-up of DeMiguel and Nogales (2008)[20]: (i) D_1 is deterministic, equal to the expected return of the asset plus five standard deviations; (ii) D_2 is binomial, equal to the expected return of the asset plus five standard deviations with probability 0.5 and to the expected return of the asset minus five standard deviations with probability 0.5; and (iii) D_3 is normal with mean equal to the expected return of the asset plus five standard deviations and same covariance matrix as N.

Figure 1 provides the histograms of the distribution of the first risky asset for the simulated data with no deviation and 5% deviation according to D_1, D_2 and D_3. Table 2 characterizes the efficient frontier by providing the mean, variance and Sharpe ratio of the tangency portfolio for the same data.

3.2.2 Comparison of the strategies

I now discuss the performance and stability of the different portfolio rules listed in Table 1. The rules are evaluated (out-of-sample) over a sample of size 120 months. Four simulation designs are considered: no deviation with respect to normality and 5% deviation according to D_1, D_2 and D_3; two sizes for the rolling window, 60 and 120 months; the risk aversion parameter is set to 1. I successively discuss P-value rules, feasible investment rules, and the cost of feasibility.

(i) P-value rules:

Tables 3 and 4 report out-of-sample mean, standard deviation, CE and turnover for six P-value rules with respective benchmarks $c_1 = 0.1Q^0_{MV}$, $c_2 = 0.5Q^0_{MV}$, and $c_3 = 0.9Q^0_{MV}$ and their associated historical counterparts. These rules are labeled accordingly Pi and Pi-h with i $= 1, 2$ or 3. In addition, I also provide: i) numerical values for these targets; ii) the out-of-sample probability for the associated rule to defeat its target; and iii) the probability for the associated rule to be reinterpreted as more conservative than the mean-variance investor, or the probability that $\tilde{\eta} > \eta$.

- *Historical vs true benchmarks.* Historical benchmarks are calculated as percentages (respectively 10%, 50%, and 90%) of the MV-sample performance $\hat{Q}_{MV}$ based on the first available sample of size 60. Such benchmarks are held constant throughout the experiments. The discrepancies between true and historical benchmarks can be quite significant; historical benchmarks are usually twice as large. I conducted robustness checks on different samples and did not systematically observe such overestimations. In addition, when $\hat{Q}_{MV}$ is estimated over a larger sample, historical and true benchmarks are much closer to each other.

- *Smaller vs larger size of the rolling window.* Overall, when the size of the rolling window is larger, P-value investment rules are more conservative with larger probabilities $P(\tilde{\eta} > \eta)$, more stable with smaller turnovers, and more successful with larger CE and larger probabilities of defeating the associated benchmark. I interpret the unusual behavior of P3-h as resulting from a benchmark that is simply too high (see also the discussion that follows).

- *Smaller vs larger benchmarks.* My experiments show that choosing a high benchmark increases the risk profile of the associated investor. However, such riskier investment strategies are not always associated with better performance. Note that the simulation design with no deviation with respect to normality appears a bit at odds with respect to the three other ones. Since empirical data are more likely to depart from normality, I mainly discuss the simulation designs that deviate from normality. More specifically, I compare investment rules P1-h and P2-h. On one hand, P1-h always defeats its benchmark while always being reinterpreted as more risk-averse than the mean-variance investor. On the other hand, P2-h defeats its target with probabilities ranging from 0.5 (when $T_w = 60$) to 0.75 (when $T_w = 120$). In addition, P2-h appears much more risky than P1-h, as the associated probabilities $P(\tilde{\eta}_2 > \eta)$ are always equal to zero. However, P1-h always performs better in terms of CE and is always much more stable and affordable with turnovers twice as small.

To conclude, P-value investment rules with smaller benchmarks tend to perform best overall. In the remaining discussion, only P1-h is considered.

(ii) Feasible rules:

Table 5 reports out-of-sample mean, standard deviation, SR, CE and turnover for 8 feasible investment rules. To compare SR performances, I rely on the test procedure proposed by Ledoit and Wolf (2008) (hereafter LW). To test the null hypothesis that the difference between the SR of two investment strategies is zero, LW suggest a studentized time series bootstrap confidence interval, as well as an algorithm to select the block size.[21] I report the p-values associated with testing the difference between the SR of P1-h and any other strategy k.

- *P1-h vs other two-fund rules.* In terms of SR performance, P1-h always outperforms the other two-fund rules. The smallest p-values of the LW-test are associated with HRW and KZ: they range below 0.2 for KZ and below 0.25 for HRW for simulation designs D_1 and D_2. In terms of CE-performance, simulation design with no deviation from normality appears (again) at odds with respect to the three other ones. P1-h actually outperforms all the other rules for simulation designs with deviations D_1, D_2 and D_3, but performs worst for simulation designs without deviation. Finally, P1-h is much more stable and affordable than any two-fund rule with a turnover two to three times smaller.

Note also that the performances of HRW and KZ are quite close to each other. HRW tends to perform slightly better in terms of SR, while KZ tends to perform slightly better in terms of CE. Consequently, ignoring the risk of the variance appears as a reasonable simplifying assumption, at least in these experiments where the ratio N/T is kept small and equal to $1/6$. In any case, both HRW and KZ are outperformed by P1-h.

- *P1-h vs other rules.* Compared to the remaining investment rules, P1-h always provide the largest SR-performance except for the design without deviation where KZ3 performs best. However, these differences are only significant against the investment rule GUW. For the other rules, EQ and KZ3, the p-values associated with the LW-test are above 0.7. In terms of CE-performance, P1-h always outperforms its competitors, except for simulation design with no deviation.

(iii) Infeasible rules:

Table 6 reports the out-of-sample losses (in percentages) from using a specific rule instead of the (infeasible) optimal rule MV^0. These losses are measured both in terms of SR and CE performances and are provided as a function of the size of the rolling window T_w. Four infeasible rules are considered, MV, HRW, KZ and KZ3, as well as their feasible counterparts.

Such analysis documents the cost of feasibility.

Within each panel, the losses are decreasing functions of the size of the rolling window T_w, both for infeasible and feasible rules. The losses are higher for feasible rules than for infeasible ones. However, the discrepancies between the different rules within a group (infeasible or feasible) and across groups decrease as T_w increases. This means that, when T_w increases, the cost of feasibility decreases. In addition, the rules tend to perform more similarly.

For a rolling window of size 60, which was considered above, the costs of feasibility of HRW and KZ are approximately 2 to 3 % for the SR-performance and 8 to 10% for the CE-performance. Interestingly, these infeasible two-fund rules would not outperform P1-h. Note also that even if the infeasible investment rule KZ3 tend to outperform P1-h, this is usually not the case with its feasible counterpart as shown above.

3.3 Empirical results

I now consider two empirical datasets and discuss the performance and stability of the different portfolio strategies listed in Table 1. The first dataset consists of monthly excess returns over the 90-day T-Bill on ten industry portfolios in the United States, while the second one consists of 49 industry portfolios in the United States. Considering sets of 10 and 49 industry portfolios allows me to vary the ratio of the number of assets and the sample size, N/T, in order to study its implications on the estimation risk of the variance. Both datasets are available on Kenneth French's web site.

I compare the out-of-sample performances of 10 investment rules over 10 years from October 2000 to October 2010.[22] More specifically, I consider three P-value investment rules based on three historical targets, $c_1 = 0.1Q_{MV}$, $c_2 = 0.5Q_{MV}$, and $c_3 = 0.9Q_{MV}$ where the performance Q_{MV} given in equation (2) is based on observations from $t = 1$ to T_w. I also consider the four two-fund rules highlighted in section 3, namely MV, HRW, KZ, and B, as well as the additional rules, EQ, KZ3, and GUW.

Tables 7 and 8 report the out-of-sample mean, standard deviation, SR (and p-value associated with the LW-test of the difference between the SR of P1-h and the SR of any other strategy), CE and turnover for rolling window of size 60. In addition, numerical values are provided for the three historical targets, as well as the out-of-sample probabilities for the associated rule to defeat its target and to be reinterpreted as more conservative than the mean-variance investor.

(i) 10 Industry portfolios:

- *P-value rules.* Similarly to the results observed on simulated data, the P-value investment rule associated with the smallest benchmark performs best. P1-h defeats its target with probability almost 1 and is always reinterpreted as more conservative than the mean-variance investor. It has the largest CE-performance and the smallest turnover. As a result, P1-h appears as a reliable and affordable investment strategy. In the remaining discussion, this is the only P-value investment rule I discuss.

- *P-value vs other two-fund rules and other competitors.* P1-h outperforms all the other rules, both in terms of CE and SR performances. The SR performance of P1-h is significantly better than any two-fund rule at level of confidence 0.85. The only rule that is not significantly outperformed by P1-h is KZ3. However, note that P1-h clearly outperforms KZ3 in terms of CE. In addition, P1-h is one of the most stable and affordable investment strategy.

Note also that there is little difference between the performances of HRW and KZ, suggesting (again) that ignoring the estimation risk of the variance is a reasonable simplifying assumption. The ratio N/T is actually twice as large as the one considered with simulated data, respectively $10/60$ and $4/60$.

(ii) 49 Industry portfolios:

- *P-value rules.* The P-value investment rule associated with the smallest benchmark performs best as previously observed. The main difference with the previous dataset is that P1-h only defeats its target with probability 0.21. Note also that the benchmark itself is much higher than in the previous case. This is likely due to less reliable estimators, because when larger sizes of rolling window are considered, the probability of defeating the benchmark rises accordingly.

- *P-value vs other two-fund rules and other competitors.* The SR performance of P1-h is significantly better than any other rule at level of confidence above 0.90, except for KZ3 which is associated to a p-value of the LW-test equal to 0.21. However, the CE performance of P1-h is clearly outperformed by KZ (and KZ3), which is also the most stable and affordable investment strategy. When I only consider competitors that ignore the estimation risk of the variance, namely HRW, this is not the case anymore. P1-h strongly outperforms HRW in terms of CE performance, while being much more stable. Consequently, ignoring the estimation risk of the variance does not seem as appropriate as with the previous dataset, or with simulated data. The ratio N/T is actually five times larger than with the previous dataset, respectively $49/60$ and $10/60$.

To conclude, similarly to what the simulation results show, the empirical datasets favor the conservative behavior of the P-value investment rule associated with the smallest benchmark. P1-h provides a reliable and affordable investment strategy. In addition, ignoring the estimation risk of the variance appears to be a reasonable simplifying assumption, as long as the ratio N/T is kept small, typically $1/6$. This result confirms the findings of Kan and Zhou (2007).

4 Extensions

The main contribution of this paper is to derive a closed-form formula for the P-value investment rule under two simplifying assumptions: 1) the performance is measured by the Markowitz' mean-variance criterion; and 2) the estimation risk of the variance is ignored. However, the proposed methodology is much more general and the following extensions are discussed next: estimated (or random) benchmarks; more elaborate performance measures; relaxing assumption 1; comparison with the Value-at-Risk.

(i) Estimated (or random) benchmark:

So far, only deterministic benchmarks have been considered. The usual practice is rather to include a random benchmark such as a value-weighted index in order to *track* its performance.

$\tilde{c}$ denotes the estimated performance of the benchmark portfolio. It is assumed to be jointly asymptotically normally distributed with the asset returns, with mean c, variance σ_c^2 and covariance vector σ_{Rc} with the asset returns. It follows that, under assumption 1, $(\hat{Q}_P(\theta) - \tilde{c})$ is asymptotically normally distributed with the following mean and variance,

$$(Q_P(\theta) - c) \quad \text{and} \quad (\theta'\Sigma\theta - 2\theta'\sigma_{Rc} + \sigma_c^2)\,.$$

The associated P-value investment rule maximizes the p-value of the following one-sided test, $H_0 : Q_P(\theta) > \tilde{c}$. Clearly, considering a random target does not change much the setup defined previously. Therefore, the reader should refer to the relevant sections if needed. In particular, ignoring the estimation risk of the variance allows the maximization problem of the p-value to be simplified as follows,

$$\theta_p(\tilde{c}) = \arg\max_{\theta} \left[\frac{\theta'\hat{\mu} - \frac{\eta}{2}\theta'\hat{\Sigma}\theta - \tilde{c}}{\sqrt{\theta'\hat{\Sigma}\theta - 2\theta'\hat{\sigma}_{Rc} + \hat{\sigma}_c^2}} \right],$$

where $\hat{\sigma}_{Rc}$ and $\hat{\sigma}_c^2$ are consistent estimators of σ_{Rc} and σ_c^2 respectively. The associated optimal P-value investment rule is then given as

$$\theta_p(\tilde{c}) = \alpha \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\eta} + (1-\alpha)\hat{\Sigma}^{-1}\hat{\sigma}_{Rc}\,, \tag{12}$$

where α is a real number defined implicitly in the Appendix.

With an estimated benchmark, the P-value investment rule is a three-fund rule investing in the riskless asset, the sample tangency portfolio, and a portfolio that accounts for the correlations between the benchmark portfolio and the excess returns. The relative weights between these three funds depend on the correlation between these two portfolios and on their risk-return trade-offs.

To conclude, introducing an estimated benchmark is fairly straightforward and leads to a feasible and intuitive investment rule.

(ii) More elaborate performance measures:

Considering the simple mean-variance portfolio performance measure provides the closed-form intuitive P-value investment rule (9). However, more elaborate performance measures can easily be incorporated, for instance to account for the effects of higher moments like the skewness. In general, higher moments cannot be neglected unless there is a reason to believe that the asset returns are normally distributed, or that the utility function is quadratic, or that these moments are irrelevant to the investor's decision (see for instance Samuelson (1970)). Several alternative criteria for portfolio selection based upon higher order moments have been developed (see for instance Wang and Xia (2002)); so far, no procedure seems to have clearly emerged. When such elaborate performance measures are linear combinations of the relevant moments of the portfolio returns' distribution, say like the mean-variance-skewness discussed in Briec, Kerstens, and Jokung (2007), a strategy very similar to the one developed in Section 1 will provide the P-value investment rule. In this case, the associated test statistic $\kappa(\theta)$ remains asymptotically pivotal, since its asymptotic distribution does not depend on θ. As a result, the maximization of the p-value can be equivalently rewritten as the maximization of $\kappa(\theta)$ as done in (7). However, it is obviously less likely that such a maximization will always lead to a closed-form investment rule.

With even more elaborate performance measures, it is not clear whether an asymptotically pivotal statistic like $\kappa(\theta)$ exists. Suppose it does not and the asymptotic distribution of the test statistic $\kappa^*(\theta)$ now depends on the vector of weights θ, say f_θ. The p-value now writes

$$\text{P-value}(\theta) = \int_{-\infty}^{\kappa^*(\theta)} f_\theta(u)du\,.$$

The associated optimization problem has to be solved numerically and it is quite unlikely that the optimal investment rule will be known in closed-form.

To conclude, more elaborate portfolio performance measures can always be considered. However, in practice, such refinements will likely be paid in terms of tractability and interpretability of the associated investment rule.

(iii) Relaxing assumption 1:

The asymptotic distribution of the test statistic $\kappa(\theta)$ naturally follows from assumption 1, which ensures the validity of a central limit theorem for the excess returns. Accordingly, the asymptotic test (5) has been considered. If such an assumption is not credible, the procedure actually remains valid with a finite sample test based on the empirical distribution, or any resampling distribution. Although most test statistics used in econometrics are asymptotically pivotal (and usually with known asymptotic distributions), most of them are not pivotal in finite samples. As a result, the p-value associated with the finite sample test is likely to write

$$\text{P-value}(\theta) = \int_{-\infty}^{\kappa^*(\theta)} \overline{f}_{n,\theta}(u)du \, ,$$

where the empirical density function $\overline{f}_{n,\theta}$ of the portfolio performance measure $\hat{Q}(\theta)$ now (likely) depends on the vector of weights θ. Accordingly, the numerical complexity of the associated optimization problem increases quite a lot. It is now very unlikely that the optimal investment rule will be known in closed-form.

To conclude, given the tractability and interpretability of the resulting investment rule, ensuring the validity of a CLT for the returns appears to be a pretty weak assumption.[23]

(iv) Comparison with the Value-at-Risk:

The P-value investment rule maximizes the p-value of a one-sided test, ensuring that the portfolio performance is above the given threshold. Such a selection method can be linked to a well-known financial risk measure, the Value-at-Risk (VaR hereafter). Recall that, for a given level α, the VaR represents an estimate of the level of loss on a portfolio, which is expected to be equaled or exceeded with the given (small) probability α. Mathematically, the Var is defined as

$$\texttt{VaR}(\alpha) = \inf \left\{ l \in \mathbb{R} \; s.t. \; P(L_P > l) \leq 1 - \alpha \right\} \, ,$$

where L_P denotes the portfolio loss over a given period of time, typically a day or two weeks. In addition, risk regulations usually dictate the choice of the level of confidence α, typically 1% or 5%.

As a result, the VaR summarizes the distribution of possible portfolio losses by a quantile for a well-chosen level of confidence. The goal of the P-value selection method is obviously to select a portfolio (or investment rule) rather than to summarize its distribution. More importantly though, the P-value selection method sets the benchmark (or quantile in the language of the VaR) and not the level of confidence. Choosing the benchmark appears to be directly in line with institutional money managers' concerns.

5 Conclusion

In this paper, I develop an alternate framework to provide a feasible investment rule, namely the P-value investment rule, that accounts for estimation risk. The P-value investment rule maximizes the p-value of a one-sided test, ensuring that the portfolio performance is above a given threshold. When the portfolio performance is measured by the Markowitz' mean-variance criterion and when the estimation risk of the variance is ignored, the optimal investment rule is known in closed-form. More specifically, with a fixed benchmark, the P-value investment rule is a two-fund rule where the share of wealth invested in the risky assets is sample dependent. According to this strategy, less is invested in the risky assets in profitable financial environments, or when the sample performance tends to be high. The resulting investment strategy is more stable and more affordable than the mean-variance investment rule as shown in the comparative study. With an estimated (or random) benchmark, the P-value investment rule is a three-fund rule investing in the riskless asset, the sample tangency portfolio, and a portfolio that accounts for the correlations between the benchmark portfolio and the excess returns. The relative weights invested in the three funds depend on the correlation between these two portfolios and on their risk-return trade-offs.
In the comparative study, I consider both simulated and empirical data and compare ten investment strategies according to their out-of-sample performances as measured by the Sharpe ratio and the certainty equivalent; in addition, I also evaluate their stabilities over time through the turnover. Overall, the P-value investment rule performs very well, especially for smaller sample sizes. More specifically, both simulated and empirical data seem to favor P-value rules associated with smaller benchmarks. These strategies are more conservative because they invest less in the risky assets; they perform consistently well and are both stable and affordable. As a result, it seems that investors are better off choosing a lower benchmark.

Finally, ignoring the estimation risk of the variance appears to be a reasonable simplifying assumption as long as the number of risky assets relative to the sample size is kept small, typically 1/6 in the above comparative study.

A Proofs of the main results

• Proof of equation (9) $\theta_p(c)$: The first order conditions can be reinterpreted as a function of the (feasible) vector of the mean-variance weights θ_{MV} defined in equation (2) as follows:

$$
\begin{aligned}
(\hat{\mu}-\eta\hat{\Sigma}\theta_p)\sqrt{\theta_p'\hat{\Sigma}\theta_p} - \frac{\theta_p'\hat{\mu}-\eta/2\theta_p'\hat{\Sigma}\theta_p - c}{\sqrt{\theta_p'\hat{\Sigma}\theta_p}}\hat{\Sigma}\theta_p = 0 \;&\Leftrightarrow\; \hat{\mu}-\eta\hat{\Sigma}\theta_p - \frac{\theta_p'\hat{\mu}-c}{\theta_p'\hat{\Sigma}\theta_p}\hat{\Sigma}\theta_p + \frac{\eta}{2}\hat{\Sigma}\theta_p = 0 \\
&\Leftrightarrow\; \hat{\mu}-\frac{\eta}{2}\hat{\Sigma}\theta_p - \frac{\theta_p'\hat{\mu}-c}{\theta_p'\hat{\Sigma}\theta_p}\hat{\Sigma}\theta_p = 0 \qquad \text{(A.1)} \\
&\Leftrightarrow\; \hat{\Sigma}^{-1}\hat{\mu}-\frac{\eta}{2}\theta_p - \frac{\theta_p'\hat{\mu}-c}{\theta_p'\hat{\Sigma}\theta_p}\theta_p = 0 \\
&\Rightarrow\; \exists\, d\in\mathbb{R}\ s.t.\ \theta_p = d\hat{\Sigma}^{-1}\hat{\mu}\,. \qquad \text{(A.2)}
\end{aligned}
$$

Pre-multiply (A.1) by θ_p' to get: $\eta\theta_p'\hat{\Sigma}\theta_p = 2c$.

Plug (A.2) in the above equation to get: $\eta d^2\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} = 2c$.

After plugging the 2 resulting candidates in the objective function, I conclude that the positive root leads to the maximum. Hence: $\theta_p = d\hat{\Sigma}^{-1}\hat{\mu}$ with $d = \sqrt{\frac{2c}{\eta(\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu})}}$.

• Proof of equation (10) c^*:

$$
Q_P(\theta_p(c)) \;=\; \theta_p'(c)\mu_0 - \frac{\eta}{2}\theta_p'(c)\Sigma_0\theta_p(c) = \frac{\sqrt{2c}}{\sqrt{\eta\hat{\gamma}^2}}\hat{\mu}'\hat{\Sigma}^{-1}\mu_0 - \frac{c}{\hat{\gamma}^2}\hat{\mu}'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}\hat{\mu}\,,
$$

where $\hat{\gamma}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$. I now maximize it with respect to c. The associated first order conditions are

$$
\frac{1}{\sqrt{2c^*\eta}}E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\mu_0}{\sqrt{\hat{\gamma}^2}}\right) = E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}\hat{\mu}}{\hat{\gamma}^2}\right) \implies c^* = \frac{1}{2\eta}\times\frac{\left[E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\mu_0}{\sqrt{\hat{\gamma}^2}}\right)\right]^2}{\left[E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}\hat{\mu}}{\hat{\gamma}^2}\right)\right]^2}\,.
$$

The associated optimal vector of weights is the following

$$
\theta_p(c^*) = \frac{\left|E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\mu_0}{\sqrt{\hat{\gamma}^2}}\right)\right|}{\left|E\left(\frac{\hat{\mu}'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}\hat{\mu}}{\hat{\gamma}^2}\right)\right|}\frac{1}{\sqrt{\hat{\gamma}^2}}\frac{1}{\eta}\hat{\Sigma}^{-1}\hat{\mu}\,.
$$

Note in particular that if $\tilde{\mu_0}$ and Σ_0 were known, I would get $\theta_p(c^*) = \theta_{MV}$, which corresponds to the best portfolio rule in absence of estimation risk.

- Proof of equation (12):

The FOC write:

$$
\begin{aligned}
& \hat{\mu} - \eta\hat{\Sigma}\theta - \frac{\theta'\hat{\mu} - \frac{\eta}{2}\theta'\hat{\Sigma}\theta - \tilde{c}}{\theta'\hat{\Sigma}\theta - 2\theta'\hat{\sigma}_{Rc} + \hat{\sigma}_c^2}(\hat{\Sigma}\theta - \hat{\sigma}_{Rc}) = 0 \\
\Leftrightarrow \quad & \hat{\mu} - \eta\hat{\sigma}_{Rc} - \eta(\hat{\Sigma}\theta - \hat{\sigma}_{Rc}) = \frac{\theta'\hat{\mu} - \frac{\eta}{2}\theta'\hat{\Sigma}\theta - \tilde{c}}{\theta'\hat{\Sigma}\theta - 2\theta'\hat{\sigma}_{Rc} + \hat{\sigma}_c^2}(\hat{\Sigma}\theta - \hat{\sigma}_{Rc}) \qquad \text{(A.3)} \\
\Leftrightarrow \quad & \hat{\Sigma}^{-1}(\hat{\mu} - \eta\hat{\sigma}_{Rc}) = \left[\eta + \frac{\theta'\hat{\mu} - \frac{\eta}{2}\theta'\hat{\Sigma}\theta - \tilde{c}}{\theta'\hat{\Sigma}\theta - 2\theta'\hat{\sigma}_{Rc} + \hat{\sigma}_c^2}\right](\theta - \hat{\Sigma}^{-1}\hat{\sigma}_{Rc}) \\
\Rightarrow \quad & \exists\, \alpha \in \mathbb{R} \ s.t.\ \theta_p(\tilde{c}) = \alpha\frac{\hat{\Sigma}^{-1}\hat{\mu}}{\eta} + (1-\alpha)\hat{\Sigma}^{-1}\hat{\sigma}_{Rc}\,. \qquad \text{(A.4)}
\end{aligned}
$$

Pre-multiply (A.3) by θ' to get

$$
0 = \frac{\eta}{2}(\theta'\hat{\Sigma}\theta)^2 - \frac{3\eta}{2}(\theta'\hat{\sigma}_{Rc})(\theta'\hat{\Sigma}\theta) + (\theta'\hat{\mu})(\theta'\hat{\sigma}_{Rc}) + (\theta'\hat{\Sigma}\theta)[\eta\hat{\sigma}_c^2 - \tilde{c}] + [\tilde{c}\hat{\sigma}_{Rc} - \hat{\sigma}_c^2\hat{\mu}]'\theta\,.
$$

Define $\alpha = 1/(1+\delta)$ and plug (A.4) in the above equation to get

$$
\begin{aligned}
& \delta^4\left[-\eta C^2 + C(E+1) - \sigma_c^2\eta B\right] + \delta^3\left[-\frac{5\eta}{2}BC - \frac{3\eta}{2}C^2 + 2E(B+C) + C - \sigma_c^2\eta B\right] \\
+ \quad & \delta^2\left[-\eta B^2 - \frac{\eta}{2}AC - \frac{9\eta}{2}BC + E(A+C) + 4BC + 3(B+C) - 3\sigma_c^2\eta(A+B)\right] \\
+ \quad & \delta\left[\frac{\eta}{2}AB - 3\eta B^2 - \frac{3\eta}{2}AC + 2(BC+AE) + C - \sigma_c^2\eta(B+3A) + 3B\right] \\
+ \quad & \left[\frac{\eta}{2}A^2 - \frac{3\eta}{2}AB + AE + B - \sigma_c^2\eta A\right] = 0\,,
\end{aligned}
$$

where A, B, C, and E are defined as follows

$$
A = \frac{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}}{\eta^2}\,, \quad B = \frac{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\sigma}_{Rc}}{\eta}\,, \quad C = \hat{\sigma}_{Rc}'\hat{\Sigma}^{-1}\hat{\sigma}_{Rc}\,, \quad E = \eta\sigma_c^2 - \tilde{c}\,.
$$

B Results of the simulation and empirical studies

I now provide additional description of the investment rules considered in this paper.

- Two-fund rule of Kan and Zhou:

$$
\theta_{KZ}^0 = \frac{1}{\eta}\left[\left(\frac{(T-N-1)(T-N-4)}{T(T-2)}\right)\left(\frac{\gamma^2}{\gamma^2 + N/T}\right)\right]\hat{\Sigma}^{-1}\hat{\mu} \ \text{ with } \ \gamma^2 = \mu'\Sigma^{-1}\mu\,.
$$

Kan and Zhou (2007) recommend the following feasible rule θ_{KZ} where γ^2 is replaced by

$$\hat{\gamma}_a^2 = \frac{(T-N-2)\hat{\gamma}^2 - N}{T} + \frac{2(\hat{\gamma}^2)^{N/2}(1+\hat{\gamma}^2)^{-(T-2)/2}}{TB_{\hat{\gamma}^2/(1+\hat{\gamma}^2)}(N/2,(T-N)/2)}\,, \tag{B.1}$$

with $\hat{\gamma}^2 = \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$ and $B_x(a,b)$ is the incomplete beta function

$$B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}dy\,.$$

- Three-fund rule of Kan and Zhou:

$$\theta_{KZ3}^0 = \frac{c_3}{\eta}\left[\left(\frac{\psi^2}{\psi^2+N/T}\right)\hat{\Sigma}^{-1}\hat{\mu} + \left(\frac{N/T}{\psi^2+N/T}\right)\mu_g\hat{\Sigma}^{-1}\iota\right]\,, \tag{B.2}$$

with $\mu_g = \frac{\iota'\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\iota}$, $c_3 = \left(\frac{T-N-4}{T}\right)\left(\frac{T-N-1}{T-2}\right)$, $\psi^2 = (\mu-\mu_g\iota)'\,\Sigma^{-1}\,(\mu-\mu_g\iota)$.

Kan and Zhou (2007) recommend replacing μ_g and ψ^2 by

$$\hat{\mu}_g = \frac{\hat{\mu}'\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota} \text{ and } \hat{\psi}_a^2 = \frac{(T-N-1)\hat{\psi}^2-(N-1)}{T} + \frac{2(\hat{\psi}^2)^{(N-1)/2}(1+\hat{\psi}^2)^{-(T-2)/2}}{TB_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2,(T-N+1)/2)}\,.$$

- Sequential min-max of Garlappi, Uppal and Wang:

$$\theta_{GUW} = \frac{1}{\eta}d\frac{T-1}{T}\hat{\Sigma}^{-1}\hat{\mu} \text{ with } d = \begin{cases} 1-(\epsilon/\hat{\gamma}^2)^{1/2} & if \quad \hat{\gamma}^2 > \epsilon \\ 0 & if \quad \hat{\gamma}^2 \leq \epsilon \end{cases} \text{ and } \epsilon = N\mathcal{F}_{N,T-N}^{-1}(p)/(T-N)\,, \tag{B.3}$$

where $\mathcal{F}_{N,T-N}^{-1}$ is the inverse of the cumulative distribution function of a central F-distribution with $(N, T-N)$ degrees of freedom and p is a probability. I use $p = .99$.

References

Bawa, V., S. Brown, and R. Klein (1979): *Estimation Risk and Optimal Portfolio Choice*, Amsterdam: North Holland.

Black, F. and R. Litterman (1992): "Global Portfolio Optimization," *Financial Analysis Journal*, 48, 28–43.

Brandt, M. (2004): "Portfolio Choice Problems," *in Handbook of Financial Econometrics*, y. Aït-Sahalia and L.P. Hansen, eds.

Briec, W., K. Kerstens, and O. Jokung (2007): "Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach," *Management Science*, 53, 135–149.

Cavadini, F., A. Sbuelz, and F. Trojani (2002): "A simplified way of incorporating Model Risk, Estimation Risk and Robustness in Mean Variance Portfolio Management," *Technical Report, University of Southern Switzerland.*

Das, S. and R. Uppal (2004): "Systeme risk and international portfolio choice," *Journal of Finance*, 59, 2809–2834.

DeMiguel, V. and F. Nogales (2009): "Portfolio Selection with Robust Estimation," *Operations Research*, 57, 560–577.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *The Review of Financial Studies*, 20, 41–81.

Goriaev, A., F. Palomino, and A. Prat (2003): "Mutual fund tournament: Risk taking incentives induced by ranking objectives," *Working paper, New Economic School.*

Harvey, C., J. Liechty, M. Liechty, and P. Müller (2010): "Portfolio Selection with Higher Moments," *Quantitative Finance*, forthcoming.

Horst, J., F. de Roon, and B. Werker (2006): "An Alternative Approach to Estimation Risk," *in Advances in Corporate Finance and Asset Pricing, Elsevier Amsterdam*, luc Renneboog, eds.

Jorion, P. (1996): *Value-at-Risk: the new Benchmark for controlling Market Risk*, Chicago, Irwin: McGraw-Hill, second ed.

Kan, R. and G. Zhou (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *Journal of Financial and Quantitative Analysis*, 42, 621–656.

MacKinlay, A. and L. Pastor (2000): "Asset pricing models: implications for expected returns and portfolio selection," *Review of Financial Studies*, 13, 883–916.

Markowitz, H. (1959): *Portfolio Selection: Efficient Diversification of Investments*, New-York: Wiley.

Merton, R. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Meucci, A. (2005): *Risk and Asset Allocation*, New-York: Springer-Verlag.

Michaud, R. (1998): *Efficient Asset Management: a Practical Guide to Stock Optimization and Asset Allocation*, Boston, Massachusetts: Harvard Business School Press.

Pastor, L. and R. Stambaugh (2000): "Comparing Asset Pricing Models: an Investment Perspective," *Journal of Financial Economics*, 56, 335–381.

Samuelson, P. (1970): "The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments," *Review of Economic Studies*, 37, 537–542.

Stutzer, M. (2003): "Portfolio Choice with Endogenous Utility: a Large Deviations Approach," *Journal of Econometrics*, 116, 365–386.

Wang, S. and Y. Xia (2002): *Portfolio selection and asset pricing*, Berlin: Springer.

Zellner, A. and V. Chetty (1065): "Prediction and Decision Problems in Regression Models from the Bayesian Point of View," *Journal of the American Statistical Association*, 60, 608–616.

Notes

[1]It has been extended since to a multi-periods framework: see e.g. Campbell and Viceira (2002).

[2]Kan and Zhou (2007) provide an extensive study of the financial consequences of ignoring estimation risk.

[3]When the number of assets is moderate compared to the number of observations (and the time-span is fixed) mean asset returns are harder to estimate: see Merton (1980) and Kan and Zhou (2007).

[4]I thank the Editor, René Garcia, for pointing this out.

[5]The concept of *corrected risk aversion* has been introduced by ter Horst, de Roon, and Werker (2006).

[6]Brandt (2004) provides a broad survey on general issues related to portfolio choice.

[7]Of course, by construction, the (infeasible) optimal mean-variance investment rule outperforms any two-fund rule, especially Kan and Zhou's; by construction, the latter also outperforms any P-value investment rule. However nothing is guaranteed when one considers feasible versions of the optimal mean-variance and Kan and Zhou's rules as shown in the comparative study in section 3.

[8]Others have also departed from the classical mean-variance framework: Garlappi, Uppal and Wang (2007) propose a sequential max-min method where the worst performance (when the unknown parameters fall into some confidence interval) is maximized with respect to the portfolio weights; Harvey, Liechty, Liechty and Muller (2004) adopt a Bayesian setting under the assumption that the returns follow a skew-normal distribution.

[9]More elaborate measures of portfolio performance are discussed in section 4.

[10]Random or estimated targets are discussed in section 4.

[11]The procedure remains similar for any other set of consistent estimates. Note that if $\{(R_t - \mu_0)\}_t$ is not a martingale difference sequence, the estimator $\hat{\Sigma}$ has to take into account the serial correlation, like the HAC (Heteroskedastic and Autocorrelation Consistent) estimator.

[12]For convenience, I consider here the asymptotic test. See the discussion in section 4 about alternative testing procedures.

[13]According to Kan and Zhou's (2007) study, the above claim is only acceptable when the ratio of the number of assets and the sample size (that is N/T) is *small*: their study reports the presence of an interactive effect between both estimation errors. See also the simulation and empirical studies in section 3.

[14]The information ratio (IR) is a well-known risk-adjusted measure of performance. The usual practice is to include a random benchmark such as a value-weighted index: see the discussion in section 4. I thank the Editor, René Garcia, for pointing this out.

[15]This is not really surprising since one maximizes the expected performance with respect to c.

[16]This (simple) definition of the turnover implies by construction a value of zero for the naive equi-weighted

portfolio.

[17]Other values were also considered for robustness.

[18]Das and Uppal (2004) calibrate a jump diffusion process to historical returns on the indexes for six countries. They found that on average a jump occurs every 20 months. This corresponds to 5% of the data that deviate from the joint normal distribution.

[19]The factor model is similar to the one used in MacKinlay and Pastor (2000).

[20]They also consider another deterministic distribution, which is really close to D_1.

[21]Computer programming code (in R) is freely available at http://www.iew.uzh.ch/chairs/wolf/team/wolf/publications_en.html

[22]Other periods have been considered for robustness checks and the results were quite similar.

[23]More explicit assumptions would involve some mixing assumptions, or even weaker assumptions like near-epoch dependency. This discussion is beyond the scope of this paper.

Infeasible rules (only for simulated data)	
MV^0	(infeasible) Mean-variance
KZ^0	Kan and Zhou (infeasible) 2-fund
HRW^0	ter Horst, de Roon and Werker (infeasible) 2-fund
$KZ3^0$	Kan and Zhou (infeasible) 3-fund
Feasible rules (both for simulated and empirical data)	
P1	P-value with target 10% Optimum (with simulated datasets)
P1-h	P-value with target 10% estimated MV-CE (with empirical datasets)
P2	P-value with target 50% Optimum (with simulated datasets)
P2-h	P-value with target 50% estimated MV-CE (with empirical datasets)
P3	P-value with target 90% Optimum (with simulated datasets)
P3-h	P-value with target 90% estimated MV-CE (with empirical datasets)
MV	Feasible counterpart of MV^0
HRW	Feasible counterpart of HRW^0
KZ	Feasible counterpart of KZ^0
B	Bayesian with diffuse priors
EQ	Equi-weighted portfolio
KZ3	Feasible counterpart of $KZ3^0$
GUW	Garlappi, Uppal and Wang

Table 1: Portfolio rules considered in the simulated and empirical experiments.

	No Deviation	Deviation D_1	Deviation D_2	Deviation D_3
Mean	0.2530	0.2818	0.1679	0.2705
Variance	0.2530	0.2818	0.1679	0.2705
SR	0.5030	0.5309	0.4097	0.5201

Table 2: Mean, variance and Sharpe ratio of the tangency portfolio for the simulated data with no deviation, and 5% deviation according to deviation distributions D_1, D_2 and D_3.

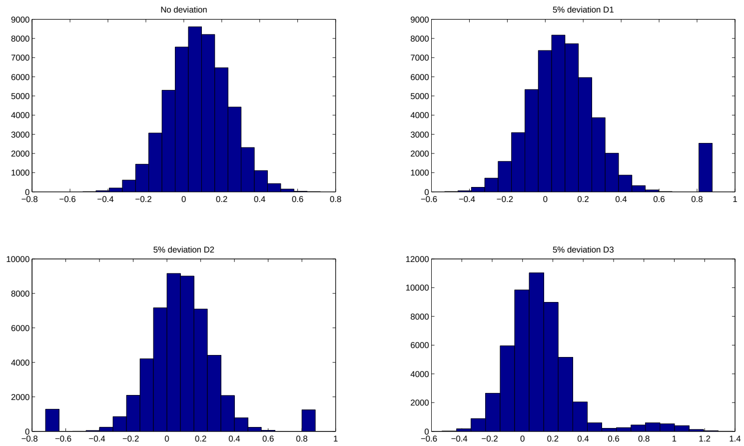

Figure 1: Histograms of the distribution of the first risky asset for simulated data with (i) no deviation (top left); (ii) 5% deviation according to D_1 (top right)), D_2 (bottom left), and D_3 (bottom right).

No deviation	**Rolling window of size 60 months**						**Rolling window of size 120 months**					
	True benchmark			**Historical benchmark**			**True benchmark**			**Historical benchmark**		
Rule	P1	P2	P3	P1-h	P2-h	P3-h	P1	P2	P3	P1-h	P2-h	P3-h
Target	0.0127	0.0633	0.1139	0.0311	0.1557	0.2802	0.0127	0.0633	0.1139	0.0311	0.1557	0.2802
P(Target defeated)	1.0000	1.0000	1.0000	1.0000	1.0000	0.4750	1.0000	1.0000	1.0000	1.0000	1.0000	0.0083
$P(\tilde{\eta} > \eta)$	1.0000	0.6417	0	1.0000	0	0	1.0000	1.0000	0	1.0000	0	0
Mean	0.1088	0.2433	0.3264	0.1707	0.3817	0.5121	0.0923	0.2063	0.2768	0.1447	0.3237	0.4342
Std dev.	0.1630	0.3644	0.4889	0.2557	0.5718	0.7671	0.1524	0.3407	0.4571	0.2389	0.5345	0.7171
CE	0.0955	0.1769	0.2069	0.1380	0.2182	0.2179	0.0807	0.1483	0.1723	0.1161	0.1808	0.1771
Turnover	0.0521	0.1165	0.1563	0.0817	0.1827	0.2452	0.0279	0.0624	0.0837	0.0437	0.0979	0.1313

Deviation D_1	**Rolling window of size 60 months**						**Rolling window of size 120 months**					
	True benchmark			**Historical benchmark**			**True benchmark**			**Historical benchmark**		
Rule	P1	P2	P3	P1-h	P2-h	P3-h	P1	P2	P3	P1-h	P2-h	P3-h
Target	0.0141	0.0705	0.1268	0.0293	0.1465	0.2637	0.0141	0.0705	0.1268	0.0293	0.1465	0.2637
P(Target defeated)	1.0000	1.0000	0.5667	1.0000	0.4833	0.0583	1.0000	1.0000	0.8250	1.0000	0.7667	0.0083
$P(\tilde{\eta} > \eta)$	1.0000	1.0000	0.6250	1.0000	0	0	1.0000	1.0000	1.0000	1.0000	0	0
Mean	0.0744	0.1664	0.2233	0.1073	0.2400	0.3220	0.0784	0.1752	0.2351	0.1130	0.2527	0.3390
Std dev.	0.1958	0.4379	0.5875	0.2824	0.6315	0.8472	0.1883	0.4210	0.5648	0.2715	0.6070	0.8144
CE	0.0553	0.0706	0.0507	0.0675	0.0406	-0.0369	0.0606	0.0866	0.0756	0.0762	0.0684	0.0074
Turnover	0.0850	0.1900	0.2549	0.1225	0.2740	0.3676	0.0365	0.0815	0.1094	0.0526	0.1175	0.1577

Table 3: Out-of-sample mean, standard deviation, CE and Turnover for 6 P-Value rules with respective benchmarks $c_1 = 0.1Q^0_{MV}$, $c_2 = 0.5Q^0_{MV}$, and $c_3 = 0.9Q^0_{MV}$ and their associated historical counterparts. Numerical targets (in terms of CE), the percentage of times it is defeated and the out-of-sample probability of getting a corrected risk aversion parameter larger than the actual one are also provided. Results are for simulated factor model with no deviation (top panel) and 5% deviation D_1 (bottom panel); rolling window of size 60 months (left panel) and 120 months (right panel); risk aversion parameter of 1.

Deviation D_2	**Rolling window of size 60 months**						**Rolling window of size 120 months**					
	True benchmark			**Historical benchmark**			**True benchmark**			**Historical benchmark**		
Rule	P1	P2	P3	P1-h	P2-h	P3-h	P1	P2	P3	P1-h	P2-h	P3-h
Target	0.0084	0.0420	0.0755	0.0219	0.1094	0.1969	0.0084	0.0420	0.0755	0.0219	0.1094	0.1969
P(Target defeated)	1.0000	0.8667	0.6417	1.0000	0.4250	0.2333	1.0000	1.0000	0.9250	1.0000	0.6417	0.2083
$P(\tilde{\eta} > \eta)$	1.0000	0.9167	0.2250	1.0000	0	0	1.0000	1.0000	0.8917	1.0000	0	0
Mean	0.0389	0.0869	0.1166	0.0628	0.1403	0.1883	0.0446	0.0997	0.1337	0.0720	0.1609	0.2159
Std dev.	0.1485	0.3319	0.4454	0.2397	0.5360	0.7191	0.1388	0.3103	0.4163	0.2241	0.5010	0.6721
CE	0.0278	0.0318	0.0174	0.0340	-0.0033	-0.0703	0.0349	0.0515	0.0471	0.0469	0.0354	-0.0100
Turnover	0.0757	0.1693	0.2271	0.1222	0.2733	0.3667	0.0299	0.0668	0.0896	0.0483	0.1079	0.1447

Deviation D_3	**Rolling window of size 60 months**						**Rolling window of size 120 months**					
	True benchmark			**Historical benchmark**			**True benchmark**			**Historical benchmark**		
Rule	P1	P2	P3	P1-h	P2-h	P3-h	P1	P2	P3	P1-h	P2-h	P3-h
Target	0.0135	0.0676	0.1217	0.0294	0.1469	0.2643	0.0135	0.0676	0.1217	0.0294	0.1469	0.2643
P(Target defeated)	1.0000	1.0000	0.5917	1.0000	0.5083	0.0500	1.0000	1.0000	0.9000	1.0000	0.7667	0.0083
$P(\tilde{\eta} > \eta)$	1.0000	1.0000	0.7000	1.0000	0	0	1.0000	1.0000	1.0000	1.0000	0	0
Mean	0.0705	0.1576	0.2115	0.1039	0.2323	0.3117	0.0757	0.1692	0.2270	0.1116	0.2494	0.3345
Std dev.	0.1868	0.4177	0.5604	0.2753	0.6155	0.8258	0.1803	0.4032	0.5410	0.2659	0.5943	0.7971
CE	0.0531	0.0704	0.0545	0.0660	0.0429	-0.0293	0.0594	0.0879	0.0807	0.0762	0.0728	0.0167
Turnover	0.0800	0.1788	0.2399	0.1179	0.2635	0.3536	0.0342	0.0765	0.1026	0.0504	0.1127	0.1512

Table 4: Out-of-sample mean, standard deviation, CE and Turnover for 6 P-Value rules with respective benchmarks $c_1 = 0.1Q^0_{MV}$, $c_2 = 0.5Q^0_{MV}$, and $c_3 = 0.9Q^0_{MV}$ and their associated historical counterparts. Numerical targets (in terms of CE), the percentage of times it is defeated and the out-of-sample probability of getting a corrected risk aversion parameter larger than the actual one are also provided. Results are for simulated factor model with 5% deviation D_2 (top panel) and D_3 (bottom panel); rolling window of size 60 months (left panel) and 120 months (right panel); risk aversion parameter of 1.

	Two-fund rules					Other rules		
	MV	**HRW**	**KZ**	**B**	**P1-h**	**EQ**	**KZ3**	**GUW**
No Deviation								
Mean	0.5059	0.4504	0.3585	0.4478	0.1707	0.0908	0.3837	0.1581
Std dev.	0.7668	0.6867	0.5482	0.6788	0.2557	0.2464	0.5720	0.2651
SR	0.6597	0.6559	0.6539	0.6597	0.6676	0.3684	0.6709	0.5963
(p-value)	(0.566)	(0.47)	(0.412)	(0.55)	-	(0.018)	(0.818)	(0.03)
CE	0.2118	0.2146	0.2082	0.2174	0.1380	0.0604	0.2202	0.1229
Turnover	0.3383	0.3263	0.2694	0.2995	0.0817	0	0.2590	0.2130
Deviation D_1								
Mean	0.2336	0.1880	0.1398	0.2068	0.1073	0.1890	0.1698	0.0272
Std dev.	0.6315	0.5219	0.3979	0.5590	0.2824	0.5253	0.4588	0.1414
SR	0.3699	0.3601	0.3514	0.3699	0.3801	0.3598	0.3701	0.1927
(p-value)	(0.502)	(0.258)	(0.194)	(0.5)	-	(0.76)	(0.726)	(0.012)
CE	0.0342	0.0517	0.0606	0.0505	0.0675	0.0510	0.0646	0.0173
Turnover	0.3146	0.2727	0.2162	0.2785	0.1225	0	0.2096	0.0839
Deviation D_2								
Mean	0.1226	0.0882	0.0609	0.1085	0.0628	0.1244	0.1030	0.0097
Std dev.	0.5428	0.4356	0.3288	0.4805	0.2397	0.5336	0.4002	0.1100
SR	0.2259	0.2025	0.1851	0.2259	0.2618	0.2331	0.2573	0.0885
(p-value)	(0.304)	(0.248)	(0.16)	(0.314)	-	(0.764)	(0.908)	(0.052)
CE	-0.0247	-0.0066	0.0068	-0.0069	0.0340	-0.0180	0.0229	0.0037
Turnover	0.3045	0.2631	0.2064	0.2696	0.1222	0	0.1913	0.0621
Deviation D_3								
Mean	0.2242	0.1799	0.1336	0.1985	0.1039	0.1835	0.1638	0.0247
Std dev.	0.6078	0.5002	0.3803	0.5380	0.2753	0.5094	0.4446	0.1303
SR	0.3690	0.3598	0.3513	0.3690	0.3774	0.3602	0.3684	0.1899
(p-value)	(0.598)	(0.41)	(0.266)	(0.562)	-	(0.782)	(0.802)	(0.018)
CE	0.0395	0.0549	0.0613	0.0538	0.0660	0.0537	0.0650	0.0163
Turnover	0.3069	0.2671	0.2123	0.2717	0.1179	0	0.2044	0.0859

Table 5: Out-of-sample mean, standard deviation, SR, p-value of the Ledoit and Wolf's (2008) test of the difference between SR of P1-h and any other rule, CE and Turnover for 8 feasible rules listed in Table 1. Results are for simulated factor model with no deviation and 5% deviation according to D_1, D_2 and D_3, rolling window of size 60 months, and risk aversion parameter of 1.

	Infeasible			Feasible				Infeasible			Feasible			
T_w	HRW^0	KZ^0	$KZ3^0$	MV	HRW	KZ	KZ3	HRW^0	KZ^0	$KZ3^0$	MV	HRW	KZ	KZ3
	No deviation - SR performance							No deviation - CE performance						
60	19.9311	19.9311	8.7263	19.9311	22.3542	24.2117	16.2281	41.2564	36.1176	16.8718	61.2594	51.5069	43.9226	32.1924
120	12.7790	12.7790	6.9475	12.7790	13.7393	14.1967	10.2282	24.8862	23.9357	13.6166	29.0267	27.1888	26.4314	19.7929
180	5.4339	5.4339	2.9233	5.4339	6.1027	6.3000	4.8300	10.5741	10.8587	5.7695	11.3367	11.8620	12.4357	9.4308
240	3.2202	3.2202	0.8197	3.2202	3.5529	3.6245	1.6821	6.3590	6.6765	1.6382	6.5658	6.9886	7.4531	3.4750
	Deviation D_1 - SR performance							Deviation D_1 - CE performance						
60	18.8450	18.8450	14.3998	18.8450	20.9413	22.5478	18.3820	38.1454	34.1784	27.4454	53.4540	44.8758	40.2570	34.9792
120	10.0570	10.0570	7.7069	10.0570	10.5940	10.8020	10.0064	19.9529	19.1032	14.8949	23.4916	21.2343	20.4379	19.2302
180	8.0537	8.0537	6.5869	8.0537	8.3677	8.4418	8.2249	15.9076	15.4641	12.8161	17.5547	16.5819	16.1754	15.8658
240	6.0733	6.0733	5.4877	6.0733	6.2539	6.2847	6.5265	12.1493	11.8015	10.7835	13.2204	12.5327	12.1952	12.7142
	Deviation D_2 - SR performance							Deviation D_2 - CE performance						
60	21.6151	21.6151	6.4485	21.6151	24.7188	27.3431	14.0718	43.8355	38.7996	13.2937	78.5373	59.6369	49.5177	31.0490
120	9.4654	9.4654	-0.3998	9.4654	11.0591	11.9687	4.8249	19.3083	18.0781	-0.7743	27.9601	24.0506	22.8630	10.7116
180	2.4227	2.4227	-3.9240	2.4227	3.2983	3.6349	-0.4476	5.0250	4.8041	-7.9889	8.1894	7.1851	7.1432	-0.5643
240	0.3230	0.3230	-2.9226	0.3230	0.7652	0.8800	-0.9805	0.8722	0.6449	-5.9272	2.9365	2.0172	1.7705	-1.6231
	Deviation D_3 - SR performance							Deviation D_3 - CE performance						
60	18.1441	18.1441	13.6149	18.1441	20.2065	21.8393	17.3991	36.9726	33.0286	26.3104	53.2753	43.8230	39.1319	33.3872
120	9.5865	9.5865	7.3212	9.5865	10.0524	10.2416	9.4940	19.0282	18.2542	14.2013	22.6751	20.1575	19.4357	18.2890
180	7.9361	7.9361	6.5950	7.9361	8.1894	8.2519	8.1624	15.6954	15.2480	12.8764	17.4468	16.2433	15.8254	15.7709
240	5.6895	5.6895	5.2532	5.6895	5.8256	5.8502	6.2743	11.2882	11.0563	10.3003	12.2509	11.5687	11.3581	12.1964

Table 6: Out-of-sample losses (in percentages) from using a specific rule instead of the optimal (infeasible MV^0) as a function of the size of the rolling window T_w: SR losses (left panel) and CE losses (right panel). Results are for simulated factor model with no deviation and 5% deviations D_1, D_2 and D_3; risk aversion parameter of 1.

	Two-fund rules							Other rules		
	MV	**HRW**	**KZ**	**B**	**P1-h**	**P2-h**	**P3-h**	**EQ**	**KZ3**	**GUW**
Target					0.0807	0.4035	0.7264			
P(Target defeated)					0.9917	0.3967	0.1322			
$P(\tilde{\eta} > \eta)$					1.0000	0.3719	0.1240			
Mean	0.6133	0.5158	0.3146	0.4927	0.2732	0.6110	0.8197	-1.7064	0.4173	0.1502
Std dev.	1.4214	1.2393	0.7866	1.1418	0.5790	1.2948	1.7371	6.2408	0.8969	0.4930
SR	0.4315	0.4162	0.4000	0.4315	0.4719	0.4719	0.4719	-0.2734	0.4653	0.3047
(p-value)	(0.14)	(0.144)	(0.15)	(0.146)	(-)	(-)	(-)	(0.04)	(0.88)	(0.122)
CE	-0.3968	-0.2522	0.0052	-0.1591	0.1056	-0.2272	-0.6890	-21.1804	0.0151	0.0287
Turnover	0.0950	0.0799	0.0491	0.0763	0.0436	0.0975	0.1308	0	0.0476	0.0252

Table 7: Out-of-sample mean, standard deviation, SR, p-value of the Ledoit and Wolf's (2008) test of the difference between SR of P1-h and any other rule, CE and Turnover for 8 feasible rules listed in Table 1. Numerical targets, the percentage of times it is defeated and the out-of-sample probability of getting a corrected risk aversion parameter larger than the actual one are also provided. Results are for dataset with 10 Industry portfolios, window of size 60, and risk aversion parameter of 1.

	Two-fund rules							Other rules		
	MV	**HRW**	**KZ**	**B**	**P1-h**	**P2-h**	**P3-h**	**EQ**	**KZ3**	**GUW**
Target					0.1866	0.9328	1.6790			
P(Target defeated)					0.2149	0.1901	0.1488			
$P(\tilde{\eta} > \eta)$					1.0000	1.0000	0.8843			
Mean	5.7671	5.3693	0.0890	0.9454	1.0187	2.2776	3.0557	-1.5881	0.1281	0.8436
Std dev.	29.1732	28.0544	0.5774	4.7825	4.0761	9.1136	12.2270	5.5778	0.6513	8.3754
SR	0.1977	0.1914	0.1541	0.1977	0.2499	0.2499	0.2499	-0.2847	0.1966	0.1007
(p-value)	(0.072)	(0.05)	(0.066)	(0.086)	(-)	(-)	(-)	(0.026)	(0.214)	(0.078)
CE	-419.7703	-388.1560	-0.0777	-10.4907	-7.2888	-39.2508	-71.6937	-17.1441	-0.0840	-34.2302
Turnover	12.8885	12.0922	0.2147	2.1129	2.1534	4.8147	6.4595	0	0.2713	2.5930

Table 8: Out-of-sample mean, standard deviation, SR, p-value of the Ledoit and Wolf's (2008) test of the difference between SR of P1-h and any other rule, CE and Turnover for 8 feasible rules listed in Table 1. Numerical targets, the percentage of times it is defeated and the out-of-sample probability of getting a corrected risk aversion parameter larger than the actual one are also provided. Results are for dataset with 49 Industry portfolios, window of size 60, and risk aversion parameter of 1.