Robust Mean-Variance Portfolio Selection

Cédric Perret-Gentil and Maria-Pia Victoria-Feser

No 2003.02

Cahiers du département d'économétrie
Faculté des sciences économiques et sociales
Université de Genève

Février 2003

Département d'économétrie
Université de Genève, 40 Boulevard du Pont-d'Arve, CH -1211 Genève 4
http://www.unige.ch/ses/metri/

First version: April 2003

Robust Mean-Variance Portfolio Selection*

Abstract

This paper investigates model risk issues in the context of mean-variance portfolio selection. We analytically and numerically show that, under model misspecification, the use of statistically robust estimates instead of the widely used classical sample mean and covariance is highly beneficial for the stability properties of the mean-variance optimal portfolios. Moreover, we perform simulations leading to the conclusion that, under classical estimation, model risk bias dominates estimation risk bias. Finally, we suggest a diagnostic tool to warn the analyst of the presence of extreme returns that have an abnormally large influence on the optimization results.

Cédric Perret-Gentil
Union Bancaire Privée, 96-98 rue du Rhône, CH-1204 Geneva, Switzerland
e-mail: cedric.perret-gentil@switzerland.org

Maria-Pia Victoria-Feser (corresponding author)
HEC, University of Geneva, 102 bd Carl-Vogt, CH-1211 Geneva, Switzerland
e-mail: maria-pia.victoriafeser@hec.unige.ch

*Cédric Perret-Gentil gratefully acknowledges Laurent Reiss and François-Serge Lhabitant at Union Bancaire Privée for their support. Maria-Pia Victoria-Feser gratefully acknowledges the financial support of the Swiss National Science Foundation (grant 610-057883.99).

1 Introduction

The seminal work by Markowitz (1952, 1959) opened the era of modern finance, and the mean-variance framework is the root of modern investment theory. As Britten-Jones (1999) notes: "Mean-variance analysis is important for both practitioners and researchers in finance. For practitioners, theory suggests that mean-variance efficient portfolios can play an important role in portfolio management applications. For researchers in finance, mean-variance analysis is central to many asset pricing theories as well as to empirical tests of those theories; however, practitioners have reported difficulties in implementing mean-variance analysis. For example, Black and Litterman (1992) note that, 'when investors have tried to use quantitative models to help optimize the critical allocation decision, the unreasonable nature of the results has often thwarted their efforts' (p.28)."

To compute the mean-variance efficient frontier and use the information it provides to select the unique optimal portfolio for a given level of risk (or return), we have to know the stochastic mechanism generating the returns for a given set of securities. In its standard formulation, the mean-variance efficient frontier (MVEF) model makes the assumption that the securities' returns are independently and identically distributed (i.i.d.) as a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the vector of the securities' mean returns and $\boldsymbol{\Sigma}$ is the covariance matrix of the securities' returns.[1] However, the uniqueness of the MVEF solution (i.e. the uniqueness of the optimal weights of the securities in the portfolio given a specific level of risk or return) depends on the implicit assumption that the inputs $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are *known*, whereas they must be *estimated* and therefore are subject to *statistical error*.

[1] The stochastic process generating the data can also include serial dependence, see e.g. Merton (1969, 1971). In practice however, the multivariate normal i.i.d. model is widely used.

The MVEF is computed after estimating the model (estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$), and the resulting optimal portfolios might then be heavily biased by this statistical error occurring during the estimation process.

This type of error is called *estimation risk* (see for instance Michaud (1989)). It can be shown that if we construct, by means of simulations, a confidence region of statistically equivalent portfolios for each optimal portfolio on the efficient frontier, these statistically equivalent portfolios may have significantly (and even radically) different structures in terms of the weights of the different securities. Moreover, on average, the Sharpe ratios[2] of the selected portfolios are not only significantly below the true Sharpe ratio, but also below the Sharpe ratio of an arbitrarily-chosen portfolio, for instance an equally weighted portfolio (see Jobson and Korkie (1981)). To solve this particular problem, Jorion (1986) suggested the use of Stein's (1956) Bayes-Stein shrinkage estimator for the mean return. Other authors have also studied estimation risk in the context of portfolio selection and suggested several types of correction.[3]

There is however another type of statistical error, which is concerned with the distribution of the data generating the model. As it is assumed that the observed returns are realizations of a multivariate normal distribution, what happens if this assumption is slightly violated, for instance when one or more of the securities have unexpected (non-normal) high or low returns? This model deviation called *model risk* was already addressed in the context of the MVEF model by Victoria-Feser (2000). More recently Cavadini, Sbuelz, and Trojani

[2] The Sharpe ratio is calculated as the difference between the portfolio's mean return and the relevant risk-free rate mean, divided by the standard deviation of the portfolio's returns.

[3] See for instance Barry (1974), Bawa, Brown, and Klein (1979), Alexander and Resnick (1985), Chopra and Ziemba (1993), and more recently ter Horst, de Roon, and Werker (2002).

(2002) have considered model risk and estimation risk simultaneously in the same context of portfolio choices, showing that model risk generates greater bias than estimation risk.

In the statistical literature this problem of model risk is referred to as a problem of *statistical robustness*. Robustness is concerned with the stability of estimators of parameters from a given model when model misspecification exists, and in particular in the presence of outlying observations.[4] Robust estimators have been extensively developed in statistics since the pioneering work of Huber (1964) and Hampel (1968).

A fundamental tool used for studying statistical robustness is the *influence function* proposed by Hampel (1968, 1974). The influence function is useful to determine analytically and numerically the *stability properties* of an estimator in case of model misspecification. In this paper, the influence function is used for studying the behavior of the estimator of the optimal portfolio weights, as well as for building a diagnostic tool to detect outlying returns. We also investigate by means of real market data and simulations how the violation of the multivariate normal assumption can seriously affect the optimality characteristics of the solution of the MVEF model when computed with sample mean and sample covariance estimators.

The contributions of this paper to the literature are therefore twofold. First, we show analytically that the necessary and sufficient condition for the mean-variance portfolio optimizer to be robust to local nonparametric departures from multivariate normality is that the estimators of the model's parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be robust with bounded influence functions. We suggest such robust estimators and show their remarkable behavior in the presence of

[4] Outlying observations can be defined as data points that have an infinitesimal probability of being generated by the model generating the rest of the data.

outlying observations in a simulation study. Second, we present a diagnostic tool based on the influence function for detecting the outlying data from the sample that have an abnormally large influence on the optimization process. This tool assesses the quality of the data before their use in the optimizer, and is of particular interest to the analyst.

The paper is organized as follows. In section 2, we present two examples of mean-variance optimization applied to real market data. We show that a robust optimizer can lead to optimal portfolios that are different from the ones given by a classical (non-robust) optimizer, and that this difference is often due to only a few outlying returns from the sample (i.e. under slight model deviations). Section 3 contains the theoretical aspects of this paper. We first present the basic concepts of robust statistics and then use them to study the robustness properties of a portfolio optimizer. In particular, we show analytically that if the model parameters are not robustly estimated, the resulting optimizer can be seriously biased, leading to sub-optimal portfolio choices. We also develop a diagnostic tool based on the influence function for detecting influential returns from the sample. In section 4, a simulation study is performed to investigate the stability properties of the robust optimizer when compared with its classical counterpart. Section 5 concludes.

2 Market data examples

2.1 Diversified portfolio

This example replicates the kinds of security that investors may hold in their portfolios. The data set is composed of series of returns on bonds, stocks and alternatives (hedge funds) represented by the following indices.[5] For the bonds, we use a Merrill Lynch index available

[5] Even if series of returns from the financial reality may exhibit autocorrelation, this case is of interest and raises a few important questions.

on Bloomberg (IND H355 <GO>). This index includes government and corporate bonds with ratings ranging from AAA to A, focuses on maturities between 1 and 10 years and is diversified across the following markets: United States, Canada, United Kingdom, Euroland, Switzerland, Denmark, Sweden, Japan and Australia. The well-known Morgan Stanley Capital International (MSCI) world developed market index is used for the stocks, and its data are collected from Datastream. The two above-mentioned indices are total-return indices (i.e. including reinvestment of coupons/dividends). The alternative investment part of the portfolio is represented by the Credit Suisse First Boston (CSFB) / Tremont hedge fund index, and data come from their web site (www.hedgeindex.com). For the three above-mentioned asset classes, monthly returns span January 1997 to December 2002, for a total of 72 observations.[6] The currency of all indices is the unhedged US dollar (USD).

The returns of these three series are presented in Figure 1. We notice that some returns of the stocks are extreme, for instance data point 20, with a very low return, and a few data points between observations 50 and 70. On the other hand, data points 7, 36 and 38 exhibit three rather high returns for the alternatives. These extreme returns heavily bias the sample mean, sample covariance and sample correlation estimators, and this bias can be seen by comparing the results of the above estimators, which we characterize as *classical* estimators, with those of *robust* estimators for the mean, covariance and correlation of the return series. We suggest the use of the translated-biweight S-estimator (TBS estimator) proposed by Rocke (1996) as robust estimator (for details, see section 3). All calculations are done with S-Plus and its numerical optimizer NUOPT.

[6] In the following of this paper, logarithmic returns are used for calculation.

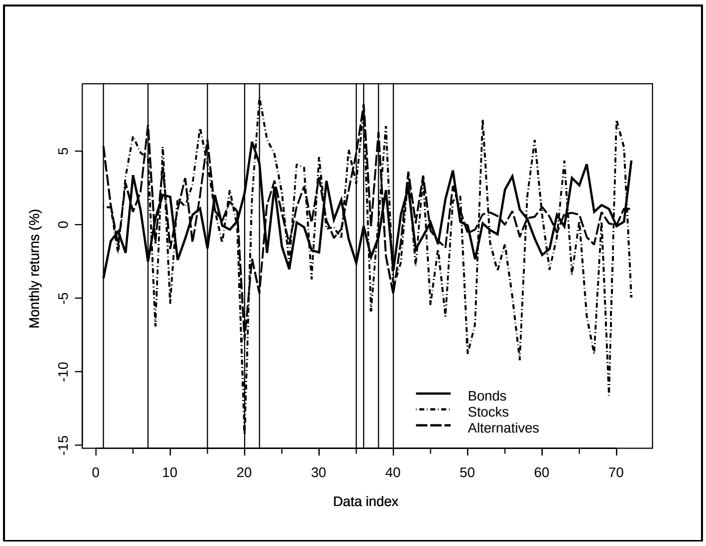

Figure 1. Diversified portfolio: serial plot of return series. This figure shows the (unhedged) USD logarithmic monthly returns of the three series used to build the diversified portfolio. Observations span January 1997 to December 2002 for a total of 72 indexed data. The 9 vertical lines correspond to the 9 most influential returns detected by the gross error sensitivity (see later).

The comparison between the classically estimated correlation matrix and its robust counterpart is presented in Figure 2, in which the ellipses represent the correlations of the (assumed) bivariate normal distribution between the different pairs of return series. The corresponding values of the correlations between the securities are shown in the lower triangular part of the Figure. Classical correlations are the numbers below, robust correlations are the ones above. Correlations in terms of numbers are different between the robust and the classical estimates, especially in the case of bonds and alternatives. Moreover, taking into account the differences in classical and robust mean return estimates (for bonds, stocks and alternative investments respectively), we find $\hat{\boldsymbol{\mu}}^{\text{cla}} = (0.31\%, 0.09\%, 0.77\%)$ and $\hat{\boldsymbol{\mu}}^{\text{rob}} = (0.30\%, -0.12\%, 0.68\%)$. It can be seen that the estimated mean returns differ substantially for the stocks. The vectors of standard deviation of returns also exhibit differences:

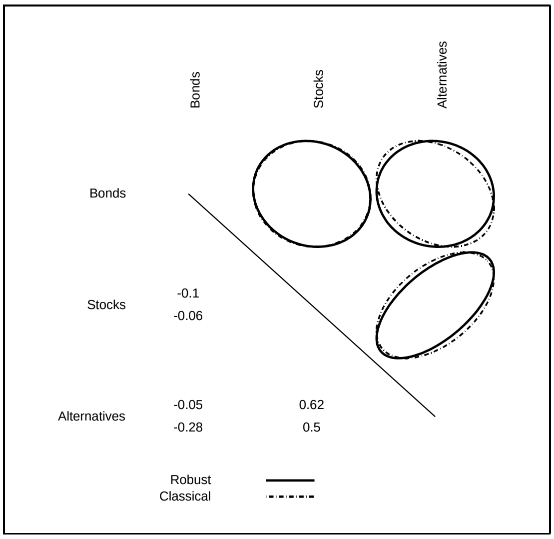

Figure 2. Diversified portfolio: correlation structure. This figure shows a double comparison between the results of classical and robust estimation. On the lower triangular part, classical correlations estimated by means of the sample correlation estimator (below) are compared with robust correlations estimated by means of the TBS estimator (above). On the upper triangular part, a comparison of the correlation structure is shown where the classical and robust ellipses represent the correlations of the (assumed) bivariate normal distribution between the different pairs of return series.

classical standard deviations are $(2.03\%, 4.95\%, 2.53\%)$, and robust standard deviations are $(2.07\%, 5.25\%, 1.70\%)$ for bonds, stocks and alternatives respectively.

To assess the effect of these differences in estimation on the results of the portfolio optimizer, the classical (calculated with the classically estimated $\hat{\boldsymbol{\mu}}^{\text{cla}}$ and $\widehat{\boldsymbol{\Sigma}}^{\text{cla}}$) and the robust (calculated with the robustly estimated $\hat{\boldsymbol{\mu}}^{\text{rob}}$ and $\widehat{\boldsymbol{\Sigma}}^{\text{rob}}$) mean-variance efficient frontiers[7] are presented in Figure 3.

Classically and robustly estimated efficient frontiers give an important insight into the statistical properties of the data. The two efficient frontiers are distinct from each other,

[7] Calculation in this example has been done without allowing for short selling, as is mainly the case in managed portfolios. This restriction is lifted in section 4.

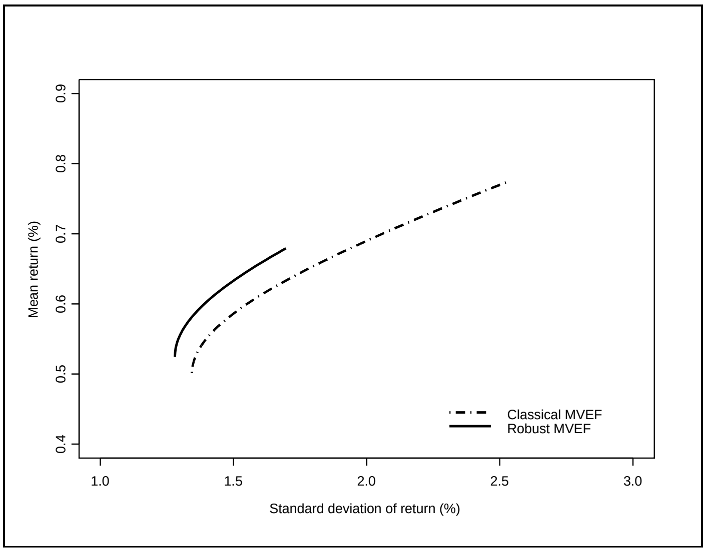

Figure 3. Diversified portfolio: classically and robustly estimated efficient frontiers. This figure shows the classical mean-variance efficient frontier (MVEF) when the mean and covariance of returns are estimated using the sample mean and sample covariance estimators, as well as the robust MVEF when the mean and covariance of returns are estimated using the TBS estimator. Short selling is not allowed.

meaning that the influential data points that affect classical estimation are *treated* by robust estimation, so that their influence is kept *under control*. The robust efficient frontier being higher and more to the left than the classical one clearly indicates a reduction in the volatility structure of the inputs due to their robust estimation. Other interesting information is given by the composition of the so-called *minimum variance portfolios*, which are the portfolios characterized by the minimum risk on the two efficient frontiers. The classical and robust weights of the three assets (bonds, stocks and alternatives respectively) in these particular portfolios are $\hat{\mathbf{p}}^{\mathrm{cla}} = (0.59, 0, 0.41)$ and $\hat{\mathbf{p}}^{\mathrm{rob}} = (0.41, 0, 0.59)$. The fact that the weights of bonds and alternatives are the exact reverse of each other is due to chance, but more worrying is that portfolio compositions that should represent the same reality (i.e. minimum risk) are quite different from each other. Stocks are obviously given a zero weight by the optimizer for such a low-risk portfolio, due to their high standard deviation of returns over the period.

A further step in the analysis is to check which observations are considered outliers and responsible for this shift of the efficient frontier. To do so, we use a diagnostic tool called *gross error sensitivity*, which we present in section 3. Briefly, the gross error sensitivity applied to our case is a tool to detect outlying returns that have an abnormally high influence on the estimator of the optimal portfolio weights. The gross error sensitivity for the diversified portfolio data set is given in Figure 4. The 9 returns that appear the most influential

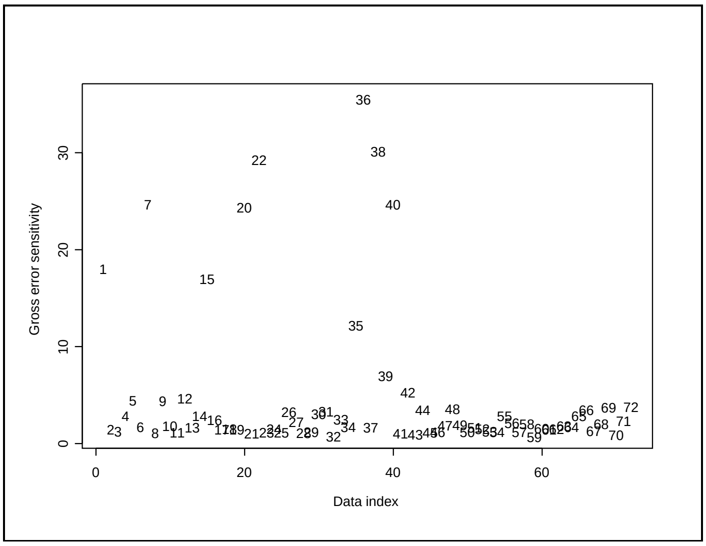

Figure 4. Diversified portfolio: gross error sensitivity. This figure shows the relative influence of each of the 72 returns on the estimator of the optimal portfolio weights as detected by means of the gross error sensitivity diagnostic tool. As this measure relates to a specific portfolio on the classical mean-variance efficient frontier, the level of standard deviation of returns has been set to 1.5%.

(compared with the majority) have also been highlighted in the time series plot of Figure 1 by means of vertical lines. The striking feature is that these 9 returns do not always correspond to the most extreme returns of the different series, and that makes them difficult to find by simple visual inspection. For example, when looking at Figure 1, data point 20 (composed of the three series) seems more statistically outlying than data point 36, which is not true when looking at Figure 4. In fact, the power of the gross error sensitivity is that it takes into account the multivariate structure of the model and highlights the influential data points according to the specific estimator used.

In short, we have found that a few outlying observations in the data have a strong influence on the composition of the resulting optimal portfolios.[8]

2.2 Low volatility hedge fund portfolio

Let's now turn to another example in which we remain in the same asset class.[9] We suggest building a portfolio composed only of alternatives (hedge funds) and we decide to include the three hedge fund strategies that exhibit the lowest standard deviation of returns over a defined past period. To do so, we use the CSFB/Tremont hedge fund index family, which is composed of a main index (used in the previous example) and 9 major sub-indices each reflecting a specific hedge fund strategy.[10]

When using the full available return history in (unhedged) USD from January 1994 to December 2002, the three hedge fund strategies with lowest standard deviation of returns are *convertible arbitrage* (CA), *event-driven* (ED) and *fixed income arbitrage* (FIA) with respective monthly standard deviation of returns of (1.41%,1.86%,1.20%). We refer the reader to Lhabitant (2002) for specific details and characteristics of these strategies.

[8] These 9 outlying points represent 12.5% of the data.

[9] As in the previous case, no direct conclusion can be drawn from this specific example as returns may exhibit serial dependence. However, a few interesting questions can be raised.

[10] Due to the particularity of hedge fund investing (low liquidity, lock-up periods, etc.), monthly portfolio re-balancing implicitly underlying the MVEF model estimated on monthly data might sometimes be difficult to achieve.

The returns of these three strategies are presented in Figure 5. Compared with the

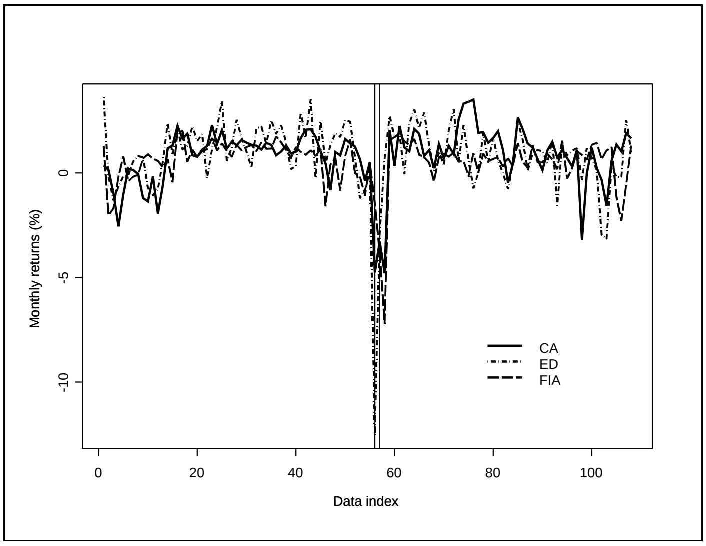

Figure 5. Low volatility hedge fund portfolio: serial plot of return series. This figure shows the (unhedged) USD logarithmic monthly returns of the three series used to build the low volatility hedge fund portfolio. Observations span January 1994 to December 2002 for a total of 108 indexed data. The 2 vertical lines correspond to the 2 most influential returns detected by the gross error sensitivity (see later).

preceding example, data seem more stable, with the notable exception of a few data points just before point 60, which correspond to the Russian financial crisis of August 1998.

Presented in Figure 6 is a comparison of the correlation structure under classical and robust estimation. Again, there are differences between classical and robust correlation estimates. The same is true for the classical and robust mean vectors (for convertible arbitrage, event-driven and fixed income arbitrage strategies respectively) which are $\hat{\boldsymbol{\mu}}^{\mathrm{cla}} = (0.81\%, 0.83\%, 0.54\%)$ and $\hat{\boldsymbol{\mu}}^{\mathrm{rob}} = (1.14\%, 1.26\%, 0.85\%)$: major differences exist between the classical and robust mean vectors. Hence, even with series exhibiting low standard deviation of returns, classical and robust estimations do not give the same results, and this

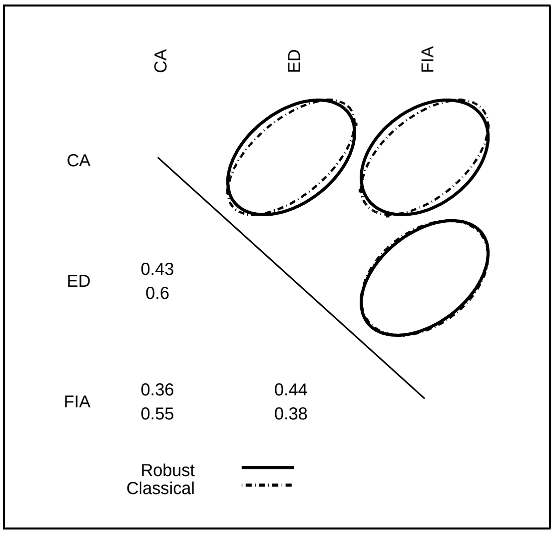

Figure 6. Low volatility hedge fund portfolio: correlation structure. This figure shows a double comparison between the results of classical and robust estimation. On the lower triangular part, classical correlations estimated by means of the sample correlation estimator (below) are compared with robust correlations estimated by means of the TBS estimator (above). On the upper triangular part, a comparison of the correlation structure is shown where the classical and robust ellipses represent the correlations of the (assumed) bivariate normal distribution between the different pairs of return series.

difference is further evidenced by classical and robust efficient frontiers[11] shown in Figure 7.

As in the previous example, the robust efficient frontier is located higher than and to the left of the classical one, indicating the presence of outlying data points. Looking at the composition of the minimum variance portfolios, the classical and robust weights (for convertible arbitrage, event-driven and fixed income strategies respectively) are $\hat{\mathbf{p}}^{\text{cla}} = (0.26, 0.09, 0.65)$ and $\hat{\mathbf{p}}^{\text{rob}} = (0.31, 0, 0.69)$. As can be seen, the difference in weights is mainly concentrated on the first two hedge fund strategies. There again, it is interesting to identify the outly-

[11] Calculation has been done without allowing for short selling

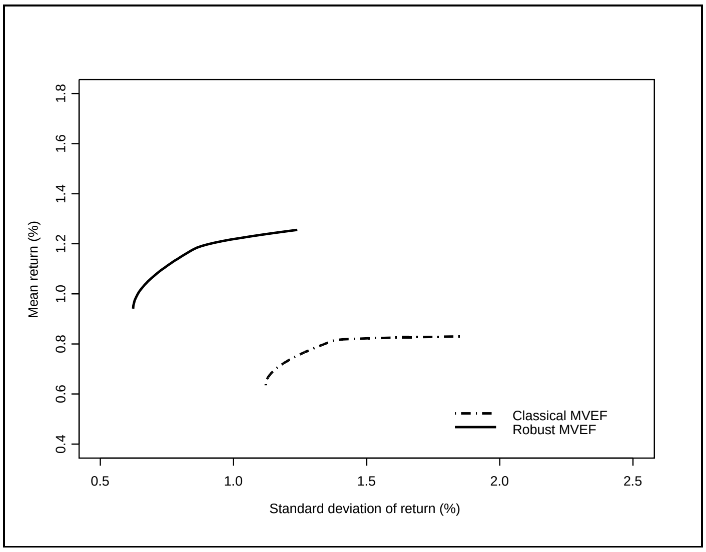

Figure 7. Low volatility hedge fund portfolio: classically and robustly estimated efficient frontiers. This figure shows the classical mean-variance efficient frontier (MVEF) when the mean and covariance of returns are estimated using the sample mean and covariance estimators, as well as the robust MVEF when the mean and covariance of returns are estimated using the TBS estimator. Short selling is not allowed.

ing returns responsible for such a change, and we make use of the gross error sensitivity whose results are reported in Figure 8. The 2 most influential return points have also been highlighted in the serial plot of Figure 5 by means of vertical lines.

Data points 56 and 57 correspond to the very low returns recorded during the Russian financial crisis, and they are detected as having a very strong relative influence on the classical estimates of the optimal portfolio weights.

In short, in this case of series exhibiting rather low standard deviation of returns, the influence of just a few outlying data points from the sample can be very strong on the classical estimates needed to calculate the efficient frontier and thus lead to different portfolio

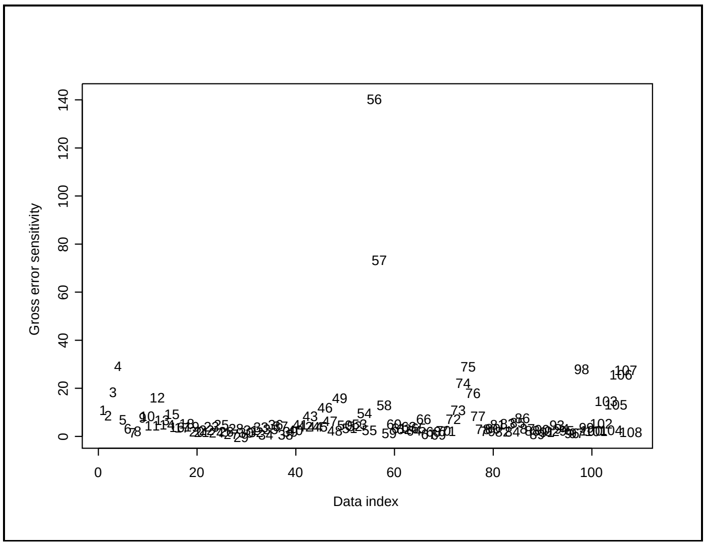

Figure 8. Low volatility hedge fund portfolio: gross error sensitivity. This figure shows the relative influence of each of the 108 returns on the estimator of the optimal portfolio weights as detected by means of the gross error sensitivity diagnostic tool. As this measure relates to a specific portfolio on the classical mean-variance efficient frontier, the level of standard deviation of returns has been set to 1.3%.

choices.[12]

At this point, we may still wonder what would be a strong enough reason for choosing a robust portfolio composition rather than a classical one, as both kinds of optimal portfolio could be considered acceptable. In fact, the reason for preferring a robust portfolio composition will become obvious in section 4, and will be strongly motivated by the sensitivity of the sample mean and covariance estimated MVEF model to the data as shown by a simulation study.

In the next section we review the basic concepts of robust statistics and apply them to the MVEF model.

[12] These 2 strongly outlying observations represent 1.85% of the total 108 observations.

3 Robustness properties of the MVEF model

3.1 Basic concepts of robust statistics

The pioneering work of Huber (1964) and Hampel (1968) has laid the ground for the theory of robust statistics. In short, as a generalization of classical theory, robust statistics takes into account the possibility of model misspecification (i.e. model deviation). This theory and its results are valid *at* the model as well as *in a neighborhood* of the model,[13] which is not the case for classical statistics, which is only valid *at* the model.

Let's denote

$$\{G_\varepsilon | G_\varepsilon = (1-\varepsilon)F_\theta + \varepsilon W\}, \tag{1}$$

with W an arbitrary distribution and $\varepsilon \in [0,1]$, the set of all distributions defining a neighborhood of the parametric model F_θ. This neighborhood includes all possible misspecified distributions around F_θ. G_ε can be considered a mixture distribution between F_θ and the contamination distribution W, and one particular case is when $W = \Delta_{\mathbf{z}}$, the distribution that gives a probability of one to a point $\mathbf{z}$ chosen arbitrarily.[14] In this case, the neighborhood of the model featuring all local nonparametric departures from F_θ is given by

$$\{F_\varepsilon | F_\varepsilon = (1-\varepsilon)F_\theta + \varepsilon \Delta_{\mathbf{z}}\}. \tag{2}$$

Hence F_ε generates observations from F_θ with probability $(1-\varepsilon)$ and observations equal to an arbitrary point $\mathbf{z}$ with probability ε.

One way of assessing the robustness properties of an estimator $\hat{\theta}$ of θ is to study its (asymptotic) stability properties in a neighborhood of the model considering a distribution

[13] In the presence of outlying observations acting as local nonparametric departures from the model, the distributional assumptions are violated and we therefore end up in a *neighborhood* of the model.

[14] $\mathbf{z}$ can be a scalar or a vector.

of type G_ε, and there is no loss of generality in focusing on the particular case of F_ε since Hampel et al. (1986) showed that the maximal bias on $\hat{\theta}$ is precisely obtained at $W = \Delta_{\mathbf{z}}$.

Considering the case when ε tends towards zero,[15] we get the so-called *influence function* (IF) suggested by Hampel (1968, 1974) and further developed by Hampel et al. (1986). The IF gives the influence of an infinitesimal amount of contamination $\mathbf{z}$ on the value of the estimator viewed as a function of the underlying distribution. The influence function is then defined as

$$\mathrm{IF}(\mathbf{z}, \hat{\theta}, F_\theta) = \lim_{\varepsilon \downarrow 0} \left[\frac{\hat{\theta}(F_\varepsilon) - \hat{\theta}(F_\theta)}{\varepsilon} \right],$$

and, when the derivative exists, as

$$\mathrm{IF}(\mathbf{z}, \hat{\theta}, F_\theta) = \left. \frac{\partial}{\partial \varepsilon} \hat{\theta}(F_\varepsilon) \right|_{\varepsilon=0}. \tag{3}$$

The IF is the directional derivative of the estimator $\hat{\theta}$ in a single point contamination direction $\Delta_{\mathbf{z}}$. Depending on the situation, this directional derivative can be scalar, vector or matrix valued.

The IF is a powerful tool for assessing the robustness properties of estimators. Indeed, Hampel et al. (1986) show that *only* the IF is needed to fully describe the asymptotic bias of an estimator caused by a contamination, implying that an estimator with a bounded IF has automatically a bounded asymptotic bias. Therefore, an estimator with a bounded IF is robust in a general neighborhood of the parametric model defined by (1).

The IF can also be used as a diagnostic tool to detect observations that have a large influence on the estimator $\hat{\theta}$. The *gross error sensitivity* (GES) is such a tool. It is defined

[15] We consider an infinitesimal amount of contamination, and therefore remain striclty in a *close* neighborhood of the model.

as the Euclidean norm of the influence function defined under (3) as

$$\mathrm{GES}(\mathbf{x}, \hat{\theta}) = \left(\mathrm{IF}(\mathbf{x}, \hat{\theta}, F_\theta)^{'}\mathrm{IF}(\mathbf{x}, \hat{\theta}, F_\theta)\right)^{1/2}. \tag{4}$$

An observation $\mathbf{x}$ with a large GES is then considered an influential observation. To compute the GES, the true parameter value θ has to be known, which is seldom the case in practice. The value of θ has then to be estimated in a robust way so as to ensure that this diagnostic tool is not biased by the outlying observations it is supposed to detect.

3.2 Classical MVEF model estimation

Let's suppose that there are N securities to choose from and let $\mathbf{p} = (p_1, \ldots, p_N)^{'}$ be the vector of portfolio weights, such that $\sum_{i=1}^{N} p_i = 1$. Recalling that $\boldsymbol{\mu}$ is the vector of size N containing the mean returns of the securities, and that $\boldsymbol{\Sigma}$ is the $(N \times N)$ covariance matrix of the returns of the securities, the portfolio mean can be written as

$$R(\mathbf{p}) = \boldsymbol{\mu}^{'}\mathbf{p},$$

and the portfolio variance as

$$S(\mathbf{p}) = \mathbf{p}^{'}\boldsymbol{\Sigma}\mathbf{p}.$$

For a given value of the risk aversion parameter λ, the mean-variance optimization selects the portfolio $\mathbf{p}^*$ which maximizes in $\mathbf{p}$

$$R(\mathbf{p}) - \frac{\lambda}{2}S(\mathbf{p}),$$

subject to

$$\mathbf{e}_N^{'}\mathbf{p} = \mathbf{1},$$

where $\mathbf{e}_N$ of size $(N \times 1)$ is a vector of ones. The set of optimal portfolios for all possible values of the risk aversion parameter λ defines the mean-variance efficient frontier. Depending on the situation, the constraint of no short selling ($\mathbf{p} \geq 0$) might be added as well as other constraints.

In the unconstrained case, the solution is explicit and it is well known that the optimal portfolio weights for a given value of λ are given by

$$\mathbf{p}^* = \frac{1}{\lambda} \mathbf{\Sigma}^{-1} \left(\boldsymbol{\mu} - \eta \mathbf{e}_N \right), \tag{5}$$

with

$$\eta = \left(\mathbf{e}_N^{'} \mathbf{\Sigma}^{-1} \boldsymbol{\mu} - \lambda \right) \Big/ \mathbf{e}_N^{'} \mathbf{\Sigma}^{-1} \mathbf{e}_N. \tag{6}$$

As can be seen, $\mathbf{p}^*$ depends directly on $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, implying that the resulting estimated optimal portfolio weights are directly affected by potential estimation bias in the mean and covariance of returns. This is reflected in the following theorem, where we demonstrate that the influence function of the estimator of the optimal portfolio weights depends directly on the influence functions of both the estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$.

Theorem 1. *The (asymptotic) bias of the resulting estimator $\hat{\mathbf{p}}^*$ of the optimal portfolio weights only depends on the (asymptotic) bias of the estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$.*

Proof. We make use of the IF to describe the behavior of the asymptotic bias of $\hat{\mathbf{p}}^*$ in a neighborhood of the model.
Let's first write $\hat{\mathbf{p}}^* = \hat{\mathbf{p}}^*(\hat{\boldsymbol{\mu}}, \widehat{\mathbf{\Sigma}})$, as the estimated optimal portfolio weights given by (5) jointly with (6) make $\hat{\mathbf{p}}^*$ depends on both $\hat{\boldsymbol{\mu}}$ and $\widehat{\mathbf{\Sigma}}$. Equation (5) then becomes

$$\hat{\mathbf{p}}^*(\hat{\boldsymbol{\mu}},\widehat{\mathbf{\Sigma}}) = \frac{1}{\lambda} \widehat{\mathbf{\Sigma}}^{-1} \left[\hat{\boldsymbol{\mu}} - \left(\mathbf{e}_N^{'} \widehat{\mathbf{\Sigma}}^{-1} \hat{\boldsymbol{\mu}} - \lambda \right) \mathbf{e}_N \Big/ \mathbf{e}_N^{'} \widehat{\mathbf{\Sigma}}^{-1} \mathbf{e}_N \right], \tag{7}$$

and, under local nonparametric departures from the model as in (2), we have

$$\hat{\mathbf{p}}^*(\hat{\boldsymbol{\mu}},\widehat{\mathbf{\Sigma}}) = \frac{1}{\lambda} \widehat{\mathbf{\Sigma}}(F_\varepsilon)^{-1} \left[\hat{\boldsymbol{\mu}}(F_\varepsilon) - \left(\mathbf{e}_N^{'} \widehat{\mathbf{\Sigma}}(F_\varepsilon)^{-1} \hat{\boldsymbol{\mu}}(F_\varepsilon) - \lambda \right) \mathbf{e}_N \Big/ \mathbf{e}_N^{'} \widehat{\mathbf{\Sigma}}(F_\varepsilon)^{-1} \mathbf{e}_N \right]. \tag{8}$$

Taking the derivative of (8) with respect to ε in $\varepsilon = 0$ to derive the IF of $\hat{\mathbf{p}}^*$ for the mean-variance optimal portfolio for a given value of λ, we get

$$\begin{aligned}\mathrm{IF}(\mathbf{z}, \hat{\mathbf{p}}^*, F_\theta) &= -\boldsymbol{\Sigma}^{-1}\mathrm{IF}(\mathbf{z}, \widehat{\boldsymbol{\Sigma}}, F_\theta)\mathbf{p}^* + \\ &\quad \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\Big\{ \mathrm{IF}(\mathbf{z}, \hat{\boldsymbol{\mu}}, F_\theta) + \Big[\mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\mathrm{IF}(\mathbf{z}, \widehat{\boldsymbol{\Sigma}}, F_\theta)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \\ &\quad \mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\mathrm{IF}(\mathbf{z}, \hat{\boldsymbol{\mu}}, F_\theta)\Big]\,\mathbf{e}_N \Big/ \mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\mathbf{e}_N - \\ &\quad \Big[\mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\mathrm{IF}(\mathbf{z}, \widehat{\boldsymbol{\Sigma}}, F_\theta)\boldsymbol{\Sigma}^{-1}\mathbf{e}_N\Big]\,\Big(\mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \lambda\Big)\,\mathbf{e}_N \Big/ \Big(\mathbf{e}_N^{'}\boldsymbol{\Sigma}^{-1}\mathbf{e}_N\Big)^2\Big\}. \quad (9)\end{aligned}$$

Hence, the IF of $\hat{\mathbf{p}}^*$ depends directly on the IFs of both $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$. This means that the estimator of the optimal portfolio weights directly inherits the stability properties of the estimator of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. ▮

Therefore, unless the mean vector and covariance matrix are robustly estimated, the mean-variance optimizer can lead to portfolio compositions heavily influenced by just a few outlying observations from the sample.

The mean and covariance of the returns are in practice often estimated by their sample counterpart, i.e. by the maximum likelihood (ML) estimator under the multivariate normal model. In this case, the estimator of $\boldsymbol{\mu}$ can be explicitely expressed as

$$\hat{\boldsymbol{\mu}}_{\mathrm{ML}}^{'} = \frac{1}{T}\mathbf{e}_T^{'}\mathbf{Y},$$

where $\mathbf{Y}$ is the $(T \times N)$ matrix containing the returns of each security (columnwise) for T periods, and given by

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^{'} \\ \vdots \\ \mathbf{y}_T^{'} \end{bmatrix} = [\mathbf{y}_1 \ldots \mathbf{y}_N] = \begin{bmatrix} y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ y_{T1} & \cdots & y_{TN} \end{bmatrix}.$$

Moreover, the ML estimator of the covariance matrix $\boldsymbol{\Sigma}$ is calculated as

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \frac{1}{T}\left(\mathbf{Y} - \mathbf{e}_T\hat{\boldsymbol{\mu}}_{\mathrm{ML}}^{'}\right)^{'}\left(\mathbf{Y} - \mathbf{e}_T\hat{\boldsymbol{\mu}}_{\mathrm{ML}}^{'}\right),$$

where $\mathbf{e}_T$ of size $(T \times 1)$ is a vector of ones. Under *normality*, the maximum likelihood estimators are the most efficient. However, in the presence of outlying observations, how can they cope with model deviation and what kind of influence have outlying data on them?

To answer this central question we make use of the influence function. As shown in Hampel et al. (1986), the respective IF for $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ are given by

$$\mathrm{IF}(\mathbf{z}, \hat{\boldsymbol{\mu}}_{\mathrm{ML}}, F_\theta) = -\boldsymbol{\mu} + \mathbf{z}, \tag{10}$$

and

$$\mathrm{IF}(\mathbf{z}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}, F_\theta) = -\boldsymbol{\Sigma} + (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})'. \tag{11}$$

We can easily see that both IFs for the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unbounded: for an infinite number of data points $\mathbf{z}$, both IFs may become *arbitrarily large*.

Making use of the above results, the following theorem derives the explicit influence function of the estimator of the optimal portfolio weights when ML estimation is used for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Theorem 2. *When sample mean and covariance are used to estimate the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the (asymptotic) bias on the resulting estimator $\hat{\mathbf{p}}^*$ of the optimal portfolio weights can be infinite under infinitesimal departures from multivariate normality.*

Proof. The explicit expression for the IF of $\hat{\mathbf{p}}^*$ is given by replacing in (9) the influence functions of the respective ML estimators reported in (10) and (11), and we find

$$\begin{aligned} \mathrm{IF}(\mathbf{z}, \hat{\mathbf{p}}^*, F_\theta) &= (\mathbf{I} - A(\mathbf{z})\boldsymbol{\Sigma})\,\mathbf{p}^* + \\ & \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\left[\mathbf{z} - \boldsymbol{\mu} + \frac{1}{b}\left(d(\mathbf{z}) - c(\mathbf{z})\right)\mathbf{e}_N - \frac{1}{b^2}\left(k(\mathbf{z}) - b\right)(l - \lambda)\,\mathbf{e}_N\right], \end{aligned} \tag{12}$$

where

$$\begin{aligned} A(\mathbf{z}) &= \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}, \\ b &= \mathbf{e}_N'\boldsymbol{\Sigma}^{-1}\mathbf{e}_N, \\ c(\mathbf{z}) &= \mathbf{e}_N'\boldsymbol{\Sigma}^{-1}\mathbf{z}, \\ d(\mathbf{z}) &= \mathbf{e}_N' A(\mathbf{z})\boldsymbol{\mu}, \end{aligned}$$

$$
\begin{aligned}
k(\mathbf{z}) &= \mathbf{e}_N^{'} A(\mathbf{z}) \mathbf{e}_N, \\
l &= \mathbf{e}_N^{'} \mathbf{\Sigma}^{-1} \boldsymbol{\mu}.
\end{aligned}
$$

The IF of $\hat{\mathbf{p}}^*$ under ML estimation is clearly unbounded in $\mathbf{z}$, which means that the asymptotic bias of the estimated optimal portfolio weights can become arbitrarily large under model deviation. ∎

In short, the ML-estimated mean-variance optimizer is *not robust to model risk*. It is therefore of interest to identify, before estimating the model, the returns from the sample that will have an abnormally large influence on the optimization results under maximum likelihood estimation. However, it is difficult to *visually* identify these specific returns, due to the multivariate structure of the model and the statistical properties of the ML estimator. To address this issue, we suggest using as a diagnostic tool the gross error sensitivity given in (4) in conjunction with the influence function given in (12), recalling that $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ have to be estimated in a robust way for the reason already mentioned above. This produces graphs like those in Figures 4 or 8, and influential data points are then easy to identify. However, it should be stressed that there are an infinite number of IF (and then of GES) since there are an infinite number of optimal portfolio weighting schemes. The value of the risk aversion parameter λ has to be fixed so as to characterize the portfolio weighting scheme of interest, and to be able to compute the GES.[16]

3.3 Robust MVEF model estimation

As demonstrated above, the estimator used for the mean vector and the covariance matrix determines the robustness properties of the estimator of the optimal portfolio weights in the

[16] We found however that the choice of the value for λ doesn't have a decisive impact on the detection of outlying returns.

optimizer. We should then choose a robust estimator with a bounded influence function to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.[17]

We suggest using Rocke's (1996) *translated biweight S-estimator* (TBS estimator), which belongs to the class of S-estimators (see Rousseeuw and Leroy (1987)). The TBS estimator has the advantage of being quite efficient compared with other robust estimators. Moreover, its robustness against relatively large quantities of outlying data can be controlled (see below). The TBS estimator is defined as the solution in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} w_i^{\mu}(\mathbf{y}_i - \boldsymbol{\mu}) &= 0, \\ \frac{1}{n}\sum_{i=1}^{n} w_i^{\delta}\boldsymbol{\Sigma} - w_i^{\eta}(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})' &= 0, \end{aligned}$$

where the weights w_i^{μ}, w_i^{δ} and w_i^{η} depend on two control parameters (see again Rocke (1996)).

The first control parameter is the *breakdown point* ε^*, and the second is the *asymptotic rejection probability* (ARP) α. The breakdown point of an estimator is the maximal amount of model deviation (or model risk) it can withstand before it breaks down, i.e. before it can take arbitrary values under model deviations such as in (2) (see Hampel et al. (1986)). The ARP can be interpreted as the probability for an estimator, in large samples and under a reference distribution, to give a null (or nearly null) weight to extreme observations.

The TBS estimator is a consistent estimator of mean and covariance. However, compared with the classical ML estimator, the TBS estimator is less efficient at the model. This loss of efficiency is the *price to pay* for its robust behavior in a neighborhood of the model and for its property of safeguarding the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ against the influence of model

[17] See for instance Victoria-Feser (2000) for a review and discussion of the desirable properties of such a robust estimator.

deviations. A robust mean-variance portfolio optimizer is then obtained by estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by means of the TBS estimator, and by using the resulting estimates in (7).

In the next section, a simulation study is performed to investigate the behavior of mean-variance optimal portfolios when classical and robust estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are used in the presence of contaminated data.

4 Simulation study

We here extend the example of the low volatility hedge fund portfolio by means of a simulation study.

Let the population of returns be i.i.d. and generated by a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. To do so, we consider the robust TBS estimators ($\hat{\boldsymbol{\mu}}_{\mathrm{TBS}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{TBS}}$) computed on the *original* low volatility hedge fund portfolio data set as being the parameters of this new trivariate normal population of returns for the convertible arbitrage, event-driven and fixed income arbitrage strategies.[18] We later refer to this new population as the *true* model. From this population, we simulate 20 samples of size 120 representing 10 years of trivariate normal monthly returns.[19] These 20 uncontaminated samples remain the same and are used throughout the following study.

As the aim of performing this simulation study is to assess the stability properties of the results of the MVEF model when data departs from normality, we work with uncontaminated

[18] Taking as population parameters the classical $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ estimators is another option which leads to the same conclusion. However, by working with their robust counterparts we can highlight an interesting point related to model risk and estimation risk.

[19] 20 samples ensure clarity in subsequent graphics, as we found that a larger number of samples leads to similar results. The size of 120 has been chosen to reflect an average of what is commonly seen in practice. Due to the lack of history on many indices or securities, it is indeed questionable to expect that portfolio optimization carried out in the financial reality could be based, on average, on much more data.

and *contaminated* samples. The contaminated samples are constructed by applying transformation to the returns of the uncontaminated samples. Four types of return contaminations have been studied.

1. *Substitutive contamination*: random *replacement* of a given percentage of the 120 returns by a specific value. This value is the sum of the true mean of the series and of a positive or negative multiple of the true standard deviation of the series.

2. *Additive contamination*: random *addition* of a specific value to a given percentage of the 120 returns. This value is a positive or negative multiple of the true standard deviation of the series.

3. *Multiplicative contamination*: random *multiplication* of a given percentage of the 120 returns by a specific value.

4. *Point mass multiplicative contamination*: random *multiplication* of a given percentage of the 120 returns by a specific value. The contamination occurs for each of the three series on the *same* data point(s).

The percentage of data contaminated and the multiplicative coefficients used in the four types of contamination above may vary from one series to another.[20] It should be emphasized that these contaminations are carried over independently on each of the 3 series except in the point mass multiplicative contamination case.

We suggest contaminating 3% of the returns in each of the three return series making

[20] All above-described contamination types are based on transformation of specific returns. Other types of contamination including direct transformation on the parameters $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ have not been considered as they seem less relevant for replicating the contamination possibly occurring in real market data.

the total contamination on each sample add up to 9%. We also choose 5 as the value for the multiplicative coefficient used for all series. As the fourth type of contamination is considered later, alternately applying each of the first three types of contamination to the 20 uncontaminated samples, we end up with 60 contaminated samples.

In what follows, classical estimation always refers to the use of the ML estimator, whereas robust estimation refers to the use of the TBS estimator (with breakdown point $\varepsilon^* = 0.35$ and asymptotic rejection probability $\alpha = 0.001$).[21] Classical and robust estimation of the parameters are performed to allow computation of the so-called classical and robust efficient frontiers (i.e. the frontiers calculated by means of the classically or robustly estimated parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).

4.1 Behavior of efficient frontiers

First, let's compute the 20 classical and 20 robust efficient frontiers on the 20 normal uncontaminated simulated samples. The results are shown in Figure 9.[22]

The classical and robust efficient frontiers computed on the original data set are the same as the ones in Figure 7, and are reported as reference. As anticipated, the classical and robust simulation results are very similar, as no contamination is present in the data.[23] The dispersion of both the 20 classical and 20 robust efficient frontiers indicates that even when the distributional normality assumption is not violated, sampling variability (even with series with rather low standard deviation of returns) is enough to make efficient frontiers move

[21] A breakdown point of 35% and an asymptotic rejection probability of 0.1% strike a good balance between robustness and efficiency.

[22] The efficient frontiers of Figure 9 have been computed with the constraint of no short selling.

[23] The population parameters being the robust mean and covariance estimates ($\hat{\boldsymbol{\mu}}_{\mathrm{TBS}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{TBS}}$), classical and robust efficient frontiers calculated on normal uncontaminated simulated samples cluster around the *true* robust mean-variance efficient frontier reported in bold.

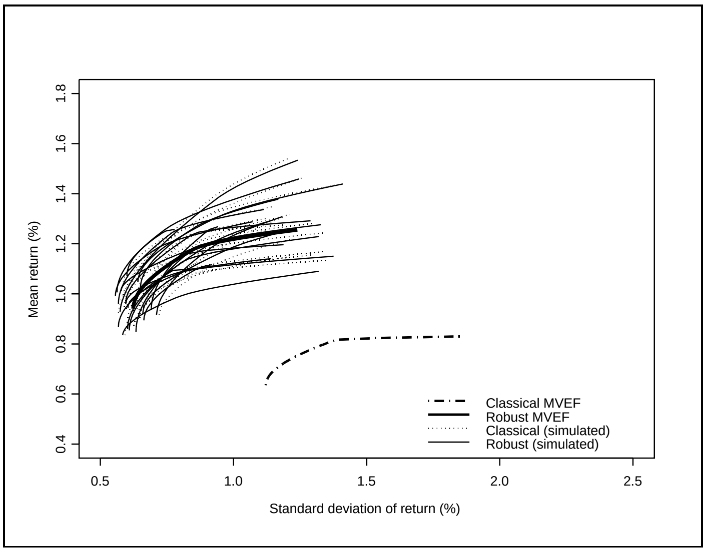

Figure 9. Classical and robust uncontaminated efficient frontiers. This figure shows 20 classical and 20 robust efficient frontiers computed on the same 20 normal uncontaminated simulated samples. Short selling is not allowed. The size of each sample is 120. The classical and robust MVEF computed on the original data set are reported as reference.

around the true one, making this situation an illustration of estimation risk. This implies that, for a given level of standard deviation of returns, the impact in terms of performance is far from being negligible when considering portfolios alternately located on each of the simulated efficient frontiers. However, this sampling variability is far smaller than the bias between classical and robust estimation on the original data represented by the classical and robust MVEF in Figure 9 .

Let's now turn on to the results when classical and robust estimation is carried over to contaminated samples. The plot of Figure 10 shows the particular results of the *substitutive contamination* type with independent contamination of 3% and a multiplicative coefficient of 5 applied to each series. As in Figure 9, classical and robust efficient frontiers computed on the original data set are reported as reference. The striking feature of Figure 10 is that

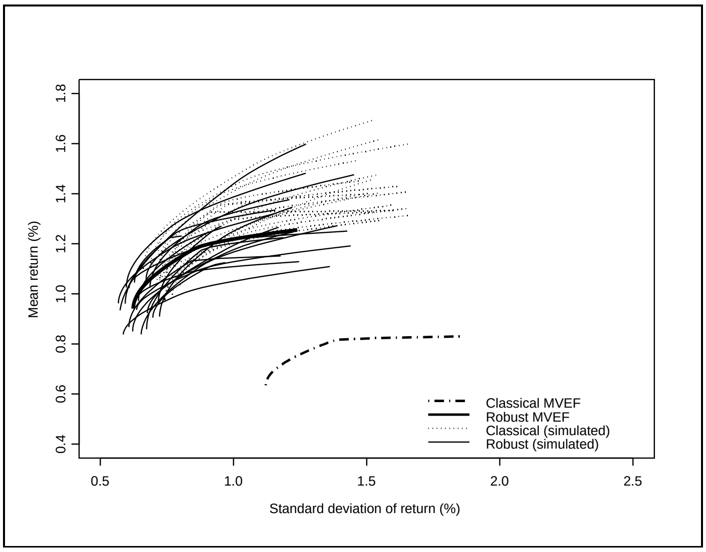

Figure 10. Substitutive contamination: classical and robust efficient frontiers. This figure shows 20 classical and 20 robust efficient frontiers computed on the same 20 contaminated simulated samples. Short selling is not allowed. The size of each sample is 120. Substitutive contamination is applied with an independent contamination of 3% on each series and a multiplicative coefficient of 5. The classical and robust MVEF computed on the original data set are reported as reference.

the 20 robust efficient frontiers are at the same position as in the previous (uncontaminated) case, while the 20 classical efficient frontiers have all been clearly shifted higher and to the right. They no longer cluster around the true efficient frontier.

This type of contamination indeed implies a lateral shift due to variance increase in the sample, and a vertical shift due to the positive multiplicative value of 5 increasing the probability of replacing returns by higher ones as each of the three original return series exhibits a positive mean. Hence, for a fixed level of return, standard deviation is overestimated. This illustrates the possibly misleading results obtained when using classical estimation to compute efficient frontiers, as the position of the latter in Figure 10 wrongly suggests a change in the true model. In fact, this bias is only caused by the bad influence of the 9% of outlying data on the estimation process.

Similar results are obtained with substitutive, additive and multiplicative contamination with the parameters mentioned above. When the multiplicative coefficient is negative, we found that efficient frontiers shift lower. And when a mix of positive and negative coefficients are used, we found that the vertical shift almost disappears and only a horizontal shift remains. In all cases however, the horizontal shift occurs to the right as the variance of the data increases due to the contamination.

The above results have been obtained with independent contamination of 3% on each series and with a multiplicative coefficient of 5. However, we may argue that this type of contamination is too high or too low, or that the choice of the multiplicative coefficient is inadequate compared with data contamination encoutered in financial reality. That's why we suggest looking again at the original data set and specifically at the plot of Figure 8 showing the gross error sensitivity for the low volatility hedge fund portfolio. As already mentioned, the gross error sensitivity clearly identifies two (trivariate) returns as having a heavy influence on the estimates of the optimal portfolio weights, namely data points 56 and 57, corresponding to the Russian financial crisis of late Summer 1998. These two data points represent 1.85% of the whole sample. We compute the median of each of the three series of returns (i.e. convertible arbitrage, event-driven and fixed income arbitrage strategies) over the whole period, as well as the median (in this case equal to the arithmetic mean) of each of the three series for these two outlying returns.[24] The ratio of the medians (by series) equals $(-4, -7, -3)$, for the convertible arbitrage, event-driven and fixed income arbitrage

[24] The median is preferred to the mean for its robustness properties.

strategies respectively. This means for instance, in the case of event-driven strategies, that the median monthly return of the two outlying data points is seven times larger (in absolute terms) than the median monthly return on the whole period.[25]

We are now able to artificially reproduce part of the contamination present in the original data set of the low volatility hedge fund portfolio, and we use the above-mentioned *point mass multiplicative contamination* to study the behavior of the efficient frontiers in that specific case. The contamination is set at 2.5% and the multiplicative coefficients are set at $(-4, -7, -3)$ for the respective strategies.[26] From now on, only this contaminated sample and the uncontaminated sample are used in the study.

The results displayed in Figure 11 are disturbing: while, as in the previous (contaminated) case, the 20 robust efficient frontiers seem unaffected by contamination, the 20 classical efficient frontiers shift sharply lower and extend to the right, to the point that they even overlap the classical efficient frontier computed on the original data set. In this case, when using classical estimation, model risk clearly dominates estimation risk, and the danger when making portfolio choices is of ending up on an efficient frontier strongly influenced by the characteristics of just a few outlying observations. Needless to say, portfolio choices made in this context may lead to sub-optimal decisions. On the other hand, robust efficient frontiers show no apparent bias, and thus more accurately represent the statistical properties of the non-outlying 97.5% of data of the 20 samples.

This case of point mass multiplicative contamination illustrates an interesting point,

[25] This shows that the previously used value of 5 was (in absolute term) a realistic choice for the multiplicative coefficient.

[26] The contamination of 2.5% is chosen to obtain exactly 3 contaminated data points among the 120 data points of each sample. It should be noticed that this contamination is smaller than that used before.

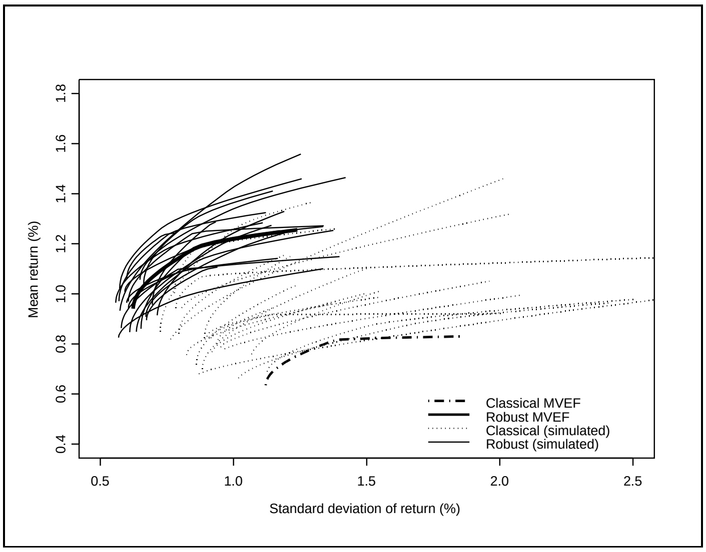

Figure 11. Point mass multiplicative contamination: classical and robust efficient frontiers. This figure shows 20 classical and 20 robust efficient frontiers computed on the same 20 contaminated simulated samples. Short selling is not allowed. The size of each sample is 120. Point mass multiplicative contamination is applied with a contamination of 2.5% and respective multiplicative coefficients equal to $(-4, -7, -3)$. The classical and robust MVEF computed on the original data set are reported as reference.

namely the non-continuous behavior of classical estimators in the presence of outlying data. Here, a mere 2.5% of outlying data has a far greater impact on the results than the previous 9% of contaminated data.

As shifts in efficient frontiers imply changes at the level of the underlying portfolios in terms of mean return, standard deviation and optimal weights of the different securities, a closer look at the portfolios themselves is also of interest.

4.2 Behavior of portfolios

Until now, we have considered portfolio optimization with a constraint of no short selling for better visual identification of shifts of efficient frontiers, and replicate what is often done in practice. We suggest lifting this restriction to be fully in line with the results of section 3, derived in the unconstrained case, and when the solution for the optimal portfolio weights is an explicit expression. We now focus on the behavior of minimum variance portfolios, but we found similar results in the case of portfolios with a given level of standard deviation of returns.

Figure 12 shows the boxplots of standard deviation of returns for minimum variance portfolios. Black diamonds represent the standard deviation of returns of the *true* minimum

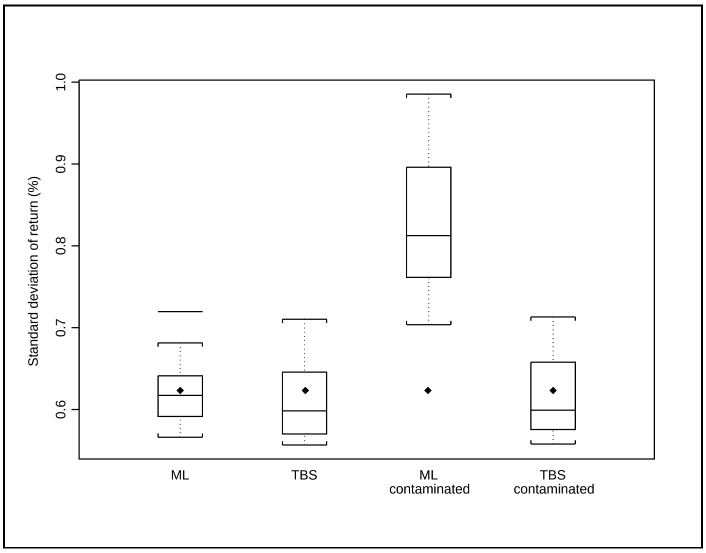

Figure 12. Minimum variance portfolios: boxplots of standard deviation of returns. This figure presents the boxplots of standard deviation of returns when classical and robust efficient frontiers are alternately computed on uncontaminated and contaminated samples. Short selling is allowed. The size of each sample is 120. Point mass multiplicative contamination is applied with a contamination of 2.5% and respective multiplicative coefficients equal to $(-4, -7, -3)$. Starting from the left, boxplots 1 and 2 show the variability in standard deviation of returns of the minimum variance portfolios when (classical) ML estimation, respectively (robust) TBS estimation is used on 20 uncontaminated samples. Boxplots 3 and 4 show the same information when estimation is carried over to 20 contaminated samples.

variance portfolio located on the true robust MVEF. The two boxplots on the left show that the characteristics of the standard deviation of returns for minimum variance portfolios resulting from 20 classical and 20 robust estimations on uncontaminated samples are very similar, with the exception of a little loss of efficiency for the robust estimation. On the contrary, under contamination, the 20 classical estimations of efficient frontiers lead to minimum variance portfolios with a large bias in standard deviation of returns (see boxplot 3), whereas the 20 robust estimations of efficient frontiers exhibit similar minimum variance portfolio characteristics (see boxplot 4) to those found in the uncontaminated case.[27] This sensitivity of results shows again the lack of robustness of the classical ML-estimated MVEF model. Just a few outlying data (here 2.5%) are enough to heavily bias the estimation and make the optimization process give misleading results.

Another way of looking at this problem is to focus on the *composition* of the minimum variance portfolios by examining the vector $\hat{\mathbf{p}}$ of estimated optimal portfolios weights. The boxplots of the minimum variance portfolio weights for the 3 hedge fund strategies under classical estimation are shown in Figure 13. While classical estimation on the 20 uncontaminated samples (from the left, boxplots 1, 3, and 5) leads to boxplots only reflecting sampling variability (i.e. estimation risk), classical estimation on the 20 contaminated samples (from the left, boxplots 2, 4 and 6) shows very significant variability (i.e. model risk) of the optimal weights within these minimum variance portfolios. Moreover, in the case of event-driven (ED) and fixed income arbitrage (FIA) strategies, the central 50% of the boxplots does not even overlap with the true weights represented by the black diamonds. Portfolio weights are

[27] We found similar results for the behavior of the *mean return* of minimum variance portfolios.

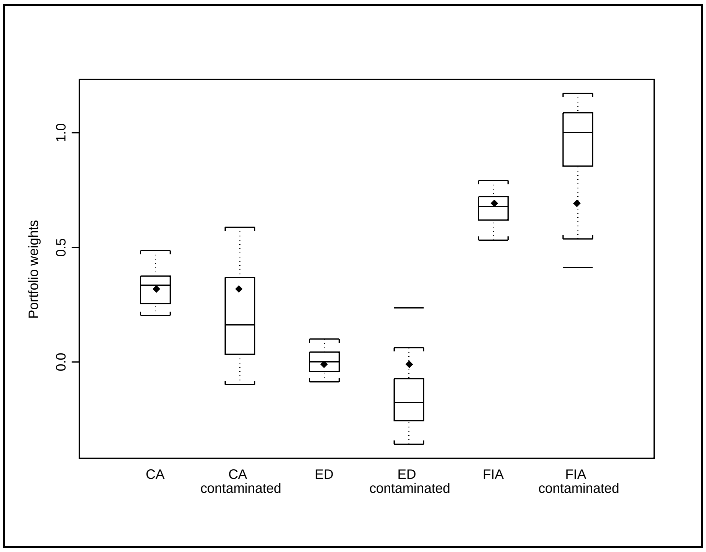

Figure 13. Minimum variance portfolio: composition under classical estimation. This figure presents the boxplots of the weights for minimum variance portfolios under classical estimation on uncontaminated and contaminated samples. Short selling is allowed. The size of each sample is 120. Point mass multiplicative contamination is applied with a contamination of 2.5% and respective multiplicative coefficients equal to $(-4, -7, -3)$. Starting from the left, boxplots 1 and 2 are each based on the 20 optimal weights of the convertible arbitrage (CA) strategy within the minimum variance portfolios located on classically estimated efficient frontiers computed on uncontaminated and contaminated samples. Boxplots 3 and 4 show the same kind of information for the event-driven (ED) strategy. Boxplots 5 and 6 do the same for the fixed income arbitrage (FIA) strategy.

heavily biased, due to the extreme sensitivity of the model to only a few outlying data when estimated in a classical way. Once again, model risk dominates estimation risk.

In the case of robust estimation, results are far more stable, as shown in Figure 14. Uncontaminated and contaminated boxplots look very similar and show the definitive advantage of using robust estimation in the case of the MVEF model. Even if both estimations may lead to portfolios behaving in a specific way according to given market conditions, the use of robust estimation makes the results of the model less sensitive to a few outlying ob-

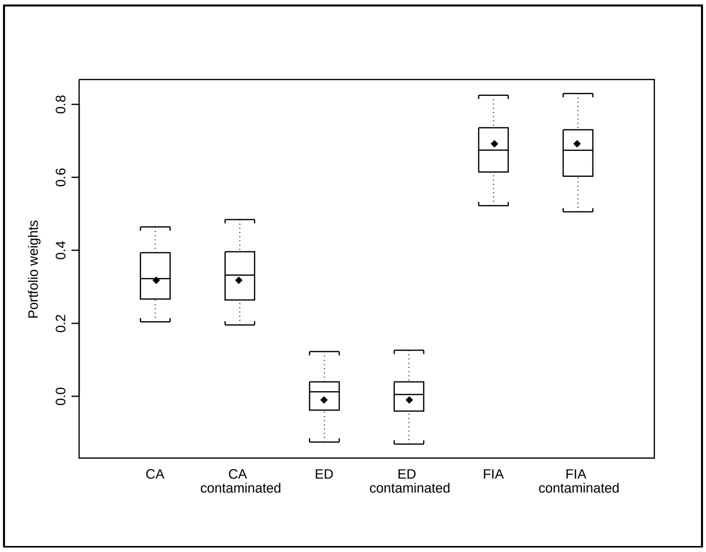

Figure 14. Minimum variance portfolio: composition under robust estimation. This figure presents the boxplots of the weights for minimum variance portfolios under robust estimation on uncontaminated and contaminated samples. Short selling is allowed. The size of each sample is 120. Point mass multiplicative contamination is applied with a contamination of 2.5% and respective multiplicative coefficients equal to $(-4, -7, -3)$. Starting from the left, boxplots 1 and 2 are each based on the 20 optimal weights of the convertible arbitrage (CA) strategy within the minimum variance portfolios located on robustly estimated efficient frontiers computed on uncontaminated and contaminated samples. Boxplots 3 and 4 show the same kind of information for the event-driven (ED) strategy. Boxplots 5 and 6 do the same for the fixed income arbitrage (FIA) strategy.

servations: model risk is kept under control. This feature is of particular importance when analyzing real data, since the presence (or absence) of just a small percentage of data from the sample shouldn't have a decisive impact on optimal portfolio choices.

5 Conclusion

We investigated the properties of the maximum likelihood estimated mean-variance portfolio optimizer and found that this model is not robust to deviations from the assumption of multivariate normality. We showed analytically that the influence function of the estimator of the optimal portfolio weights, when computed with the maximum likelihood estimator as is often the case, is unbounded, meaning that even a single outlier may take these weights beyond any predefined value.

We introduced the gross error sensitivity as a powerful diagnostic tool for detecting the specific returns from the sample that bias estimation of the optimal portfolio weights. We also highlighted that outlying observations may be characterized otherwise than by extreme returns, making them difficult to find without using this diagnostic tool. Moreover, to address the problem of non-robustness of the classical maximum likelihood estimator, we suggested replacing it by the translated-biweight S-estimator. This estimator is robust to local departures from normality and ensures that the resulting mean-variance optimal portfolios truly reflects the statistical properties of the majority of the data.

The simulation study makes clear that the classically estimated mean-variance efficient frontier model suffers from model risk when data underlying its computation are not exactly generated by a multivariate normal distribution, and that model risk dominates estimation risk.

Finally, it should be stressed that although we have here considered the multivariate normal stochastic process as generating the (majority of the) independently and identically distributed data, the same concepts of statistical robustness can be applied to more sophisticated models.

As normality is the exception rather than the rule in the financial reality, the use of robust statistics in quantitative portfolio management opens the way to fruitful research.

References

Alexander, Gordon J., and Bruce G. Resnick, 1985, More on estimation risk and simple rules for optimal portfolio selection, *Journal of Finance* 40, 125–133.

Barry, Christopher B., 1974, Portfolio analysis under uncertain means, variances, and covariances, *Journal of Finance* 29, 515–522.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation risk and optimal portfolio choice* (North Holland, Amsterdam, The Netherlands).

Black, Fisher, and Robert Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* 48, 28–43.

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

Cavadini, Fabiano, Alessandro Sbuelz, and Fabio Trojani, 2002, A simplified way of incorporating model risk, estimation risk and robustness in mean variance portfolio management, Working paper, University of Southern Switzerland.

Chopra, Vijay K., and William T. Ziemba, 1993, The effect of errors in means, variances and covariances on optimal portfolio choice, *Journal of Portfolio Management* 19, 6–11.

Hampel, Frank R., 1968, Contribution to the theory of robust estimation, Ph.D. thesis, University of California, Berkeley, CA.

Hampel, Frank R., 1974, The influence curve and its role in robust estimation, *Journal of the American Statistical Association* 69, 383–393.

Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel, 1986, *Robust statistics: The approach based on influence functions* (John Wiley & Sons, New York, N.Y.).

Huber, Peter J., 1964, Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35, 73–101.

Jobson, John D., and Robert M. Korkie, 1981, Putting Markowitz theory to work, *Journal of Portfolio Managment* 7, 70–74.

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

Lhabitant, François-Serge, 2002, *Hedge funds, myths and limits* (John Wiley & Sons, New York, N.Y.).

Markowitz, Harry M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Markowitz, Harry M., 1959, *Portfolio selection: Efficient diversification of investments* (Yale University Press, New Haven, CT).

Merton, Robert C., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51, 247–256.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Michaud, Richard O., 1989, The Markowitz optimization enigma: Is 'optimized' optimal?, *Financial Analysts Journal* 45, 31–42.

Rocke, David M., 1996, Robustness properties of S-estimators of multivariate location and shape in high dimension, *The Annals of Statistics* 24, 1327–1345.

Rousseeuw, Peter J., and Annick M. Leroy, 1987, *Robust regression and outlier detection* (John Wiley & Sons, New York, N.Y.).

Stein, Charles M., 1956, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution; in Jerzy Neyman, ed.: *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, vol. 1 (University of California Press, Berkley and Los Angeles, CA).

ter Horst, Jenke R., Frans A. de Roon, and Bas J. Werker, 2002, Incorporating estimation risk in portfolio choice, Working paper, Tilburg University, The Netherlands.

Victoria-Feser, Maria-Pia, 2000, Robust portfolio selection, Cahiers de recherche HEC 2000.14, University of Geneva, Switzerland.