CA' FOSCARI UNIVERSITY OF VENICE

Department of Economics

Master's Degree programme – Second Cycle (D.M. 270/2004) - in

Economics and Finance

Final Thesis

The mean-variance framework: is there a superior portfolio selection strategy?

Supervisor:
Prof. Marco Corazza

Graduand:
Simone Desimio
Matr. number 850773

Academic Year 2015-2016

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Corazza, for his precious advices and indications, for his patience and for all the support he gave me during the months I spent projecting and writing this thesis.

I'm also sincerely thankful to all the people and friends who, with their encouragement, insightful comments and technical assistance, not only made the completion of this work possible, but also inspired me in improving it further. I also want to thank all the mates and friends who walked alongside with me through this unforgettable academic journey, making it even more special.

Finally, how not greet my parents and family, who allow me to live such an experience, always letting me focus on the things I cared the most. I will never thank you enough for your silent support, hard work and unconditional love.

Contents

Introduction

More than 60 years has passed since Markowitz proposed the mean-variance framework for solving portfolio selection problems, providing the grounds for the rise of Modern Portfolio Theory. One of the biggest credit of Markowitz's work, has been to recognize the importance of diversification, which may help investors in achieving better combinations of risk and return, where these features are measured respectively by the standard deviation and the expected value of a certain portfolio's return.

In spite of the theoretical attractiveness of this approach, however, it has never been intensively used by practitioners, due to some drawbacks when coming to an actual application with real-world data. In particular, as it has been often reported in literature, mean-variance strategies tend to perform poorly out of the sample and, moreover, they're usually outperformed by naïve diversification rules, as the one which splits investor's wealth equally among all the available assets.

The main fallacies of the Markowitz model are related both to the estimation of the parameters required, which are represented by the vector of expected returns and by the variance-covariance matrix, and to the instability of the optimal solutions over time.

About the first issue, which is known in literature as "estimation error", empirical evidence have shown that historical estimates, that are commonly used for setting the model's parameters, are not usually affordable estimators of the return distribution moments. The challenge then, is to find "trustable" estimators, which should be able to improve the performances of mean-variance solutions, at least when transaction costs are not taken into account.

As it usually happens, in fact, even if we are able to obtain a precise estimation of the efficient frontier, mean-variance strategies remain outperformed by naïve diversification rules, when considering trading expenses. What you should, then, is to focus also in decreasing portfolios turnover in order to reduce the magnitude of transaction costs.

In the context of the present work, I will illustrate the main fallacies implied by the use of the mean-variance framework for portfolio selection problems, describing some of the most famous and reliable solutions among the ones developed in literature. In particular, I will focus on the use of shrinkage estimators, as the ones provided either by Jorion and by Ledoit and Wolf, and on the effect related to the application of both nonnegativity and turnover constraints.

The purpose of this study is to verify the capability of such methods to actually deal with estimation risk and with the instability of mean-variance combinations,

allowing for an improvement in risk-adjusted performances, both in absence and in presence of transaction costs. In addition, the analysis investigates about the existence of a superior portfolio selection strategy, which should be preferred among all the others. In order to do that, 49 different asset allocation models will be tested on 3 different datasets, containing 252 monthly observations each. The paper is organized as follows.

In section 1, I will provide an overview of the mean-variance framework, also describing some of the possible statistical tools, alternative to the mean and the variance, for measuring central tendency and financial risk.

In section 2, I will address the main fallacies of the Markowitz's approach, proposing possible ideas for the selection of the target return required by mean-variance optimization procedures, and practical solutions for dealing with both estimation risk and instability in optimal portfolios' weights.

In section 3, I will describe the datasets considered, the portfolio selection strategies being tested and the methodologies used for evaluating their performances. Results will be reported in section 4, as well as some final considerations.

Proofs of theorems and lemmas used during the analysis will be provided in appendix A, while the Matlab code created for running optimizations and performing comparisons will be made available in appendix B. Robustness checks are left to appendix C, while appendix D contains some interesting figures, which may add completeness to the present study.

1 The Markowitz optimization problem

1.1 Origins of Modern Portfolio Theory

Before the advent of the so called "Modern Portfolio Theory" (MPT), strategies for selecting assets and creating portfolios were based mainly on John Burr William's theories, who wrote, in 1938, "The Theory of Investment Value". In his work, Williams, set the theoretical framework for the creation of the Dividend Discount Model, arguing that an investor's objective should be to find undervalued assets to buy, assuming that their price will, sooner or later, converge to the true value. Before the 50's, fundamental analysis was considered the best and only approach for creating portfolios, with fund managers worried in discovering new and unexploited investment opportunities with scarce or no regard risk. For fundamental analysts, in fact, financial risk was no more than a variable to add when discounting dividends, with measures that, although allowing for an immediate preferential order, fail to represent investors' tastes and rationality.

As Harry Markowitz pointed out,*"Williams asserts that the value of a stock should be the present value of its future dividends* [...]" but, "[...] *if you're only interested in the expected value of a security, you must only be interested in the expected value of a portfolio. If you're only interested in the expected value of a portfolio, you maximize that by putting all of your money into whichever security has the greatest expected return. But that didn't make sense, because everybody knows you're not supposed to put all of your eggs into one basket*" (Markowitz 2009).[1]

The concept of diversification was nothing revolutionary or new even among people without particular experience or knowledge in the financial field. "*My ventures are not in one bottom trusted, nor to one place; nor is my whole estate upon the fortune of this present year* [...]"[2]: this has not been written by a "20th century's economist", but instead by William Shakespeare at the end of the 16th century, in his well-known "Merchant of Venice". This is not, however, the only example of diversification we can find in non-economical literature. Also in Robert Stevenson's "Treasure Island" we can find something similar, with regard to the places where the pirate John Long Silver used to hide his wealth: "*I puts it all away, some here, some there, and none too much anywheres, by reason of suspicion*".[3]

Although the blurry knowledge of the concept, however, diversification was not an issue in portfolio theory until 1952, when Harry Max Markowitz, a 25-years old student at the time, published his renowned paper "Portfolio Selection" becoming

[1] M.M. Towle (2009) "Ideas and Innovation across Multiple Disciplines: A Discussion with Nobel Laureate Harry M. Markowitz, PhD", *The Journal of Investment consulting*, Vol.10, N.1.

[2] William Shakespeare (1596-1598), "The Merchant of Venice", Act I, Scene 1, p. 2.

[3] Robert L. Stevenson (1883), "Treasure Island", Part II, Chapter 11, p. 102.

the father of the "Modern Portfolio Theory".

1.2 The Mean-Variance framework

In his work, Markowitz stated that there are two features which are common to all investors: they prefer higher returns than lower ones and they want these returns to be less uncertain and unstable as possible. With this statement, Markowitz does not assume there are not risk lovers investors trading in the market, but rather that its model answer to the needs of rational investors, who are risk averse. According to Markowitz's model, a rational investor will choose the security that he expects will offer the highest return for a given level of risk or, alternatively, the one characterized by the lowest return uncertainty for a given target return.

In its classical formulation, Markowitz chose the expected value (the mean) of the rate of return as a proxy of the expected return and the return's variance as measure of uncertainty, instability and risk. As example, let $X = x_1, x_2, ..., x_m$ be a discrete random variable describing the possible returns from a given asset, and $p = p_1, p_2, ..., p_m$ the associated probability of realization, we can define:

$$\mathrm{E}(X) = \sum_{i=1}^{m} x_i p_i, \tag{1}$$

$$\mathrm{Var}(X) = \sum_{i=1}^{m} (x_i - \mathrm{E}(X))^2 p_i. \tag{2}$$

Moving from securities to portfolios is straightforward. Defining a portfolio as a combination (or, in other words, a weighted average) of individual securities, we can easily derive the formula for the expected return and the variance of a portfolio, remembering how to compute the expected value and the variance of a weighted sum of random variables. Let R_p be the random rate of return of the portfolio, r_i be the expected rate of return of the asset i-th that has variance σ_i^2, with $i = 1, 2, ..., n$, and x_i be the percentage of wealth invested in the same i-th asset, then we can state the following:

$$\mathrm{E}(R_P) = \sum_{i=1}^{n} r_i x_i = r_P, \tag{3}$$

$$\begin{aligned} \mathrm{Var}(R_P) &= \sum_{i=1}^{n} x_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \sigma_{i,j} \\ &= \sum_{i=1}^{n} x_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \sigma_i \sigma_j \rho_{i,j} = \sigma_P^2 \end{aligned} \tag{4}$$

where $\sigma_{i,j}$ represents the covariance between the asset i-th and j-th and $\rho_{i,j}$ is the Pearson correlation coefficient.

As we can see from (4), the variance (or its square root, the standard deviation) of a portfolio is a function of the variance (or the standard deviation) of the individual securities, of the correlation between each pair of them and of the amount of wealth invested in each asset. "Other things being equal the more returns on individual securities tend to move up and down together, the less do variations in individual securities cancel out each other and hence the greater is the variability of return on the portfolio" (Markowitz, 1959)[4].

Applying Williams' ideas to a portfolio of assets (instead of applying to individual securities) we can say that, what Williams thought, was that you can focus on choosing the portfolio with the highest expected return, forgetting about the variance, because investors are able to completely get rid of it through diversification, due to the law of large numbers. As Markowitz pointed out in his "Portfolio Selection" (1952), the law of large numbers cannot be applied when considering a portfolio of securities because returns from securities are heavily inter-correlated. Markowitz, in fact, "had the brilliant insight that, while diversification would reduce risk, it would not generally eliminate it [...]" but "[...] probably the most important aspect of Markowitz's work was to show that it is not a security's own risk that is important to an investor, but rather the contribution the security makes to the variance of his entire portfolio, and that this was primarily a question of its covariance with all the other securities in his portfolio" (Rubinstein, 2002)[5].

Due to the "diversification effect" it is possible to create portfolios which dominate individual securities in a mean-variance sense, which means that investor could obtain better combinations of risk-return simply investing their wealth in different assets. We define a portfolio "efficient" if it is impossible to obtain a greater average return without incurring greater variance (or standard deviation) or, in other words, to obtain smaller variance (or standard deviation) without giving up return on the average. Rational investors, who prefer more to less and who are risk averse, will of course choose portfolios lying on the so called "efficient frontier", that is defined as the set of all the efficient portfolios and whose expression, shape and characteristics will be presented in the following pages.

The proper choice among efficient portfolios, anyway, depends on the willingness and the capability of each investor to bear risk. In the case in which safety is considered an extremely important property, expected return will be sacrificed in order to decrease uncertainty. On the other hand, in the case of a wealthy investor with lower risk-aversion, a greater level of expected return may be pursued. In

[4]Harry M. Markowitz (1959), "Portfolio Selection: Efficient Diversification of Investments", p.19.

[5]Mark Rubinstein (2002), "Markowitz's Portfolio Selection: a 50-year retrospective", *The Journal of Finance*, Vol.57 , N.3.

other words, every investors will try to maximize his own individual utility function, characterized by a particular parameter of risk aversion. We are now ready to formally describe the mean-variance optimization problem and to compute the expression for the efficient frontier.

Assume there are n available assets, with expected rates of return $\mathrm{r} = r_1, r_2, ..., r_n$ and covariances $\sigma_{i,j}$ for $i, j = 1, 2, ..., n$. A portfolio is defined as a set of weights w_i with $i = 1, 2, ..., n$ which have to sum to 1. Markowitz's solution to the portfolio selection problem starts by fixing a desired return $r_P = \pi$, which of course depends on the investor's risk aversion. He then proceeds by solving the following:

$$\begin{gathered} \min_w \sigma_P^2 \\ \text{subject to} \begin{cases} r_P = \pi \\ \sum_i w_i = 1 \end{cases} \end{gathered} \tag{5}$$

or, in matrix form:

$$\begin{gathered} \min_w \mathrm{w'Vw} \\ \text{subject to} \begin{cases} \mathrm{w'r} = \pi \\ \mathrm{w'e} = 1 \end{cases} \end{gathered} \tag{6}$$

where "w" is the vector of the portfolio weights,"V" is the variance-covariance matrix and "e" is a column vector of ones.

Given that the problem is quadratic-linear under mild assumptions on both the returns vector and on the variance-covariance matrix, we can solve it by differentiating the following Lagrangian with respect to each variable and to each Lagrangian multiplier, and by setting each equation's result equal to zero:

$$\mathrm{L} = \mathrm{w'Vw} - \lambda_1(\mathrm{w'r} - \pi) - \lambda_2(\mathrm{w'e} - 1) \tag{7}$$

From which we obtain the following first order conditions:

$$\begin{cases} \dfrac{\partial \mathrm{L}}{\partial \mathrm{w}} = 2\mathrm{w'V} - \lambda_1 \mathrm{r'} - \lambda_2 \mathrm{e'} = 0_{\mathrm{N}} \\ \dfrac{\partial \mathrm{L}}{\partial \lambda_1} = -\mathrm{w'r} + \pi = 0 \\ \dfrac{\partial \mathrm{L}}{\partial \lambda_2} = -\mathrm{w'e} + 1 = 0 \end{cases} \tag{8}$$

This lead us to the following system of equations:

$$\begin{cases} \mathrm{w'} = \dfrac{1}{2}\lambda_1 \mathrm{r'V}^{-1} + \dfrac{1}{2}\lambda_2 \mathrm{e'V}^{-1} \\ \mathrm{w'r} = \pi \\ \mathrm{w'e} = 1 \end{cases} \tag{9}$$

After substituting w′ with its formula in each equation of the system, and rearranging the results, we obtain:

$$\begin{cases} \mathrm{w}' = \frac{1}{2}\lambda_1 \mathrm{r}'\mathrm{V}^{-1} + \frac{1}{2}\lambda_2 \mathrm{e}'\mathrm{V}^{-1} \\ \frac{1}{2}\lambda_1 = \dfrac{\pi\gamma - \beta}{\alpha\gamma - \beta^2} \\ \frac{1}{2}\lambda_2 = \dfrac{\alpha - \pi\beta}{\alpha\gamma - \beta^2} \end{cases} \tag{10}$$

where $\alpha = \mathrm{r}'\mathrm{V}^{-1}\mathrm{r}$, $\beta = \mathrm{r}'\mathrm{V}^{-1}\mathrm{e}$, $\gamma = \mathrm{e}'\mathrm{V}^{-1}\mathrm{e}$. Then simply substituting $\frac{1}{2}\lambda_1$ and $\frac{1}{2}\lambda_2$ in the first expression with the respective equalities obtained from the second and from the third expression, we can derive the formula for computing the weights of an efficient portfolio (w^*) for a certain level of π:

$$\mathrm{w}^* = \frac{(\gamma\mathrm{V}^{-1}\mathrm{r} - \beta\mathrm{V}^{-1}\mathrm{e})\pi + (\alpha\mathrm{V}^{-1}\mathrm{e} - \beta\mathrm{V}^{-1}\mathrm{r})}{\alpha\gamma - \beta^2}. \tag{11}$$

We are now able to compute efficient portfolios for each value of the target return π and to draw the efficient frontier. Usually, however, the expression of the frontier it is expressed in terms of variance. As it is proved in Merton (1972)[6], simply rearranging the results in (10), we are able to obtain the following analytical expression for the efficient frontier:

$$\mathrm{wVw'} = \sigma_P^2 = \frac{\gamma\pi^2 - 2\beta\pi + \alpha}{\alpha\gamma - \beta^2}. \tag{12}$$

From this formula we could already figure out what the shape of such a frontier should be. Having the return of the portfolio (π) raised to the second power, lead the efficient frontier to have a parabolic shape, at least when the variance stays on the y-axis and the expected return stays on the x-axis, or an hyperbolic shape when we consider the standard deviation as the measure for risk. The usual representation of the efficient frontier, however, has the expected return in the y-axis and the standard-deviation (or the variance) on the x-axis, causing a 90 degrees rotation of the shapes just discussed.

From the plot in Fig.1 we can see the typical bullet shape of the efficient (and inefficient) frontier and we may also be able to identify the portfolio characterized by the lowest variance among all the available portfolios, which is called "minimum variance portfolio". How you may notice, pushing expected returns to a level that is lower than the one offered by the minimum variance portfolio, it is not possible to obtain a further reduction in standard deviation. The portion of the frontier associated with these portfolios no more represents efficient combinations of risk

[6] R.C. Merton (1972), "An Analytic Derivation of the Efficient Portfolio Frontier", *The Journal of Finance and Quantitative Analysis*, Vol.7, No.4, 1851-1872.

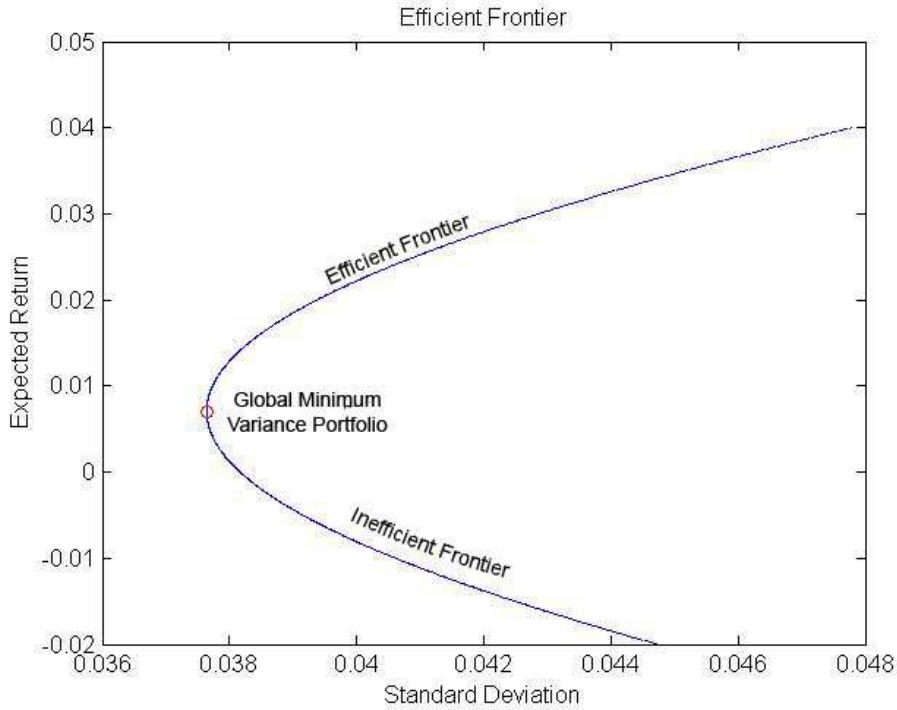

Figure 1: Example of efficient frontier

and return (it is possible to obtain an higher expected return for each level of standard deviation) and, for this reason, it is called "inefficient frontier".

It is possible to prove (see Appendix A) that the mean-variance framework is coherent with the expected utility maximization criterion if either:

- the investor's utility function is a quadratic one, i.e.

$$U(w, \mu, V, \gamma) = w'\mu - \frac{\gamma}{2} w'Vw; \tag{13}$$

- or that portfolio's return are described by an elliptical joint distribution[7], that is a family of probability distributions "characterized by the property that their equi-density surfaces are ellipsoids" (Szegö, 2005)[8].[9]

Anyway, in spite of the apparent strength of the previous conditions of coherence, it has been proven by Levy and Markowitz (1979) that a well selected point from the mean-variance efficient set "is expected to yield almost maximum expected utility if the investor's utility function is approximately quadratic, or if his a priori beliefs are approximately normal"[10]. As noted by Szegö (2005), however,

[7]See, for example, Cambinis, Huang, Simons (1981), or Owen, Rabinovitch (1983).

[8]G.Szegö (2005), Measures of risk, *European Journal of Operational Research*, 163 , 5–19.

[9]As an example remember that both the Normal and the T-distribution belong to this class of distributions, but note that the symmetry of the pdf around the mean is not a sufficient condition to define a distribution elliptical.

[10]H.Levy , H.M. Markowitz (1979), *The American Economic Review*, Vol.69, No.3, 308-317.

applying a mean-variance approach when the distribution of the returns differ significantly from an elliptical one, may lead to "sever underestimate extreme events which cause the most severe the losses".

For what concerns the model's input, Markowitz, in his 1952's work, did not assume that investors should use historical data for computing the vector of the expected returns and for setting up the variance-covariance matrix but, what he actually suggested, is to use "relevant beliefs about future performances" (Markowitz, 1952) without specifying how to build up these beliefs. "Choices based on past performances alone", he pointed out in a later work, "assume that average returns of the past are good estimates of the likely return in the future, and that the variability of returns in the past is a good measure of the uncertainty of returns in the future".[11] Even if these assumptions may seem not fully reasonable, for both academics and practitioners, historical estimates still represent the default choice for predicting the moments of the distribution of the returns. Moreover, using historical data, we are able to obtain mean-variance efficient portfolios quite easily and to test our investing strategy through time using either real or simulated data. For these reasons, in the following chapters, historical samples will be used for setting up mean-variance efficient portfolios.

1.3 Alternative measures of central tendency and financial risk

Mean and variance are not the only possible measures for central tendency and financial risk respectively. Literature is plenty of instruments that we might use for selecting portfolios. For example, the mode may be used as well as a measure of central tendency, or the semi-deviation may be considered an interesting gauge of financial risk. I think it may of interest to present e brief discussion about the most famous alternative measures that an investor may be tempted to use instead of the mean and the variance.

1.3.1 Measures of central tendency

The mean is the most used measure of central tendency in portfolio selection problems and it is the technical name for the average. The mean value of a probability distribution is called the expected value of the random variable, and its formal definition, in the case of discrete random variables, it has already been stated in (1).

The mode, on the other hand, is simply the most frequent outcome, or the one characterized by the highest probability in the case of continuous random variables. As example take a discrete random variable X, with x being the set of the possible outcomes x_i, where $i = 1, 2, ..., n$, sorted by magnitude. The mode

[11] Harry M. Markowitz (1959), "Portfolio Selection: Efficient Diversification of Investments", p.14.

will be equal to the value x_{Mo} such that:

$$P(x_{Mo}) = \max[P(x)]. \tag{14}$$

The median is the midpoint of a distribution, with half of the probability (or frequencies) associated with outcomes equal to or smaller than the median, and the other half associated with outcomes equal to or larger than the median itself. Taking the random variable X described above, we can define the median as the value x_{Me} such that:

$$P(X \leqslant x_{me}) \geqslant \frac{1}{2} \;\; ; \;\; P(X \geqslant x_{me}) \leqslant \frac{1}{2}. \tag{15}$$

In the classical Markowitz optimization problem, the mean is used as a measure of central tendency and it is usually represent on the y-axis of the when plotting efficient frontiers. Instead of the mean we may use one of these two alternatives creating, for example, mode-standard deviation efficient portfolios. Of course, using a different indicator of central tendency we may (because this is not always the case, as it will be clear later) obtain different efficient sets of portfolios with respect to the mean-variance ones. Efficient portfolios, then, are not "absolute results", but they depend upon the parameters used to assess their efficiency. The real question now is which of these measure is the best? Which one should an investor use?

One measure of central tendency (as it is true for risk) is better than another one if it is able to generate better efficient portfolios, i.e. portfolios characterized by superior out of sample performances. In some cases, however, different combinations of measures lead to the same efficient set. If the returns follow, for example, a normal distribution, a portfolio is mean-variance efficient if and only if it is also mode-variance efficient and median-variance efficient. In this case the choice of the measure is only a matter of practical and computational convenience.

Both the mode and the median, anyway, show some problems when applied to portfolio analysis. Differently from the mean, in fact, they may both fail in recognizing superior distributions when the "most likely" value, in the case of the mode, or the "central value", in the case of the median, remain the same. As an example, consider two discrete random variables:
$X = \{1, 1, 5, 7, 8\}$;
$Y = \{1, 1, 5, 7, 20\}$.
When comparing this two random variables we note that they're equal with the exception of the highest value of the distribution, that is 8, in the case of X, while it is equal to 20 in the case of Y. For this reason we can recognize Y as having a superior distribution with respect to X, even if both the mode and the median are the same for the two random variables (in this example the mode is equal to 1 and the median is equal to 5).

Another issue we may encounter when using the mode and the median as measures for central tendency, is that in a distribution it is possible to have more than one value for the mode and the median[12], while we cannot have more values for the mean. For all these reasons (and for computational advantages) the mean is usually chosen as the default measure for central tendency in portfolio analysis.

1.3.2 Measures of financial risk

After 1952 a lot of different risk measures have been developed whose detailed treatise is beyond the scope of this thesis. The most important ones when dealing with portfolio selection problems are the variance (V), the mean absolute deviation (MAD), the semi-deviation, the Value at Risk (VaR) and the Conditional Value at Risk (CVaR). The variance (or its square root) is the standard instrument used in portfolio selection problems in order to measure the risk associated to a given investment opportunity. It is, as we have already seen in (2), a measure of dispersion and it is defined as the average of the squared deviation from the mean value.
Many authors[13] have written about problems in using the variance as a measure of risk, in particular when it is applied to portfolio selection problems. Some of the most recurring issues are:

- the computational burden in estimating $n(n+1)/2$ covariances when n is particularly large;
- investors are unhappy about very low returns but they don't dislike very high returns, which means that their risk preferences are not symmetric around the mean, while variance penalizes large gains and large losses in the same way;
- generally the distribution of the returns is far from being normal, and so an investor should consider also other moments of the same distribution when taking investment decisions (i.e. skewness and kurtosis).

Of course the second and the third point are strictly related because, when considering a normal distribution of the returns, we have a symmetric shape and, when minimizing the variance, we are also minimizing risk in its most narrow (and asymmetric) sense. For this reason, in case of normal distributions, both the second and the third issue do not subsist.

Focusing on the first point, one of the most famous alternatives has been offered by Konno and Yamazaki (1991)[14], who substituted variance with mean absolute

[12] in the case of the median of a discrete random variable, it is actually a matter of how to manage the possibility not to have such a x_{Me} inside our set of possible outcomes x when, for example, this median value lies between two values x_i and x_{i+1}, which belong to the set. If we decide to compute the median as $\frac{x_i+x_{i+1}}{2}$ then it is always uniquely determined.

[13] See for example Karacabey (2005) or Konno and Yamazaki (1991).

[14] H.Konno, H. Yamazaki (1991), "Mean-Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market", *Management Science*, Vol.37, No.5, 519-531.

deviation (MAD) for measuring risk in portfolio optimization problems. MAD is defined as the average deviation from the mean in absolute terms, i.e

$$\begin{aligned} \text{MAD} &= E\left[\left|\sum_{i=1}^{n} R_i w_i - E\left[\sum_{i=1}^{n} R_i w_i\right]\right|\right] \\ &= E\left[\left|\sum_{i=1}^{n} r_i - E(R)\right|\right]. \end{aligned} \tag{16}$$

They found that MAD is able to generate portfolios similar to those which come from the Markowitz model obtaining, at the same time, an important reduction in the computational power required by the optimization problem. Money managers, in fact, do not need to calculate the covariance matrix to set up the model and, moreover, the optimization problem may be solved though linear programming, instead of the quadratic one required by the mean-variance approach. It may be interesting to note that, as it has been proved by Konno and Yamazaki (1991), when the distribution of the returns is a multivariate normal one, minimizing the MAD is equivalent to minimize the variance, and the solutions offered by the mean-MAD approach converge to the ones that we can get from the mean-variance optimization.

A possible answer to the second issue, the asymmetry of risk perception by investors, lies in using an asymmetric risk measure like the semi-variance (or its square root, the semi-deviation), that is defined as the average of the squared deviations of observed values that are lower than the mean, i.e.:

$$\text{semi-Var} = \frac{1}{n} \sum_{r_i < E(R)} (E(R) - r_i)^2. \tag{17}$$

While the variance of a random variable and of its reflection is the same, this is of course not true for the semi-variance. In the case in which all return distributions are symmetric, or have the same degree of skewness, the efficient portfolios produced using the semi-variance and the ones produced using the variance are the same. In the other cases, however, given an equal expected return and variance, if you use the semi-variance (or its square root, of course) you will choose portfolios with the greatest skewness to the right, or the lowest skewness on the left. The intuition behind this statement lies in the lower dispersion (and lower variance) that we can clearly notice, for example, on the left side (with respect to the mean) of a right-skewed Gaussian distribution, when compared with the dispersion we can find in the left side of a the standard-normal (as in Fig.2).

Moreover, as noted by Markowitz (1959), the semi-deviation may produce efficient portfolios preferable to those of the standard deviation in terms of out of sample performances but, anyway, those produced by the standard deviation are

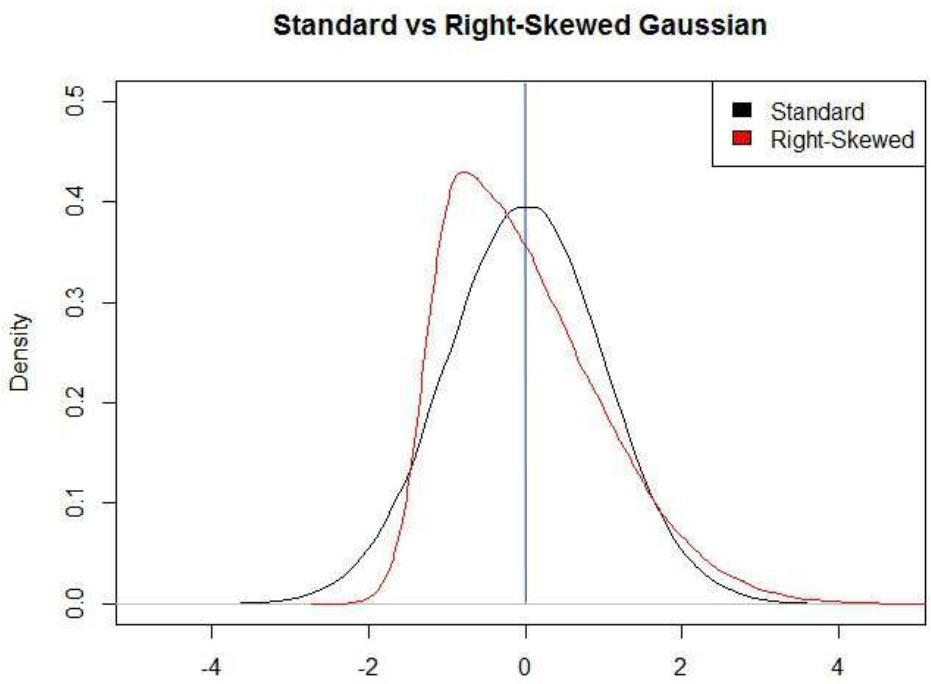

Figure 2: Comparison of the pdf$_s$ of two random samples of 10,000 observations generated respectively by a Normal(0,1) and by a Skew-Normal (0,1,2).

satisfactory, and the standard deviation itself is easier to use and interpret requiring, at the same time, lower computational power for solving the optimization problem. When using the semi-deviation the entire joint distribution of the returns must be supplied as inputs, that is far more than what is required for the usual mean-variance problem.[15] So, in order to solve the second issue, we are worsening the computational one, creating a trade-off problem. However we should remember that, in the case of portfolios characterized by very low variance (like the minimum variance ones, that will be one of the main focus of this thesis), both the higher and the lower extremes of the returns distribution are softened, resulting in portfolios with low semi-deviation too, with a solution of the previous trade-off which goes then in favor of the mean-variance optimization.

Another viable alternative for measuring risk is to use quantile-based risk measures, that are measures whose definition and computation recall, implicitly or explicitly, the notion of quantile[16]. Among this family of risk measures we find the Value at Risk (VaR) and the Conditional Value at Risk (CVaR), where the first is defined as "the worst loss that can be expected with a certain probability" (Bertsimas et. al, 2004)[17] in a certain period of time, and the latter represents

[15]Note that we cannot use the correlation coefficient when working in terms of semi-variance.

[16]Quantiles are values q_α that split the CDF of a distribution in two subsets: one greater and one lower than or equal to q_α and for which $F(q_\alpha) = \alpha$, where F represents the CDF and α is a certain probability threshold.

[17]D. Bertsimas, G. J. Lauprete, A. Samarov (2004), "Shortfall as a risk measure:properties, optimiza-

the conditional expectation of losses greater than VaR. Let X a random variable with L being the loss function associated to X ($L = -X$), we have that, for any discrete or continuous distribution:

$$P\{L \geqslant \mathrm{VaR}(\alpha)\} = \alpha \tag{18}$$

and so we can define VaR as

$$\mathrm{VaR}(\alpha) = \inf\{x : P(L > x) \leqslant \alpha\} \tag{19}$$

and the CVaR as:

$$\mathrm{CVaR}(\alpha) = \frac{\int_0^\alpha \mathrm{VaR}(u)du}{\alpha}. \tag{20}$$

CVaR is usually preferred to VaR by academics as a measure of risk[18], because it considers also what happens beyond the VaR threshold. Moreover it allows for linear programming techniques when solving the optimization problem and it is a coherent measure of risk also for non-elliptical distributions[19] while, for instance, in the case of this family of distributions, VaR fails to be subadditive (see Szegö 2005). In literature[20], however, we can find many proofs that, at least for multivariate elliptical distributions, mean-VaR and mean-CVaR portfolios are subsets of the mean-variance efficient ones, while mean-variance efficient portfolios might be considered inefficient under these two frameworks.

1.4 The Markowitz's legacy

More than 60 years after the rise of Modern Portfolio Theory (MPT) with the publication of "Portfolio Selection" in 1952, Markowitz's ideas and intuitions are still valid, used and teached. MPT's biggest credit was to show, in quantitative terms, the importance of assets' correlation and the economic value of diversification. Moreover, the establishment of MPT, has led to the creation of new financial models, constantly updated to allow for new discoveries and needs from both academicians and practitioners. One of the most important application of MPT in financial economics lies in building the foundations for the Single Index Model (Sharpe, 1963) and for the Capital Asset Pricing Model by Sharpe (1964), Lintner (1964) and Mossin (1966).

The Single Index Model may be seen as a simplified approach, with respect for example to the Markowitz's one, for explaining asset's return. Sharpe's intuition,

tion and applications, *Journal of Economic Dynamics & Control*, No.28, 1353-1381.

[18]See, for example, Szegö (2002) or Acerbi et al. (2001).

[19]Following the definition in Szegö(2005), coherent risk measures must respect the following properties: positive homogeneity, subadditivity, monotonicity and transitional invariance.

[20]See, for example, Hüllirmann (2002) or De Giorgi (2002).

just to offer a very brief overview of the model, lied in assuming that all the correlation among the return offered by different assets is due to their correlation with the market (expressed by the β). In this way, for predicting and understanding portfolio returns, instead of estimating an entire variance-covariance matrix, investors only need to estimate the correlation of their portfolio with market returns. The usual representation of the single index expression:

$$r_i = \alpha + \beta r_M + \epsilon_i \tag{21}$$

where r_i is the return from the i-th portfolio, r_M is the market return[21], β is the market factor and ϵ is the term associated to the error in the model's prediction.[22]

The CAPM, following in the footsteps of the Single Index Model in explaining assets'returns but making stronger assumptions[23], define a global equilibrium environment in which all the investors are completely diversified, have the same efficient frontier and hold the same risky portfolio (the tangency or market portfolio), with a percentage of wealth invested in the riskless asset which depends upon the investor's specific risk aversion. The expected return of an asset (it may be a single security or a portfolio) will be then determined by the risk-free rate, by the market risk premium (which is function of the the average risk aversion in the market) and by the share of market risk held by the asset (once again measures by the β).The usual representation of the CAPM formula is:

$$r_i = r_f + \beta(r_M - r_f) + \epsilon_i \tag{22}$$

being r_i the return of the i-th asset, r_f the return of the riskless asset, r_M the return of the market portfolio, β the market factor and ϵ the error associated to CAPM's estimation.

The value of Markowitz's work earned him, altogether with Miller and Sharpe, the Nobel Prize in Economic Sciences in 1990, for pioneering the research in the field of financial economics. Differently from what is commonly thought, however, Markowitz was not the first to develop the concept of diversification as we know it today, nor the first to apply mean-variance framework to portfolio selection problems. In 1940, in fact, a brilliant Italian mathematicians, Bruno De Finetti, published a paper named "Il problema dei pieni" that may be translated as "The Problem of Optimal Retention Levels in Proportional Reinsurance" (Barone, 2006)[24], which contains many ideas that, 12 years later, became milestones of the

[21]it is usually chosen equal to the return of a broad market index, belonging to the same asset class of the analyzed asset.

[22]Clearly, the expected error is assumed equal to 0.

[23]it requires, for example, that information is free and available to all investors in the market, that these investors are rational, mean-variance optimizers and that they have homogeneous expectations about the future.

[24]L. Barone (2006), "Bruno de Finetti and the Case of the Critical Line's Last Segment", *Bruno de Finetti Centenary Conference*, Accademia dei Lincei, Rome.

MPT.

The question to which De Finetti tried to answer was: "How should an insurance company behave in maximizing the profit of a portfolio of insurance contracts subject to the constraint given by the risk level it is ready to bear?". Or, in other terms: "Which contracts should it reinsure[25] and to what extent?" (Barone, 2006). This question can also be formulated in an slightly different way, in case of portfolio selection problems: "What are the optimal mean-variance weights of the portfolio that allow the insurance company to maximize profits given the default-risk constraint chosen?".

The original contributions of de Finetti has recently been acknowledged by both Mark Rubinstein (2006)[26] and Markowitz himself, who also wrote an article in 2006 titled "De Finetti Scoops Markowitz"[27] in which he analyzes both the similarities and the differences among their works. Mark Rubinstein (2006), moreover, tried to explain the reasons why the existence of De Finetti's research has remained unknown even among academicians, addressing the issue mainly to the language in which these contributions have been written, Italian, and to the separation in the Italian academic field between actuarial science experts and financial economic ones.

Anyway, even considering De Finetti a co-founder of the mean-variance approach, Markowitz still has the credit for making MPT famous and known to the public, giving asset managers a more efficient and coherent instrument for setting up financial portfolios, opening the gate to a path that still have to be completely explored.

[25]"The practice of insurers transferring portions of risk portfolios to other parties by some form of agreement in order to reduce the likelihood of having to pay a large obligation resulting from an insurance claim" (definition by Investopedia).

[26]M.Rubinstein (2006), "Bruno de Finetti and Mean-Variance Portfolio Selection Problem", *Journal of Investment Management*, Vol. 4, No. 3.

[27]H.Markowitz (2006), "De Finetti Scoops Markowitz", *Journal of Investment Management*, Vol. 4, No. 3.

2 Mean-Variance issues and alternatives

2.1 Fallacies of the Mean-variance approach

Modern Portfolio Theory represents, without any doubt, an appealing theoretical framework for describing how investors and money managers should build their portfolios. Moreover, after the recent technological progresses, which has significantly reduced the computational time needed for solving optimization procedures, it may be also considered a tempting practical tool for asset (or asset classes) selection. However, in spite of its "academic attractiveness", the mean-variance approach has not been used intensively by practitioners (see Michaud, 1989)[28] due to some shortcomings in its classical definition and interpretation, which has brought into question the usefulness of the model itself (See Frankfurter, Phillips, Seagle, 1971)[29]. In this chapter we will discussed these issues and, in particular, we will focus on the following topics:

- the choice of the target return (π) required by the mean-variance optimization procedure;
- the estimation risk related to the model's parameter (i.e. mean and variance);
- the instability of mean-variance solutions.

I will now provide a very brief description of the issues above, leaving a more detailed dissertation to the following sections.

As we have already pointed out in the previous chapter, and made explicit in expression (5) and (6), when formulating a mean-variance problem the investor has to specify a desired return (π), which plays the role of an indicator of his own risk aversion[30], and which allows the selection of a specific portfolio among the ones belonging to the efficient frontier. The need for a target return should not therefore be considered as a fallacy of the mean-variance approach, but rather as an opportunity to grant flexibility to the model itself.

For an investor, however, may be difficult to precisely specify a "coherent" target rate of return especially within an unstable financial environment, where assets estimated expected returns change significantly among different periods of time. It may be very likely, in fact, that an investor fails to identify the return associated to the level of risk that he is ready to bear, selecting a portfolio characterized by too much variance or, in the opposite case, "leaving expected return on the

[28] R.O. Michaud (1989), "The Markowitz optimization enigma: is 'optimized' optimal?", *Financial Analysts Journal*.

[29] G. M. Frankfurter, H. E. Phillips, J. P. Seagle (1971), "Portfolio Selection: The Effects of Uncertain Means, Variances, and Covariances", *The Journal of Financial and Quantitative Analysis*, Vol. 6, No. 5, pp. 1251- 1262.

[30] Of course less risk averse investors will select higher desired return. For this reason π may used as gauge of the aversion toward risk.

table".[31] Moreover, in the context of an analysis about the out-of-sample effectiveness of different portfolio selection strategies (as it is the case of this thesis), we would not be able to perform trustable tests without determining a rule for the choice of the desired return in each period of the investment horizon. Without such a rule, results would actually be subjective and dependent upon the choice of π, with better results not necessarily due to the superiority of a certain selection strategy. For all these reason, in section 2.2, some famous arbitrary rules for deciding the desired return of an optimal mean-variance portfolio will be described.

Moving to the second of the issues introduced before, i.e. estimation risk, it represents an important phenomena which has been observed since the 70's and which has been deeply debated in literature.[32] As noted by Frost and Savarino (1986), in fact, "optimal portfolio selection requires knowledge of each security's expected return, variance, and covariance with other security returns. In practice, each security's expected return, variance, and covariance with other security returns are unknown and must be estimated from available historical or subjective information. When portfolio optimization is implemented using the historical characteristics of security returns, estimation error can degrade the desirable properties of the investment portfolio that is selected."[33] Mean-variance optimization, moreover, tends to overweight assets characterized by large expected return, negative correlations and small variances which are the ones that, unfortunately, are most likely to be bear large estimation error. For this reason, traditional mean-variance optimization procedures are sometimes defined as "estimation error maximizers"[34]. Estimation error may be considered the biggest fallacy of the mean-variance approach and the main reason behind its poor out-of-sample performances, which have been repeatedly documented in literature[35]. In section 2.3, then, we will discuss some of the most famous solutions for reducing the bias caused by estimation error.

Estimation error does not only affect the out-of-sample performances of a mean-variance strategy, giving solutions to the optimization problem which are not really optimal, but it also causes portfolio weights to be very unstable over time. From a statistical perspective, we could think to this effect as related to the variance of the estimation error itself, where a greater variance of this error term leads to heavier changes in the weights of optimal portfolios. The instability of mean-variance

[31] Similarly, for an investor would also be challenging to choose the value for π which maximize his own utility function, due to the uncertainty related to the magnitude of the risk-aversion coefficient (γ). In section 2.2.2, however, a solution to the mean-variance problem based on this idea will be provided, assuming known the risk aversion coefficient.

[32] See, for example, Frankfurter (1971), Klein and Bawa (1975), Jobson and Korkie (1980), Frost and Savarino (1986).

[33] Frost and Savarino (1986), "An Empirical Bayes Approach to Efficient Portfolio Selection", *Journal of Financial and Quantitative Analysis*, Vol.21, No.3.

[34] See, e.g., Michaud (1989).

[35] See, for example, De Miguel, Garlappi, Uppal (2009), or De Miguel, Garlappi, Nogales, Uppal (2009).

portfolios, however, is not due solely to the error in estimating the model's input, but it seems to be an unavoidable feature of the mean-variance approach, related to the asset structure of the efficient frontier. As noted by Michaud (1989), in fact, even when taking a neighborhood of the optimal portfolio lying on the true efficient frontier, we can find many "statistically equivalent portfolios", with really close expected return and standard deviation, which may have a completely different asset weights. In subsequent periods then, even for very small changes in the inputs and in absence of estimation error, the usual mean-variance model may suggest us to heavily revise our portfolio, causing large transaction costs and making our investment strategy less appealing both in term of net average return and in term of transaction costs-adjusted Sharpe ratio[36]. For this reason, in section 2.4, a possible solution for dealing with the instability of the mean-variance solutions will be provided.

2.2 Setting the target return

The selection of the desired return is a key choice when solving a mean-variance optimization problem. It makes possible to obtain a unique optimal solution among all the efficient ones, which belong to the efficient frontier, taking into account the aversion toward risk of a specific investor. This aversion, in the usual representation of the mean-variance optimization problem, is no more measured by the risk aversion coefficient γ (see the formulation of a quadratic utility function given in (13)), but by the target return itself, which will be higher the larger the investor's appetite for risk. As we have already pointed out in the brief overview contained in the previous paragraph, when analyzing an asset selection strategy's performances, we need a formal rule for defining the target return in order to obtain results which are robust to any critique related to a subjective choice of π. Hereafter will be described some of the possible procedures for dealing with this issue, using the entire knowledge we can get from an historical sample.

2.2.1 Tangency Portfolios

The first alternative proposed lies in ignoring the choice of the target return, without taking into account any risk-aversion coefficient specification, but preserving the risk-return trade-off implied by the Markowitz model. We can do this by simply selecting the so called "tangency portfolio", whose definition is related to the two fund separation theorem and whose concept is fundamental when working under the CAPM framework.

Consider the case when it is possible to invest in a risk free-asset, whose return is certain and equal to r_f. With such a chance in hand, an investor does not need to worry about focusing on selecting a portfolio of risky assets which reflects his

[36] It is a risk-adjusted measure performance that may be calculated both in absence and in presence of taxes. A more formal definition will be provided in section 2.2.1.

own taste for risk. What he will actually do, following the fund separation theorem, is choosing to hold a mix of the risk-free asset and of a portfolio of only risky asset, which is known as "tangency" (also called "market portfolio" when working under the CAPM framework), with a proportion of wealth invested in this market portfolio which will then be function of the investor's risk aversion.

A useful concept when trying to define the tangency portfolio is the notion of Sharpe Ratio. This ratio, named after William Sharpe developed it in 1966[37], is maybe the most popular measure of risk-adjusted return, useful for evaluating (and for predicting) mutual funds or diversified personal portfolios' performances. It may be defined as the average return earned, in excess of the risk-free rate, per unit of total risk (measured by the portfolio's standard deviation).

$$\text{Sharpe Ratio} = \frac{r_p - r_f}{\sigma_p}. \tag{23}$$

Where r_p is the average portfolio return, r_f is the average return on the risk-free asset and σ_p is the standard deviation of the portfolio returns. Of course, the higher this ratio, the better will be our opinion about the considered portfolio's performances (or expected performances, where these quantities come from sample estimates). Under the CAPM framework, the Sharpe ratio may also be interpreted as the slope of the Capital Allocation Line, defined as the line representing the return should demand for a given level of risk.

When applying the two fund separation theorem to investment choices, what we should do is to invest a certain proportion of wealth in the riskless asset, as it has been already pointed out, putting what is left in the portfolio of only risk assets which maximize the Sharpe Ratio (i.e. the tangency or market portfolio). The tangency line associated to this risky portfolio is the one with the highest slope and, under the CAPM hypotheses, it may be also called Capital Market Line. The tangency portfolio can be graphically defined as the point of tangency between the Capital Market Line[38] and the efficient frontier (see Fig.3).

For deriving the analytical expression of the tangency portfolio, we minimize the ex-ante Sharpe Ratio of the portfolio under the constraint that the weights must sum to one, obtaining the following:[39]

$$w^T = \frac{V^{-1}(\mu - r_f \cdot e)}{e' V^{-1}(\mu - r_f \cdot e)} \tag{24}$$

In literature[40], however, sample tangency portfolios are known to perform poorly out-of-the sample and to be associated to very high levels of turnover.

[37] W.F. Sharpe (1966), "Mutual Fund Performance", *The Journal of Business*, 39, No.1, 119–138.

[38] In using this term we are not making any assumption about the validity of the CAPM, but we simply indicate the tangency line with the highest Sharpe ratio.

[39] For a more detailed description of the derivation procedure check Appendix A.

[40] See, for example, DeMiguel, Garlappi, Uppal (2009).

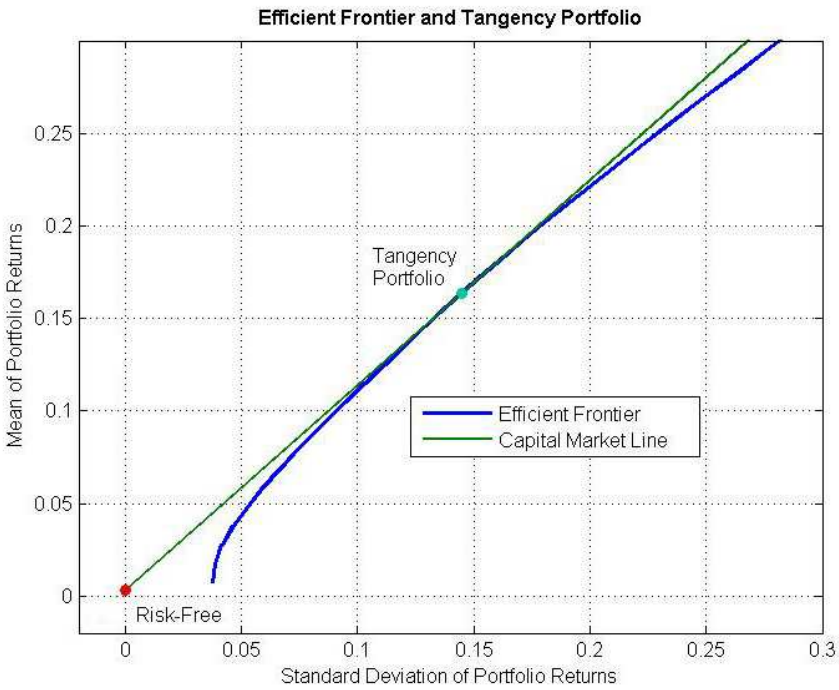

Figure 3: Efficient frontier, Tangency portfolio and Capital Market Line. It is quite clear how, combining together the riskless asset and the tangency portfolio, we obtain portfolios which dominate all the others in a mean-variance sense.

Kirby and Ostdiek[41] associate this feature to the form of the denominator of expression (24) which may be very close to zero, causing very extreme portfolio weights.

2.2.2 Utility maximization

Another possible method for obtaining a solution to the mean-variance optimization problem, without specifying on each period an ad-hoc desired return, is to directly maximize an investor's expected utility function. For reasons which has been already explained in the first chapter, the utility function here used is assumed to be quadratic, of the type described in (13). What we actually do, then, is solving the problem:

$$\begin{aligned} \max_{w} \mathrm{E(U)} = \max_{w} \quad & w'\mu - \frac{\gamma}{2} w'Vw. \\ \text{s.t. } & w'e = 1 \end{aligned} \tag{25}$$

Remembering that maximizing an equation is equivalent to minimize its negative, we proceed by differentiating the following Lagrangian:

$$L(w, \lambda) = -r'w + \gamma w'Vw - \lambda(w'e - 1) \tag{26}$$

[41] C. Kirby, B. Ostdiek (2012), "It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification", *Journal of Financial and Quantitative Analysis*, Vol.47, pp 437-467.

The solution[42] then is reached by replicating the procedure reported for the usual Markowitz optimization in (8) and (9) for the classical mean-variance optimization problem, and what we obtain is:

$$w^U = \frac{1}{\gamma} V^{-1} \left(r - \frac{e'V^{-1}r - \gamma}{e'V^{-1}e} \, e \right). \tag{27}$$

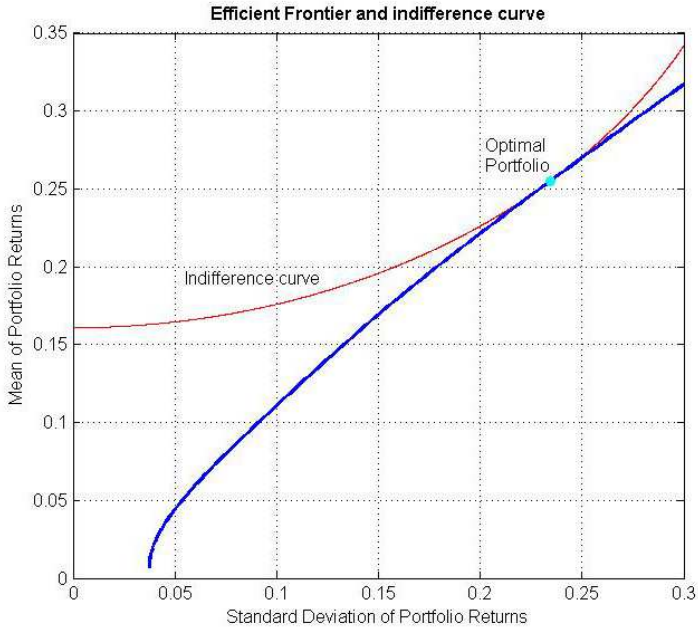

Figure 4: Indifference curve, efficient frontier and the choice of the optimal portfolio.

When plotting all the possible mean-variance combinations which have the same expected utility what we obtain is an indifference curve. The optimal portfolio will then be an efficient one, which will also lie on the highest indifference curve. Graphically, this portfolio represents the tangency point between the mean-variance efficient frontier and the highest indifference curve (see Fig.4), and it is optimal because it is expected to provide the largest utility level for a given coefficient of risk aversion.

There is, however, a well known drawback in using a direct utility maximization procedure for computing mean-variance efficient solutions, which lies in the choice of the risk aversion parameter γ. The size of this quantity is, in fact, still an empirical issue, and its estimates seems to vary with respect of the specific environment considered.[43]. Just to give an idea of the possible range for γ, you

[42]See Chapados, Bengio (2001).
[43]See, e.g., Bucciol and Miniaci (2007)

may notice that Hansen and Singleton (1983)[44] estimated it lying between 0 and 2, while Farber (1978)[45] obtained evidence of γ at least greater than 2.5. More recent researches, like the one provided by Bucciol and Miniaci (2007)[46] and by Chen et al. (2010)[47], have estimated γ to be very close to 5.

2.2.3 The maximum between $r_{1/N}$ and r_{min}

Another viable alternative for selecting a mean-variance optimal portfolio coherently with our purposes, is to fix the desired return to a level that is equal to the one of other famous portfolios which are uniquely defined on each period of time. An interesting solution used by Kourtis (2015)[48], is to set the target return equal to the maximum between the return expected from the equally weighted portfolio and one expected from the global minimum variance portfolio, i.e.

$$\pi = \max(r_{1/N};\ r_{min}). \tag{28}$$

The choice of these two "target portfolios", the equally weighted and the global minimum variance one, it is not due only to methodological needs but also it is also related to efficiency reasons. As it will be explained in section 2.3, in fact, both these two portfolios have interesting properties that an investor may be tempted to exploit. Moreover, if the equally weighted portfolio constitutes a widely used benchmark for assessing asset selection strategies performances, and so it may appear reasonable to set the desired return equal to the one we expect from this benchmark, adding the comparison with the expected return from the global minimum variance portfolio give us the chance to obtain a further reduction of risk without incurring in any loss in the expected reward. In other words, we will invest in the global minimum variance portfolio if and only if this portfolio dominates the sample mean-variance one with $\pi = r_{1/N}$.

2.3 Fighting estimation error

As it has been briefly explained at the beginning of this chapter, traditional mean-variance procedures often lead to "financially irrelevant or false optimal portfolios and asset allocation" (Michaud, 1989), due to the error implied in the estimates of both the mean and the variance. Gains from a proper diversification,

[44] L.P. Hansen, K.J. Singleton (1983), "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Stock Returns", *Journal of Political Economy*, Vol. 91(2), pp. 249-265.

[45] H.S. Farber (1978), "Individual Preferences and Union Wage Determination: the Case of the United Mine Workers", *Journal of Political Economy*, Vol. 86(5), pp. 923-942.

[46] A. Bucciol, R. Miniaci (2007), "Risk aversion and portfolio choice", *Working paper*.

[47] W. Chen, H. Chung, K. Ho, T. Hsu (2010), "Portfolio optimization models and mean-variance spanning tests", *Handbook of Quantitative Finance and Risk Management*, ch. 10, pp. 165-184.

[48] A. Kourtis (2015), "A Stability Approach to Mean-Variance Optimization", *The Financial Review*, 50, pp.301-330.

in fact, are often completely offset by estimation error, which represents the main reason for which mean-variance portfolios, although achieving resounding success among academics, perform so badly out of the sample, especially in presence of transaction costs. I will now give a more rigorous representations of the estimation error phenomenon.

In a "certainty equivalence framework" (Jorion, 1986)[49], we assume a vector of parameters θ (representing both the expected return and the variance) to equal their estimate $\hat{\theta}(y)$, whose value depends upon the set of observations y. In absence of estimation error, an investor's objective function may be defined simply as:

$$\max_q = \mathrm{E}_y[\mathrm{U}(r_p)] \mid \theta = \hat{\theta}] \tag{29}$$

where q is the vector of the optimal portfolio's weights and r_p is the expected return of such a portfolio ($r_p = q'r$). In presence of estimation error, however, the maximization problem must be described in terms of portfolio's unconditional expected utility, i.e.:

$$\max_q = \mathrm{E}_\theta \left[\mathrm{E}_{y|\hat{\theta}}[\mathrm{U}(r_p) \mid \theta] \right]. \tag{30}$$

When the distribution moments $\theta = (\mu, V)$ are assumed known, (29) and (30) coincide, and the portfolio chosen by the investor running these optimization procedures is actually optimal, with an expected utility function as:

$$F(q^*(\theta) \mid \mu, V) = F(q^{*\prime}\mu, q^{*\prime}Vq^*) = F_{MAX}. \tag{31}$$

In reality, however, the parameters θ are unknown and must be estimated. The optimization procedure will then be based upon estimated parameter, i.e. the sample estimates $\theta\hat{(y)}$ and, due to estimation error, the true expected utility will necessarily be lower than (or at least equal to) the one defined in absence of estimation error:

$$F(\hat{q}(\hat{\theta}(y)) \mid \mu, V) = F(\hat{q}'\mu, \hat{q}'V\hat{q}) \leqslant F_{MAX}. \tag{32}$$

Figure 5 gives a graphical representation of the estimation error. Running a mean-variance optimization, an investor is willing to select portfolio B, lying on the estimated efficient frontier (black dashed line) and on the highest indifference curve (red dashed linen). Due to estimation error, however, he fails to identify his true optimal portfolio, which is represented by portfolio A, that lies on the true efficient frontier (black solid line) and which is associated to the highest achievable utility level F_{MAX} (higher red solid line). What he really select, then, is not portfolio B,

[49] P. Jorion (1986), "Bayes-Stein Estimation for Portfolio Analysis", *The Journal of Financial and Quantitative Analysis*, Vol.21, No.3, pp. 279-292.

but portfolio C, that is not efficient and which is characterized by a utility level $F(\hat{q})$.

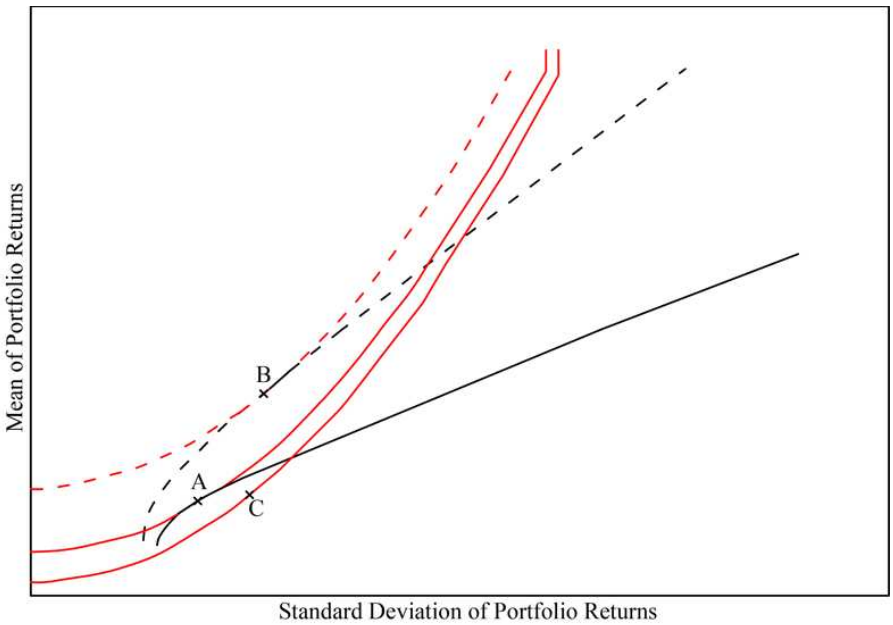

Figure 5: Estimation error in the efficient frontier; red lines represents indifference curve, black lines represent efficient frontier; dashed lines are drawn using estimated quantities, while solid lines are created using true data.

Jobson and Korkie, already in 1980, have quantified estimation error's magnitude in mean-variance optimizers, using the tangency portfolio as the solution of the optimization problem.[50] More recently, other authors like Ledoit and Wolf (2003)[51], DeMiguel, Garlappi and Uppal (2009)[52], and Kourtis (2015) have compared different sample based mean-variance strategies' performances with the ones produced by equally weighted portfolios. What they have actually found, is that, at least when no corrections for estimation risk is considered, the latter are able to produce better results[53], both in absence and in presence of transaction costs.

Before discussing possible mean-variance alternatives for dealing with estimation error, it may be useful to describe the attracting features of the equally weighted portfolio, explaining the reasons why such a simple strategy often outperforms more complex and theoretically appealing selection rules.

[50]They used a 60-months sample, assuming a multivariate normal distribution of monthly returns, following the findings in Fama (1976), and computing their results through a Monte-Carlo simulations for 20 stocks listed in the New York Stock Exchange (NYSE).

[51]O. Ledoit, M. Wolf (2003), "Improved Estimation of the Covariance Matrix of Stock Returns", *Journal of Empirical Finance*, Vol.10, Issue 5, pp. 603-621.

[52]V. DeMiguel, L. Garlappi, R. Uppal (2009) "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy", *Review of Financial Studies*, 22, 1915–1953.

[53]Usually measured in terms of Sharpe Ratio.

2.3.1 The 1/N benchmark

The equally weighted portfolio is a combination of assets in which a fraction 1/N of wealth is allocated to all the N available assets[54]. Given the simplicity of such an investment model, literature[55] often refers to it as "naïve diversification rule". There are at least three main reasons for using this rule as a benchmark:

- it is quite easy to implement because there is no need to estimate any parameter or to run any kind of optimization procedure;
- investors still use this way of allocating wealth across assets and funds;[56]
- in spite of its simplicity, it has also been documented to be capable of generating appreciable performances.

Which are anyway the features which make 1/N performances so attractive? In which cases and for which reasons we should expect it to outperform sample based mean-variance strategies? To answer these question let's try to identify the most appealing characteristics of this simple diversification rule:

- it avoids large concentrations in the same asset, even more of what is usually done by mean-variance strategies, which tends to overweight assets with extreme (positive as well as negative) returns, negative correlations and low variance;
- it is a buy-low sell-high strategy because, at each rebalancing period, assets which have increased in price are sold, while assets whose price has decreased during the investment period are bought. This process allows the recover of the original portfolio's asset structure, in which wealth is allocated equally among all the available assets;
- it never performs worse than the worst asset, because the 1/N portfolio is a long-only diversified portfolio;
- it always invests in the asset which performs best;
- given the tendency of small-firm stocks to perform better than their highly capitalized counterparts[57], 1/N catches a size premium (usually called α) investing equally in small and in big companies. This premium, however, may be quite small when limiting the investment universe to firm with large market capitalization, as the ones contained in main developed countries indexes;
- it is usually characterized by low levels of turnover, due to the small amount of trading required for recovering the 1/N structure; this characteristic may be quite important when considering transaction costs.

[54] For this reason it is usually known as 1/N portfolio, or simply 1/N.
[55] See, for example, Ledoit and Wolf (2003, 2004) or DeMiguel et al. (2009).
[56] see Benartzi and Thaler (2001), and Huberman and Jiang (2006).
[57] This phenomenon has been widely recognized in literature. See for example Fama and French (1992)

Even recognizing the attractiveness of these features, it may still seem puzzling that a model which ignore all the information available about the investment universe, can perform better than others which try to interpret and use such knowledge for obtaining more efficient results from a risk-return perspective. This strange finding is mainly due to estimation error. There are some cases, actually, in which we can expect the magnitude of this error to be quite large, and the 1/N rule to outperform usual mean-variance optimization. The main conditions which often imply large estimation error, and which have been documented by authors like Merton (1980)[58], Jagannathan and Ma (2002)[59] and DeMiguel, Garlappi and Uppal (2009), usually refer to:

- small estimation windows (usually defined with the letter "M");
- high number of available assets ("N");

A small estimation window, clearly, causes our estimate to be less precise because they're based on a small, and not always trustable, sample. When N is large, moreover, the number parameters we have to estimate increases[60], reducing the degrees of freedom at our disposal and decreasing further the accuracy of our estimates, which become more exposed to estimation risk[61]. Finally, even when the number of available assets is small, if these assets are themselves portfolios composed by individual securities, 1/N rule may outperform sample-based strategies. As noted by DeMiguel, Garlappi and Uppal (2009), given that "diversified portfolios have lower idiosyncratic volatility[62]than individual assets, the loss from naive as opposed to optimal diversification is much smaller when allocating wealth across portfolios".

2.3.2 Mean-Variance alternatives

In this paragraph some possible mean-variance alternatives for dealing with the estimation error fallacy will be proposed, i.e. strategies which are known in literature to perform well out-of-sample with respect to the 1/N benchmark.[63] These strategies vary from ignoring the return means, which are known to bear most of the estimation error[64], focusing on the global minimum variance portfolio, to the use of "shrinkage estimators" of the return distribution moments or to constrain portfolio weights to be nonnegative. The latter method, although its simplicity and popularity, will be discussed at the end of the chapter due to its applicability

[58]R. C. Merton (1980), "On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics*, Vol. 8, Issue 4, pp. 323–361.

[59]R. Jannagathan, T. Ma (2002), "Risk Reduction In Large Portfolios: Why Imposing The Wrong Constraints Helps", *Journal of Finance*, vol. 58(4), pp. 1651-1684.

[60]Notice that, as we have pointed out in the first chapter, mean-variance optimization requires the estimation of N expected returns and N(N+1)/2 covariances.

[61]See, e.g., Jannagathan and Ma (2002).

[62]It is defined as the portion of the volatility of a certain assets which is not explained by its correlation with the market. Through diversification we may be able to eliminate it almost completely.

[63]At least when not considering transaction costs.

[64]See, e.g., Merton (1980).

to all the previous strategies, even if its effectiveness is known to vary significantly among different mean-variance alternatives.[65]

2.3.2.1 The global minimum variance portfolio

In the first chapter we introduced the notion of "global minimum variance portfolio" and we have defined it is defined as the risky portfolio associated to the lowest level of variance. Minimum variance strategies have recently become quite popular in literature[66] when discussing asset allocation problems and dealing with estimation risk. But where does this popularity come from?

Following Coqueret (2015)[67] there are three possible explanations for such a success:

1. the minimum variance portfolio is not affected by the estimation error related to the expected returns;
2. it has been shown that, often, low volatility stocks do not perform worse than their high volatility counterparts;
3. investors' preferences have moved toward less risky investment solutions as an effect of the 2007-2008's financial crisis;

With regard to the first issue, clearly, reducing the number of estimated parameters, may help in reducing estimation risk. Recent works, moreover, have proved that it is far more difficult to estimate expected returns than variance-covariance matrices. Cho (2003)[68] identifies two main reasons for which second moment estimates are believed to be more accurate. First of all, when considering an i.i.d. normal distribution for the returns, the variance of the second moment gets smaller as the sample size grows. Just to give an idea of the difference, as it is reported by Cho (2003), the confidence interval for the mean of such a distribution is about 40% wider than the standard deviation's one. The second reason is related work of Merton (1980), who showed that, assuming a continuous time geometric Brownian motion, it is possible to obtain precise estimates for the variance as the sampling time intervals approach zero.

The controversial fact that low volatility stocks do not always underperform high volatility ones, outperforming them in risk-adjusted terms[69] is currently a deeply discussed topic among both academics and practitioners, and it is known

[65] See, e.g., Jannagathan and Ma (2002) or DeMiguel et al. (2009).

[66] See, for example, Ledoit and Wolf (2001, 2003, 2004), Jannagathan and Ma (2002) or Coqueret (2015).

[67] G. Coqueret (2015), "Diversified minimum-variance portfolios", *Annals of Finance*, Volume 11, Issue 2, pp 221-241.

[68] D. D. Cho (2003), "Uncertainty in Second Moments: Implications for Portfolio Allocation", *Working Paper*, SUNY at Buffalo.

[69] See, for example, Frazzini and Pedersen (2011),Van Vliet et al. (2011) or Hsu and Li (2013).

in literature as "low-volatility anomaly". This apparent market inefficiency, as well as the growth in the average aversion toward risk after the recent financial crisis, have further increased the interest in minimum-variance strategies, making them appealing to different kind of investors.

The procedure for computing the weights for the minimum-variance portfolio is quite simple and it may be performed in two different ways. The first one is running a portfolio optimization similar to the one we have used in (5) and (6) for the usual mean-variance model, but without considering the desired return constraint:

$$\min_w \sigma_P^2 \ ; \ \text{s.t.} \sum_i w_i = 1 \tag{33}$$

or, in matrix form:

$$\min_w w'Vw \ ; \ \text{s.t.} \ w'e = 1. \tag{34}$$

We then proceed as in the classical Markowitz case, differentiating the Lagrangian and posing its partial derivatives equal to zero. Solving the resulting system for w, we obtain the expression for computing the global minimum variance portfolio:

$$w^{MIN} = \frac{V^{-1}e}{e'V^{-1}e} = \frac{V^{-1}e}{\gamma}. \tag{35}$$

A second approach is to think about the minimum variance portfolio as the vertex of the parabola described by the efficient frontier in the mean-variance plane. The coordinates of such a point are:

$$\text{Minimum Variance Portfolio} = \left(\frac{1}{\gamma}; \frac{\beta}{\gamma}\right) \tag{36}$$

where β and γ are defined as in (10), with $\frac{1}{\gamma}$ being the estimated variance of the minimum variance portfolio and $\frac{\beta}{\gamma}$ being its expected return. Once having such information at our disposal, obtaining the expression for the weights of the minimum variance portfolio is straightforward. What we should do, is simply substituting $\frac{\beta}{\gamma}$ in the efficient frontier's expression defined in (12), recovering an expression equivalent to the one defined in (35).

2.3.2.2 The "shrinkage" idea

One of the most popular solution to deal with estimation error in portfolio selection is to use the so called "shrinkage estimators", which are obtained by "shrinking" the usual sample estimator toward a certain target. The problem with

sample estimators is that, even if they're unbiased[70], they are often characterized by such a large variance that cause their estimates to differ significantly from the true values. On the other side, shrinkage targets, even being usually biased, contain less variance than sample estimators. The advantage in using shrinkage estimators then, is to offer an "halfway solution" to the estimation problem, combining together a virtually unbiased estimator with large variance, with another estimator which is known to be biased, but which shows a lower dispersion.

Making an analogy with portfolio selection problems, we could say that, choosing a shrinkage estimator, we are differentiating our assets. In particular, making such a choice, we are giving up return (which we can see as related to bias) in order to obtain a reduction in risk (i.e. estimation risk). Coherently with the very foundations of MPT, then, we can presume that, due do the diversification effect, we may be able to achieve superior solutions to the risk-reward trade-off (just as for the bias-estimation risk one), resulting in more efficient portfolios, i.e. better estimators.

"Shrinkage" is not a recent concept in Statistics. It has been introduced by Charles Stein in 1956[71] and it has then been developed further by Stein himself altogether with James in 1961[72]. In 1977, Efron and Morris discussed the so called "Stein's Paradox"[73] and its application to real-life examples, as predicting the batting average for baseball players.[74] What they proved, is that when considering three or more players, a James-Stein shrinkage estimator may lead to an improvement in the forecasting precision with respect to the usual historical average. In particular, they used an estimator of the form:

$$z = \bar{y} + c(y - \bar{y}) \tag{37}$$

where y is the sample batting average of a certain player, $\bar{y}$ is the so called "grand mean", defined as the average of all the players' averages, and c represents the shrinking factor, which they set equal to:

$$c = 1 - \frac{(k-3)\sigma^2}{\sum(y-\bar{y})^2}. \tag{38}$$

where k represents the number of unknown means, σ^2 is the variance of a certain player's performances and $\sum(y-\bar{y})^2$ is the sum of squared deviations of individ-

[70]Meaning that their expected values equal the true parameter.

[71]C. Stein (1956), "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability, University of California, Berkeley*, J. Neyman ed., Volume I , pp. 197–206.

[72]W. James, C. Stein (1961), "Estimation with quadratic loss", *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, J.Neyman ed., Vol.1, pp. 361–379.

[73]Where the "paradox" was that there might exist better estimator than the arithmetic average for predicting a future outcome.

[74]B. Efron, C. N. Morris (1977), "Stein's Paradox in Statistics", *Scientific American*, 236(5), pp. 119-127.

ual averages from the grand mean. Of course, if c equals 1, James-Stein estimator converges to the sample mean, while, for c equal that approaches 0, the shrinkage estimator approaches the gran mean $\bar{y}$.

Intuitively, if the means are close to the grand mean, the estimates are shrunk further toward the grand average, otherwise not much shrinking is done. From the numerator we note that, the larger the number of football players (elements of the sample), the higher the estimation risk, and the more the estimate will be shrunk. Moreover, the greater the dispersion in the player's performances, the less informative will be the sample mean, and the stronger will be the shrinking intensity.

The gran mean used by Efron and Morris is only one of the possible shrinkage target. A good target should fulfill two requirements: it must involves only a small number of "free parameters" (Ledoit and Wolf, 2004), in order to allow for a precise estimation, but it should also reflect characteristic features of the estimated quantity, in such a way not have contain too much specification error.

For finding applications of shrinkage estimators to portfolio management, however, we have to wait until the 80's, thanks to the works of Jorion (1985)[75] and Frost and Savarino (1986). More recently, Ledoit and Wolf (2003, 2004) have developed and applied shrinkage estimators with particular regard to the variance-covariance matrix of the global minimum variance portfolio.

2.3.2.3 Jorion's "shrinked" portfolio

As introduced earlier, Jorion has been one of the first to present applications of shrinkage estimation to portfolio selection problems. Until that time, in fact, portfolio choice has been analyzed in the so called "certainty equivalence" framework, in which the underlying moments (i.e. the mean and the variance in a Markowitz's framework) are assumed known (Jorion, 1986). This assumption, however, has revealed its fallacies in the poor out-of-sample performances of traditional optimization procedures, based on usual sample estimates.

Coherently with Stein's "shrinking idea", Jorion's estimator for the exected return of a portfolio has the following structure:

$$\mu(\hat{\delta}) = \hat{\delta}\hat{\mu}_0 + (1 - \hat{\delta})\bar{\mu} \tag{39}$$

where $\bar{\mu}$ is the usual sample average,$\hat{\mu}_0$ represents the shrinkage target and $\hat{\delta}$ stays for the shrinkage intensity. Instead of assuming μ_0 and δ known or fixed a priori, as in a "strict Bayesian approach" (Jorion, 1985) like the ones proposed by Zellner

[75] P. Jorion (1985), "International Portfolio Diversification with Estimation Risk", *The Journal of Business*, Vol.58, No.3, pp. 259-278.

and Chetty (1965)[76] or Klein and Bawa (1976)[77], Jorion works under an "empirical Bayesian framework", estimating them directly from the data. More in detail, Jorion set the global minimum variance portfolio as the shrinkage target; $\hat{\mu}_0$ is then defined as:

$$\hat{\mu}_0 = w'_{MIN} r = \frac{e'V^{-1}}{e'V^{-1}e} r. \tag{40}$$

More complicated is the procedure for choosing the shrinkage intensity $\hat{\delta}$, being its value dependent upon the precision of the prior distribution, which must be estimated from the data. Leaving the issues about the derivation of such a procedure in Appendix A, the expression that Jorion's use for computing the shrinkage intensity is the following:

$$\hat{\delta} = \frac{N+2}{(N+2) + M(\bar{\mu} - \hat{\mu}_0)'\hat{V}^{-1}(\bar{\mu} - \hat{\mu}_0)} \tag{41}$$

where $\hat{V}$ is the estimator for the variance-covariance matrix provided by Zellner and Chetty (1965), which is defined as:

$$\hat{V} = \frac{M-1}{M-N-2} S \tag{42}$$

with S being the usual unbiased estimator of the variance-covariance matrix defined in (4).

Shrinkage factor, then, increases with the number of available assets (N), decrease as the length of the estimation (M) window gets larger and as the dispersion of sample means $\bar{\mu}$ from the shrinkage target μ_0 increases. These features seem quite reasonable when analyzing their relation with estimation risk. As the number of assets increases, in fact, we have more parameters to be estimated and we are more exposed to estimation risk. In this case, then, shrinkage becomes more appealing and the shrinkage intensity grows. For long estimation windows, on the other hand, parameters are estimated with more precision, and the need for strategies which are known to deal with estimation error become less impelling. Finally, coherently with the interpretation given for (38), where observed averages are not well approximated by the target-mean, shrinkage could lead to a not precise estimators and so it may seem correct to use a weaker shrinkage intensity.

Bayesian estimation provided by Jorion, however, is not limited to the first moment of the returns distribution. In practice, also the variance is unknown

[76] A. Zelner, V. K. Chetty (1965), "Prediction and Decision Problems in Regression Models from a Bayesian Point of View", *Journal of the American Statistical Association*, vol.60, issue 310, pp. 608-661.

[77] R. Klein, V. S. Bawa (1976), "The effect of estimation risk on optimal portfolio choice", *Journal of Financial Economics*, vol. 3, issue 3, pp. 215-231.

and it must be estimated. Leaving again computational issues to Appendix A, we define Jorion's estimator for the variance-covariance matrix as:

$$\hat{V}^J = \left(1 + \frac{1}{M + \hat{\lambda}}\right)\hat{V} + \frac{\hat{\lambda}}{M(M + 1 + \hat{\lambda})}\frac{ee'}{e'\hat{V}^{-1}e} \tag{43}$$

where $\hat{\lambda}$ is:

$$\hat{\lambda} = \frac{M\hat{\delta}}{1 - \hat{\delta}} \tag{44}$$

We can guess some indications from the expression in (43) analyzing some of its components. We can easily note, for example, that Jorion's estimator for the variance is increasing in the sample variance[78] ($\hat{V}$), or that uncertainty in the measure of the sample average ($\bar{\mu}$) affects it negatively.

The weights of the portfolio generated by Jorion's strategy for dealing with estimation are computed as in the usual mean-variance optimization case, where the optimal portfolio may be selected choosing, for example, one of the strategies proposed in section 2.2. The only difference between Markowitz's solution and Jorion's one, lies actually in the estimators used for assessing the expected return and the variance of the available assets.

2.3.2.4 Ledoit and Wolf's minimum variance portfolio

Another possible solution for dealing with estimation error in mean-optimizations has been proposed by Ledoit and Wolf (2003, 2004) and it is, once again, related to Stein's "shrinkage idea". Differently from Jorion's application that we have described in the previous paragraph, Ledoit and Wolf ignore return sample means, focusing only on the estimation of the variance-covariance matrix.

As it has already been explained in section 2.3.2.1, it is easier to obtain a reasonably precise estimate of the variance-covariance matrix than what it is possible to do for the vector of the expected returns, which are known to contain the largest portion of estimation error. Moreover, given the good performances of minimum variance portfolios that has often been reported in literature[79], ignoring expected returns seems to be a good starting point for obtaining efficient portfolio selection strategies. We can try, anyway, to improve further the precision of our estimates by using more efficient estimators for the variance.

Despite being unbiased, in fact, sample covariance matrices contain a lot of estimation error when the number of observations (M) is close to (or smaller than)

[78] From (42), we have that $\hat{V}$ is positively correlated with the sample variance-covariance matrix.
[79] See section 2.3.2.1 when talking about "low-volatility anomaly".

the number of available assets (N), as it is usually the case in real-life financial applications[80], at least when considering non-daily return. As it is true for the mean, most extreme (very low as well as very high) coefficients in the variance-covariance matrix often contain extreme amount of error. It may be an interesting idea, then, to use shrinkage strategies like the ones provided for the mean by Efron and Morris (1977) or Jorion (1985). The main issues, now, are toward which target to shrink and with which intensity.

A possible answer to these question is provided by the work of Ledoit and Wolf (2003, 2004) which considered a shrinkage estimator for the variance-covariance matrix based on a linear combination of the sample covariance (S) and a highly structured estimator, denoted by F.

$$V_{LW} = \delta F + (1 - \delta)S \tag{45}$$

where, once again, δ represents the shrinkage intensity, whose value lies between 0 and 1.

Ledoit and Wolf make an interesting choice for the shrinkage target F, setting it equal to an estimator from a well known 1-factor model: Sharpe's Single Index Model (SIM). We have briefly described this theoretical framework in section 1.4 but, in this context, you may appreciate how it represents a gauge shrinkage due to its popularity, to the strong consensus about the nature of the factor used (the market), to its usefulness in describing some of the features of the data being analyzed and to the heaviness of its structure, that allows for a quite precise parameters estimation. Single-index model covariance matrix, anyway, tends to contain a lot of bias related to a misspecified structural assumption[81], but little in the way of estimation error, exactly the opposite of what is true for the sample covariance matrix.

According to the SIM, returns are explained by the formula provided in section 1.4, that is reported here for allowing an easy understanding of the shrinkage computation procedure:

$$r_i = \alpha + \beta r_M + \epsilon_i \text{ ; with } i = 1, ..., N, \tag{21}$$

where residuals ϵ_i are uncorrelated both to market returns (r_M) and one another over time. The variance-covariance matrix of a portfolio under the SIM's framework is given by:

$$\Phi = \sigma_M^2 \beta\beta' + \Delta \tag{46}$$

being σ_M^2 the variance of market returns, β the vector of assets correlation with the market and Δ the diagonal matrix containing the variance of the residuals.

[80] See Jannagathan and Ma (2002) or Ledoit and Wolf (2003a).

[81] Using only 1 factor, of course, we cannot be able explain with precision a certain asset's returns.

However, as it has already been said in the previous paragraphs, such quantities are not known and must be estimated. The estimated version of SIM's variance-covariance matrix may be written as:

$$F = s_M^2 bb' + D. \tag{47}$$

For setting the shrinkage intensity (δ), Ledoit and Wolf propose to minimize a particular loss function associated to a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm. While a more complete description of this approach is provided in Appendix A, it may be useful here to describe the main steps for computing the magnitude of the shrinkage factor, without entering too deeply in mathematical details.

First of all, we define the Frobenius norm of a certain $N \times N$ symmetric matrix Z of elements z_{ij} and eigenvalues[82] λ_i, for $i, j = 1, \ldots, N$, as:

$$\|Z\|^2 = \text{Trace}(Z^2) = \sum_{i=1}^{N} \sum_{j=1}^{N} z_{ij}^2 = \sum_{i=1}^{N} \lambda_i^2 \tag{48}$$

Applying the Frobenius norm to the difference between the shrinkage estimator and the true covariance matrix we obtain the following quadratic loss function:

$$\text{L}(\delta) = \|\delta F + (1-\delta)S - \Sigma)\| \tag{49}$$

where Σ represents the true variance-covariance matrix. The optimal shrinkage factor will be the one which maximize the risk function defined as the expected value of the loss function defined in (49), i.e.:

$$\text{R}(\delta) = \text{E}[\text{L}(\delta)] = \text{E}\,\|\delta F + (1-\delta)S - \Sigma)\| \tag{50}$$

with N fixed while M goes to infinity, Ledoit and Wolf proved that δ behave like a constant over M, and that such a constant may be written as:

$$k = \frac{\xi - \rho}{\gamma} \tag{51}$$

where ξ represents the sum of asymptotic variances of the elements of the sample covariance matrix (S) multiplied by $\sqrt{M}$, i.e.:

$$\xi = \sum_{i=1}^{N} \sum_{j=1}^{N} \text{AsyVar}[\sqrt{M} s_{ij}]; \tag{52}$$

[82] An eigenvalue is defined as a scalar associated with a certain linear transformation of a vector space, having the property that there is some nonzero vector which, when multiplied by the scalar, is equal to the vector obtained by letting the transformation operate on the vector itself (definition by Merrian-Webster). For detailed information about Frobenius norms, eigenvalues and eigenvector check Kuttler (2015).

ρ represents the sum of asymptotic covariance of the elements of the shrinkage target (F) and from the sample covariance matrix (S), again scaled by $\sqrt{M}$, i.e.:

$$\rho = \sum_{i=1}^{N} \sum_{j=1}^{N} \text{AsyCov}[\sqrt{M} f_{ij}, \sqrt{M} s_{ij}]; \tag{53}$$

and γ measures the specification error contained in the shrinkage target:

$$\gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2. \tag{54}$$

If k were known we could choose k/M as the shrinkage intensity, but being it unknown we have to estimate it by estimating each one of its components. Ledoit and Wolf (2003) provide consistent estimators[83] for ξ, ρ and γ, which are reported in Appendix A. Once these quantities have been estimated, we are able to compute $\hat{k}$ and the shrinkage intensity $\hat{\delta}$.

$$\hat{\delta} = \frac{\hat{k}}{M} = \frac{1}{M} \frac{\hat{\xi} - \hat{\rho}}{\hat{\gamma}} \tag{55}$$

It may happen sometimes, although its very unlikely (Ledoit and Wolf, 2003), that $\hat{k}$ is lower than 0 or grater than 1, causing the resulting variance-covariance estimator not to be "shrinked" anymore. For this reason, the optimal shrinkage factor applied in practice by Ledoit and Wolf is the following:

$$\hat{\delta}^* = \max \left\{ 0, \min \left\{ \frac{\hat{k}}{M}, 1 \right\} \right\}. \tag{56}$$

We now have all the elements for deriving the Ledoit and Wolf's estimator defined in (45). Rewriting it in terms of $\hat{k^*}$[84], we obtain:

$$V_{LW} = \frac{\hat{k}^*}{M} F + \left(1 - \frac{\hat{k}^*}{M} \right) S \tag{57}$$

The computation of the weights for the Ledoit and Wolf's optimal portfolio does not require any other information, being their focus on global minimum variance combinations. The composition of such portfolio is then computed as in (35), with the sample variance-covariance matrix substituted by the shrinkage estimator described in (57).

[83] they're estimators that converge in probability to the true values or, in other words, that approach the true value when the sample size increase.

[84] where * denotes results achieved after running the control procedure in (56).

2.3.2.5 Nonnegativity constraints

We have already explained in section 2.1 how, one of the most direct and important consequences of estimation error, lies in the extreme values (very small or very large) that you can find in the estimated mean vectors and variance-covariance matrices. As noted by Michaud (1989), moreover, such extreme values often translate in extreme portfolio weights, causing that assets whose moments are estimated with the largest estimation error, are the ones which are usually overweighted by mean-variance optimization procedures. Shrinkage strategies try to deal with estimation error by providing better estimators of the distribution moments, but another solution, may be to directly constrain portfolio weights in order to avoid extreme concentrations, usually related to large errors in the estimates.

Frost and Savarino (1988)[85], as well as Chopra (1993)[86] show that short-sale constraints can actually help in reducing estimation error. More recently, Jagannathan and Ma (2003) apply nonnegativity constraints to global minimum variance portfolios, proving that constraining portfolio weights may help in reducing risk, with an improvement of portfolios performances. A possible downside in banning short-sales, anyway, is that, due to the non-linearity nature of the constraint, it is not possible to obtain an analytical solution for the mean-variance problem which may be written as:

$$\begin{aligned} &\max_{w} \ \mathrm{E(U)} = \max_{w} \ \ \mathrm{w}'\mu - \frac{\gamma}{2}\mathrm{w}'\mathrm{V}\mathrm{w} \\ &\text{subject to} \begin{cases} \mathrm{w}'\mathrm{e} = 1 \\ w_i \geqslant 0 \end{cases} \end{aligned} \tag{58}$$

where w_i, with $i = 1, ..., N$, represents the i-th element of the portfolio weights vector w. Fortunately, however, we may numerically solve the optimization using a quadratic programming algorithm[87], making use of the Karush-Kuhn-Tucker conditions. [88] Just in order to understand the role of short-sale constraints into mean-variance solutions, we consider the Lagrangian associated to the optimization problem defined above:

$$\begin{aligned} L =& \mathrm{w}'\mathrm{r} - \frac{\gamma}{2}\mathrm{w}'\mathrm{V}\mathrm{w} + \ddot{\lambda}(\mathrm{w}'\mathrm{e} - 1) + \lambda(\mathrm{w}' - 0) \\ =& \mathrm{w}'(\lambda + \mathrm{r}) - \frac{\gamma}{2}\mathrm{w}'\mathrm{V}\mathrm{w} + \ddot{\lambda}(\mathrm{w}'\mathrm{e} - 1) \end{aligned} \tag{59}$$

where $\ddot{\lambda}$ is the Lagrange multiplier associated to the constraint for which the

[85] P. A. Frost, J. E. Savarino (1988), "For Better Performance: Constrain Portfolio Weights", *Journal of Portfolio Management*, Vol. 15, No. 1, 29-34.

[86] V. K. Chopra (1993), "Improving optimization", *Journal of investing*, 8:51-59.

[87] See, for example, Roncalli (2010)

[88] These are first order conditions which, if certain assumptions are met, allow for optimality in the solution of a non-linear programming algorithm.

weights have to sum to 1, λ is the vector of Lagrange multipliers for the no-short-selling constraint whose elements, λ_i, are larger than or equal to 0. From the expression above, you may notice how the constrained version of the mean-variance portfolio is equivalent to an ordinary mean-variance portfolio with a different mean:

$$\tilde{r} = r + \lambda. \tag{60}$$

As noted by Kolusheva (2008)[89], the constraint for the j-th asset is likely to be binding ($\lambda_j > 0$), when the asset return is negative. In this case, short-sale constraint increases the expected return, shrinking it towards zero.

Something similar has been reported by Jagannathan and Ma (2003) for minimum variance portfolios. In particular, they found out that constructing a constrained global minimum variance portfolio from the usual sample variance-covariance matrix S, is equivalent to constructing an unconstrained minimum variance portfolio using a different estimator $\tilde{S}$, where:

$$\tilde{S} = S - (e\lambda' + \lambda' e) \tag{61}$$

where e is the usual column vector of ones and λ represents the vector of Lagrange multipliers related to the nonnegativity constraint. Notice that, when the constraint is binding for asset i, its covariances with other assets (S_{ij}, for $j \neq i$) are reduced by $\lambda_i + \lambda_j$, while its variance is reduced by $2\lambda_i$. The variance-covariance estimator defined in (61), therefore, is constructed by shrinking extreme values in original sample variance-covariance matrix toward the average covariances. The described applications from Kolusheva and Jannagathan and Ma, prove that weight constraint have a "shrinking like" effect on the return moments estimators and, similarly to the shrinkage strategies discussed in the previous paragraphs, may help in reducing estimation error.

The gain from imposing these constraints depends on the trade-off between the reduction in sampling error and the increase in specification error[90]. For portfolios constructed by using usual sample estimates, which often contain large estimation error, imposing these constraints is likely to be helpful. On the other hand, for portfolios constructed using strategies which are known for dealing with estimation error, like minimum variance portfolios or the portfolios based on shrinkage estimators, imposing such constraint is likely to hurt.[91] Similarly, when considering portfolios composed by assets which are themselves large portfolios, sampling error is usually small and, therefore, we should not expect constraints on weights to be very effective.

[89] D. Kolusheva (2008), "Out-of-sample Performances of Asset Allocation Strategies", *Working Paper*.

[90] It is related to a misspecification of the model used for the estimation procedure. Very simple models, with only a limited number of factors, usually suffer of this fallacy.

[91] See Jannagathan and Ma, 2003.

2.4 Improving the stability of Mean-Variance solutions

In section 2.3 we have seen how estimation error may erode the profitability of mean-variance selection strategies, not allowing investors to identify true optimal (and often not even efficient) solutions. However this is not the only problem related to inaccurate estimates of return distribution moments. Errors change in direction (sign) and magnitude over time, making the weights of mean-variance optimal portfolios particularly unstable over time. The larger the dispersion of such errors and the more the portfolio composition is expected to change in consecutive periods of time.[92]

Instability of mean-variance solutions, however, does not depend solely on estimation error. It has been proved, in fact, that also portfolios resulting from mean-variance optimization procedures which are known for being robust to estimation error (like the ones based on shrinkage estimators), and that usually outperform the 1/N benchmark in absence of transaction costs, heavily underperform it when transaction costs are taken into account.[93] Already in 1989, Michaud noted that even "when considering the true efficient frontier, taking a neighborhood of the optimal portfolio, we can find many statistically equivalent portfolios that, however, may have significantly or even radically different asset structures", which could mean that "portfolio structure is fundamentally not well defined".

Instability would not be such a problematic issue if investors should not pay commissions for trading on the market. The existence of transaction costs, however, strongly penalize mean-variance investment strategies which, as it has been explained, often require radical changes of portfolio composition at rebalancing dates. When portfolio turnover is particularly large, transaction costs may heavily affect the performances of mean-variance models, and this may lead investors to use the more intuitive and stable naïve diversification rule. It is simply natural to ask ourselves whether this fluctuation in weights is such an unavoidable effect or if we can, in some way, mitigate its magnitude and damages.

Fortunately, several solution to this fallacy have been proposed in literature. A traditional approach to promote stability, for example, has been to explicitly incorporate proportional transaction costs in the mean-variance framework.[94] Roughly speaking, you can penalize the turnover in the portfolio optimization problem, setting a "penalty coefficient" (k) which reflects the magnitude of transaction costs. This may be achieved by solving an optimization problem similar to the following:

[92] See, e.g., Kirby and Ostdiek (2012).

[93] See DeMiguel et al (2009) or Kourtis (2015).

[94] See Woodside et al. (2013) for a review of the recent literature about accounting for transaction costs in mean-variance problems.

$$\min_{w} \; w'Vw$$
$$\text{subject to} \begin{cases} w'r - k\,\|w - \tilde{w}\| = \xi \\ w'e = 1 \end{cases} \tag{62}$$

Due to the nonlinear form of proportional transaction costs, however, this procedure does not usually lead to a closed-form solution of the portfolio selection problem (See Kourtis, 2015).

Another possible solution, then, may be to impose an instability penalty directly to the mean-variance objective, where this penalty deals with the change in the portfolio weights in different period of time. The solution proposed by Kourtis (2015) goes in this direction. In particular, altogether with portfolio variance, he suggests that an investor should also minimize the deviation in portfolio weighs in consecutive periods. The mean-variance objective function may then be written as:

$$\min_{w} w'Vw + c(w - \tilde{w})'V(w - \tilde{w}) \tag{63}$$

where $\tilde{w}$ represents the vector of portfolio weights before trading, i.e. the weights resulting from the "old" portfolio after considering returns generated since the last rebalancing date, and c is the stability parameter. According to the expression above, moreover, investors should solve a trade-off between efficiency and stability, with an importance of the "stability issue" measured by the parameter c. For particularly large value of c, in fact, stability will be considered an extremely important feature and the portfolio resulting from the optimization procedure will be very similar to $\tilde{w}$.

In addition to provide an analytical solution to the portfolio optimization problem, the stability approach proposed by Kourtis has another attractive feature. Since it does not impose any limitation on how to estimate the efficient portfolio (the unstable one), this stabilization procedure may be applied to any sample-based asset selection strategy, even to the ones described in section 2.3, which may be useful in controlling for estimation risk. It has been proved[95], moreover, that using this stabilization method may lead to an increase in risk-adjusted performances in presence of estimation risk or when returns tend to show a positive serial correlation, even if not considering transaction costs.

An investor is free to select his personal stability parameter c, according to the importance he gives to the stability issue and, of course, to the magnitude of transaction costs. An interesting (and objective) solution proposed by Kourtis is to set c in such a way to obtain a portfolio turnover equal to the one produced by the $1/N$ strategy, where the turnover at a certain rebalancing date t, for the i-th

[95] See Kourtis (2015).

asset, may be defined as:

$$\tau_{i,t} = \|\hat{w}_{i,t} - \tilde{w}_{i,t}\| = \left\| (\hat{w}_{i,t} - \hat{w}_{i,t-1}) - \frac{\hat{w}_{i,t-1}(R_{i,t} - R_t^p)}{(1 + R_t^p)} \right\| \tag{64}$$

with $w_{i,t}$ being the percentage of wealth allocated to the i-th asset at time t after rebalancing, $\tilde{w}_{i,t}$ being the percentage of wealth invested in the i-th asset just before rebalancing, while $R_{i,t}$ and R_t^p are respectively the last period return on the i-th asset and on the whole portfolio. The portfolio turnover, then, is simply the sum of each asset turnover, i.e.

$$\tau^p = \sum_{i=1}^{N} \tau_i. \tag{65}$$

The choice of using the turnover from the 1/N benchmark as target for mean-variance portfolios turnover, is mainly due to the known stability in the weights of equally weighted portfolios[96] with respect to usual sample-based strategies. For obtaining such a results, we set c as follows:

$$\hat{c} = \frac{\tau^p - \tau^{\text{ew}}}{\tau^{\text{ew}}}. \tag{66}$$

where τ^p is the turnover associated to the efficient portfolio chosen by the investor and τ^{ew} is the turnover associated to the equally weighted one. The turnover of the stable version of a sample-based strategy is then defined as:

$$\bar{\tau} = \frac{1}{1 + \hat{c}} \tau^p = \tau^{\text{ew}}. \tag{67}$$

Once the stability parameter has been estimated we can compute the solution of the portfolio choice problem in the usual way. There is, however, another easier way to obtain represents the portfolio weights resulting from the optimization procedure. The composition of the stable portfolio, in fact, may be expressed as a linear combination of a portfolio lying on the efficient frontier[97] (w) and the one that is already held by the investor before rebalancing ($\tilde{w}$):[98]

$$\bar{w} = \frac{1}{1 + \hat{c}} w + \frac{\hat{c}}{1 + \hat{c}} \tilde{w} \tag{68}$$

At each trading period, then, the investor will move only a fraction of his wealth (equal to $\frac{1}{1+\hat{c}}$ toward the mean-variance target, leaving the remaining part ($\frac{\hat{c}}{1+\hat{c}}$) allocated as before.

[96] Of course, on each period, the investor will hold portfolios composed by all the assets in equal portions. Before rebalancing dates, however, returns may have affected such asset structure and trading is required to recover it, as it has been briefly explained in 2.3.1.

[97] Which depends upon the sample-based strategy chosen by the investor.

[98] A proof is provided in Appendix A.

3 An empirical analysis: data, models and methodologies

The aim of this work is to empirically compare different portfolio selection strategies trying to assess which is the most efficient one, i.e. the one that an investor should use when building up his portfolio, veryifing whether it is possible for mean-variance models, to outperform the 1/N benchmark even when transaction costs are taken into account. In order to practically compare different strategies and to measure their efficiency, I will use three different datasets that will be needed for estimating portfolios moments and for calculating their performances over time.

The comparison will be actually run in terms of Sharpe ratio, turnover and Sharpe ratio adjusted for transaction costs, assuming that these quantities are able to reflect the most important features of an attractive asset selection strategy, i.e. the capability of producing positive and stable returns even in presence of transaction costs. In section 3.1 I will described the datasets that will be used for the analysis, in section 3.2 I will briefly present the mean-variance alternatives considered and finally, in section 3.3 I will discuss the methodologies used for evaluating their performances.

3.1 Description of the datasets

For performing the empirical analysis to test the out-of-sample efficiency of the different portfolio selection strategies, I chose to use three different datasets, all containing monthly returns related to the period which goes from January 1995 until December 2015. All the datasets, then, will be composed of 21 years of data, divided into 252 monthly observations.

The choice of using monthly returns, instead of daily ones, is mainly due to the appealing features of the former, which have been known in literature for being at least approximately normally distributed[99] and less sensitive to market fluctuations[100], and for coherence with the recent literature about portfolio selection and management.[101]

The length of the period of analysis, on the other hand, allows the utilization of a large number of assets if desired[102], due to the number of firms listed since the beginning of the considered period, and for a sufficiently large out-of-sample period, at least when using rolling samples of 60 (or also 120) months as it is usually done in literature[103].

[99] At least more than daily ones.

[100] See, e.g., Fama (1965).

[101] See Ledoit and Wolf (2003, 2004), DeMiguel et al.(2009), Coqueret (2015) or Kourtis (2015).

[102] In the case of the present analysis, due to limitations in computational power, it has not been possible to consider a number of assets greater than 40. For N=60, just to make an example, it took about 4 hours (on an i7 processor) to perform the empirical test proposed on one of the 48 mean-variance alternatives proposed and that will be defined in section 3.2.

[103] See, e.g., Ledoit and Wolf (2003, 2004), DeMiguel et al. (2009), Kourtis (2015).

I have also chosen to focus on the US market, being all the datasets subsets of the S&P500[104], with the third one being simply an industry division of the famous market index. Such a choice is due both to the availability of long historical records for US stocks, and to the presence of a market index[105] whose structure has remained almost unchanged during the period of the analysis.[106] This has not been true, just to make an example, for the Italian market, where the MIB30[107] has been substituted, since June 2003, by the S&P MIB[108] which, in its turn, has been replaced by the FTSE MIB since June 2009, after the acquisition of Borsa Italiana by the London Stock Exchange Group. These changes in the structure of market indexes, may create homogeneity problems when performing historical analyses, at least when an harmonized solution is not provided by the Stock Exchange itself.[109] I will now provide a brief description of the datasets used, while a review is left to Table 1.

#	Description	N	Time period range	Abbrev.
1	Monthly returns from 40 stocks listed on the S&P500 since January 1995, Source: Bloomberg	40	Jan 1995 - Dec 2015	"SPL"
2	Monthly returns from 10 stocks listed on the S&P500 since January 1995, Source: Bloomberg	10	Jan 1995 - Dec 2015	"SPS"
3	Monthly returns from 10 industry portfolios of stocks listed on the S&P500, Source: Bloomberg	10	Jan 1995 - Dec 2015	"IND"

Table 1: List and description of the datasets used

The first dataset considered, that will be shortened as "SPL" (Standard & Poor's-Large), is a set of monthly returns from 40 of the 237 stocks listed on the S&P500 since January 1995. Even if the S&P500 actually counts 505 stocks[110], only a portion of them has been regularly quoted since the time chose as the starting date of the analysis. Being this dataset the one with the highest number of assets considered, coherently with what has been explained in section 2.3.1 we should expect traditional mean-variance models to perform quite poorly out of sample due to the number of parameters which has to be estimated. Moreover, being the assets considered individual stocks and not stock funds or portfolios, we should also expect estimation error to have an even greater impact on these models' performances.

[104]The S&P500 is a market index, introduced by Standard & Poor's in 1957, based on the market capitalization of the largest 500 companies listed on the NYSE or on the NASDAQ.

[105]Like, for example, the S&P500.

[106]The availability of a meaningful and coherent index over time will be useful when applying Ledoit and Wolf's solution to estimation error, which has been described in section 2.3.2.4.

[107]It was an index composed by the stocks of 30 largest Italian firms.

[108]This value weighted index was pretty similar to the actual FTSE MIB, representing the behavior of the stock prices from the 40 largest Italian firms.

[109]As it is in the case of the Italian stock market.

[110]Of the 500 companies listed, 5 of them quote two different share classes.

The second dataset proposed, that will be abbreviated as "SPS" (Standard & Poor's-Small), is composed by 10 stocks randomly chosen among the ones isted on the S&P500 since January 1995 and which have not been considered for costructing the "SPL" dataset. The purpose of using such a small subset, is to check whether, reducing the number of available assets, mean-variance solutions may perform better than how they do in the first one with respect to the 1/N benchmark.

The third dataset, shortened as "IND" (Industry), is composed of monthly returns from 10 industry portfolios based on stocks listed on the S&P500. The sectors considered are based on the first, and more general, industry partition provided by Bloomberg, and they are: Energy, Healthcare, Financials, Information Technology, Consumer- Discretionary, Consumer-Staples, Utilities, Industrials, Material and Telecommunication Services. The small number of assets, which this time are also represented by funds of stocks instead of individual securities as in the previous cases, may lead traditional mean-variance solutions to achieve performances similar to the one of the 1/N benchmark. For this dataset, then, we should expect shrinkage estimators as well as portfolio constraints to be less effective or, in certain cases, even dangerous.

3.2 Portfolio selection strategies

The code I wrote, which is reported in Appendix B, allows a would-be investor to study the efficiency of 49 different portfolio selection strategies. While a full list of such strategies is provided in table 2, here I will briefly introduce the way in which they have been chosen.

In addition to the naïve diversification rule ("1/N" or "ew"), I have included all the "traditional" mean-variance models defined in section 2.2 and 2.3.2.1:

- the one which uses the tangency rule for computing optimal portfolios ("mvT");
- the one which maximizes the expected utility ("mvU") given a risk aversion parameter γ that, coherently with the findings of both Bucciol and Miniaci (2007) and Cheng et al. (2010), I set equal to 5;[111]
- the one which sets the desired return π equal to the maximum between the expected return from the equally weighted portfolio and the one from the global minimum variance combination ("mvM");
- the one which ignores expected returns, focusing on global minimum variance portfolios ("min").

[111] Results for $\gamma = \{10, 20, 50\}$ will be provided in appendix C.

#	Model	Abbrev.
	Naive - Benchmark	
1	Equally weighted (1/N)	1/N or ew
	Traditional mean-variance alternatives	
2	Tangency portfolio	mvT
3	Utility maximization rule	mvU
4	$\pi = \max(r_{1/N}, r_{min})$	mvM
5	Minimum variance	min
	Models based on shrinkage estimators	
6	Jorion estimators for tangency portfolios	jorT
7	Jorion estimators with utility maximization rule	jorU
8	Jorion estimators with $\pi = \max(r_{1/N}, r_{min})$	jorM
9	Jorion estimators applied to minimum variance portfolios	jorMin
10	Ledoit-Wolf estimator for tangency portfolios	lwT
11	Ledoit-Wolf estimator with utility maximization rule	lwU
12	Ledoit-Wolf estimator with $\pi = \max(r_{1/N}, r_{min})$	lwM
13	Ledoit-Wolf estimator applied to minimum variance portfolios	lwMin
	Portfolios with nonnegativity contraints	
14	Tangency portfolio, no s.s.	mvT-c
15	Utility maximization rule, no s.s.	mvU-c
16	$\pi = \max(r_{1/N}, r_{min})$, no s.s.	mvM-c
17	Minimum variance portfolio, no s.s.	min-c
18	Jorion, tangency portfolio, no s.s.	jorT-c
19	Jorion, utility maximization rule, no s.s.	jorU-c
20	Jorion, $\pi = \max(r_{1/N}, r_{min})$, no s.s.	jorM-c
21	Jorion, minimum variance portfolios, no s.s.	jorMin-c
22	LW, tangency portfolios, no s.s.	lwT-c
23	LW, utility maximization rule, no s.s.	lwU-c
24	LW, $\pi = \max(r_{1/N}, r_{min})$, no s.s.	lwM-c
25	LW, minimum variance portfolios, no s.s.	lwMin-c
	Stable portfolios with turnover contraints	
26	Tangency rule, turnover constr.	mvT-t
27	Utility Maximization rule, turnover constr.	mvU-t
28	$\pi = \max(r_{1/N}, r_{min})$, turnover constr.	mvM-t
29	Minimum variance, turnover constr.	min-t
30	Jorion, tangency portfolios, turnover constr.	jorT-t
31	Jorion, utility maximization rule, turnover constr.	jorU-t
32	Jorion, $\pi = \max(r_{1/N}, r_{min})$, turnover constr.	jorM-t
33	Jorion, minimum variance portfolios, turnover constr.	jorMin-t
34	LW, tangency portfolios, turnover constr.	lwT-t
35	LW, utility maximization rule, turnover constraint	lwU-t
36	LW, $\pi = \max(r_{1/N}, r_{min})$, turnover constr.	lwM-t
37	LW, minimum variance portfolios, turnover constr.	lwMin-t
38	Tangency rule, no s.s. and turnover constr.	mvT-ct
39	Utility maximization rule, no s.s. and turnover constr.	mvU-ct
40	$\pi = \max(r_{1/N}, r_{min})$, no s.s. and turnover constr.	mvM-ct
41	Minimum variance portfolio, no s.s. and turnover constr.	min-ct
42	Jorion, tangency portfolio, no s.s. and turnover constr.	jorT-ct
43	Jorion, utility maximization rule, no s.s. and turnover constr.	jorU-ct
44	Jorion, $\pi = \max(r_{1/N}, r_{min})$, no s.s. and turnover constr.	jorM-ct
45	Jorion, minimum variance portfolios, no s.s. and turnover constr.	jorMin-ct
46	LW, tangency portfolios, no s.s. and turnover constr.	lwT-ct
47	LW, utility maximization rule, no s.s. and turnover constr.	lwU-ct
48	LW, $\pi = \max(r_{1/N}, r_{min})$, no s.s. and turnover constr.	lwM-ct
49	LW, minimum variance portfolios, no s.s. and turnover constr.	lwMin-ct

Table 2: List of the portfolio selection strategies considered

As it has been explained in section 2.3, however, most of these strategies usually suffer of large estimation error and, when applied out of sample, result quite inefficient both in absence and in presence of transaction costs. For this reason, I have also considered shrinkage solutions provided by both Jorion and Ledoit and Wolf, described respectively in section 2.3.2.3 and 2.3.2.4. Differently from what is done, for instance, in Ledoit and Wolf (2003), I structured my matlab code in such a way to allow the user to apply these solutions to all the traditional mean-variance models cited above[112]. In this way, it is possible to observe how the effect of using shrinkage estimators change among the different asset selection procedures considered, and to investigate the profitability of the "shrinked" version of these mean-variance solutions.

In addition to the effect using shrinkage estimators, I also analyze the consequences of adding nonnegativity and turnover constraints, defined respectively in section 2.3.2.5 and 2.4. In particular, I provide for a "long-only" version of all the mean-variance strategies previously defined.[113] In this way, the user may check the effectiveness that banning short sales has in dealing with estimation error and instability of mean-variance solutions, both for traditional models and for the ones which are already constructed for being robust to estimation risk.

A further step in the search for superior selection strategies[114], lies in investigating the effect of constraining portfolios turnover, following the procedure described in section 2.4. Also in this case, this routine is applied to all the mean-variance solutions previously discussed.[115] We expect, of course, this constraint to be less binding and useful when applied to long-only portfolios, due to their less extreme asset structure, which usually requires less trading activity at rebalancing dates.

3.3 Methodologies used for evaluating performances

Coherently with what is done in DeMiguel et al. (2009a,b), for estimating portfolios moments I use 60-months rolling samples. Samples, then, will always have length (denoted by M) equal to 60 months (5 years), with a composition that will change over time for including the most recent data and forgetting about the oldest. In practice, at each month t starting from $t = M + 1$, data from the previous M months will be used to estimate the parameters.[116]

Once all the elements needed for running portfolio optimization procedures has been defined, we can actually compare out-of-sample performances of differ-

[112] They are "mvT", "mvU", "mvM" and "min".

[113] Strategies based on shrinkage estimators are of course included.

[114] I.e. strategies which consistently lead to better out of sample performances than the one achieved by traditional mean-variance solutions and by the 1/N benchmark.

[115] "Shrinked" and "weights-constrained" ones included.

[116] To test whether the choice of the size of the rolling samples is critical, results for $M = 120$ will be provided in appendix C.

ent mean-variance strategies. This will be practically done using the following quantities:

- Sharpe ratio,
- average turnover,
- Sharpe ratio adjusted for transaction costs.

Even if some of these measures has already been introduced in section 2.3, it may be useful, in this context, to provide a brief review of these concepts and a description of the way in which they will be computed in the present analysis. Notice all the reported quantities[117] will be expressed in monthly terms.

3.3.1 Sharpe ratio

The Sharpe ratio is a very famous risk-adjusted measure of performance. It may be quite useful for evaluating funds or portfolios performances over time, taking into account not only the return produced by a certain strategy but also the risk implied by its use. In section 2.2.1 we defined the Sharpe Ratio as the "average return earned, in excess of the risk-free rate, per unit of total risk", when this risk is expressed in terms of return standard deviation, and we provided the following expression for computing its value:

$$\text{Sharpe Ratio} = \frac{r_p - r_f}{\sigma_p}. \tag{23}$$

where r_p is the average portfolio return, r_f is the average return on the risk-free asset and σ_p is the standard deviation of portfolio returns. Nonetheless the apparent simplicity of this formulation, there is not general accordance about how to estimate the average return from the portfolio and from the riskless assets. In particular, doubts arise in choosing between arithmetic and geometric average where, in this specific context, these quantities may be defined as follows:

$$\text{Arithmetic avg} = \frac{1}{T-M} \sum_{t=M+1}^{T} r_t \,, \tag{69}$$

$$\text{Geometric avg} = \sqrt[(T-M)]{\prod_{t=M+1}^{T} (r_t + 1)} - 1. \tag{70}$$

where adding "1" under the square root operator allows the computation of the geometric mean even in presence of negative returns. The main difference in using the geometric average, instead of the arithmetic one, lies in the capitalization effect of past returns. With the geometric average, in fact, past performances affect the

[117] Means, standard deviations and Sharpe ratios, both traditional and adjusted for transaction costs.

capability of the strategy of producing profits, as it is true for real financial applications. For this reason, I will use geometric average for computing out-of-sample average returns and Sharpe ratios produced by portfolio selection strategies.

Particular attention has to be paid when excess returns[118] are negative numbers. In this case, in fact, the Sharpe ratio will have a negative value too, and its interpretation gets more complicated. It is actually impossible to compare two negative Sharpe ratios without checking additional information, like the average portfolio return and the standard deviation associated to such returns. The reason is that a weaker magnitude of a negative Sharpe ratio is not always related to better portfolio performances, but it may also be caused by a larger returns dispersion. When displaying the results of the empirical analysis, then, also returns' geometric mean and standard deviation will be reported. This refinement may allow for a better comprehension even in the case in which Sharpe ratios are positive, helping in explaining their magnitude.

3.3.2 Turnover

Another important measure when analyzing historical returns from a certain investment strategy, is the turnover required by its implementation. The reason is quite simple: the more we change the composition of our portfolios, the higher transaction costs that we will have to pay, and the lower the net return that we may achieve. Turnover has been deeply discussed in section 2.4 and it may be defined as the change in weights between the portfolio held just before rebalancing ($\tilde{w}$), and the one that is held immediately after trading ($\hat{w}$). Recalling expression (64), the turnover of an asset is given by:

$$\tau_{i,t} = \|\hat{w}_{i,t} - \tilde{w}_{i,t}\| = \left\| (\hat{w}_{i,t} - \hat{w}_{i,t-1}) - \frac{\hat{w}_{i,t-1}(R_{i,t} - R_t^p)}{(1 + R_t^p)} \right\| \tag{64}$$

where $w_{i,t}$ represents the percentage of wealth allocated to the i-th asset at time t after rebalancing, $\tilde{w}_{i,t}$ is the percentage of wealth invested in the i-th asset just before rebalancing, while $R_{i,t}$ and R_t^p are respectively the last period return on the i-th asset and on the whole portfolio. The turnover of the entire portfolio, then, is calculated by summing up the individual turnover of all the available assets. Finally, in accordance with DeMiguel et al. (2009), we compute the turnover associated to a certain strategy as the average turnover on the out-of-sample period:

$$\bar{\tau}^p = \frac{1}{T - M} \sum_{t=1}^{T-M} \tau_t^p. \tag{71}$$

The larger the turnover implied by a certain strategy, the more its performances will be damaged by transaction costs, and the more impelling will be the need for

[118] Excess returns, usually represented with the Greek letter μ, are equal to difference between the portfolio returns and the risk-free rate.

a stabilization routine as the one described in section 2.4.

3.3.3 Sharpe ratio adjusted for transaction costs

The usual Sharpe Ratio does not consider how transaction costs affect a certain strategy performances. On the other hand, comparing different models simply using turnover does not tell us anything about the profitability of the models themselves. For making a meaningful comparison, then, we need to use a quantity which combines the information provided by these two measures, i.e. which could take into account both the capability of a portfolio selection strategies to produce profits and the amount of transaction costs that we should actually pay to implement it on real markets.

An interesting solution is to use a Sharpe Ratio adjusted for the presence of transaction costs. This measure may be computed using net returns ($\check{r}_p$) instead of the usual gross return (r_p), i.e.:

$$\text{Sharpe Ratio Adj.} = \frac{\check{r}_p - r_f}{\sigma_p}. \tag{72}$$

For calculating net returns, first of all, we have to compute the amount of transaction costs. Coherently with what is done in Balduzzi and Lynch (1999)[119] and DeMiguel et al. (2009), I assume proportional transaction costs equal to 50 basis points, where this percentage is based on the studies about transaction costs for individual stocks on the NYSE by Stoll and Whaley (1983)[120], Bhardwaj and Brooks (1992)[121] and Lesmond et al. (1999)[122]. The monetary amount of transaction costs may be then computed as:

$$TC_t = \tau_t^p \cdot E_t \cdot k \tag{73}$$

where τ_t^p is the portfolio turnover at time t, E_t is the available equity at time t and k is the percentage of proportional transaction costs. Once we have calculated TC_t we substract it from the available equity E_t in order to obtain the net available equity $\check{E}_t$ which, in turn, is used for computing portfolio net return for period t:

$$\check{r}_{p,t} = \frac{\check{E}_t - \check{E}_{t-1}}{\check{E}_{t-1}}. \tag{74}$$

[119] P. Balduzzi, A.W. Lynch (1999), "Transaction Costs and Predictability: Some Utility Cost Calculations", *Journal of Financial Economics*, 52:47–78.

[120] H. Stoll, R. Whaley (1983), "Transaction Costs and the Small Firm Effect", *Journal of Financial Economics*, 12:57–79.

[121] R. Bhardwaj, L. Brooks (1992), "The January Anomaly: Effects of Low Share Price, Transaction Costs, and Bid-Ask Bias", *Journal of Finance*, 47:553–574.

[122] D.A. Lesmond, J.P. Ogden, C.A. Trzcinka (1999), "A New Estimate of Transaction Costs", *The Review of Financial Studies*, 12:1113–1141.

The Sharpe ratio computed using net returns will be the main quantity for comparing the efficiency of mean-variance because, differently from the other measures defined in section 3.3.1 and 3.3.2, reflect all the main features that a portfolio selection strategy should have to be truly profitable when applied to real financial markets, i.e. being able to produce large returns limiting the amount of transaction costs paid.

3.3.4 Testing for statistical difference

It may be of interest, moreover, to check whether the "classic" and the adjusted Sharpe ratios resulting from a certain mean-variance strategy are statistically different from the ones produced by the 1/N benchmark. The traditional approach used in literature[123], which has been proposed by Jobson and Korkie (1981)[124] and improved by Memmel (2003)[125], assume returns to be i.i.d. normally distributed. Empirical evidence, however, suggests this assumption to be unreasonable, putting at risk the validity of the framework proposed by Jobson and Korkie.

For this reason, in the context of the present analysis, I will apply the robust methodology proposed in Ledoit and Wolf (2008)[126], which is based on a studentized bootstrap confidence interval. As it has been proved in literature, in fact, inference based on studentized bootstrap techniques is more accurate than the one based on asymptotic normality[127] both when using i.i.d data[128] and when working with time series data[129].
Just to give a general overview of how Ledoit and Wolf's test works, I will provide here a brief description of both the bootstrap method and of the studentized bootstrap confidence intervals used in their analysis.

The bootstrap[130] technique is a very famous tool used for constructing a confidence interval for a certain unknown parameter θ through simulation. This is done by letting the sample to play the role of the population and creating a very large number (B) of samples using the data from the original one.[131] It is then possible, for each new sample, to compute an estimate (usually defined as $\hat{\theta}_b^*$) of the unknown's parameter value (θ) and of the standard error[132] ($\hat{s}_b^*$) associated to

[123] See, e.g., DeMiguel et al. (2009a) or Gasbarro et al. (2007).
[124] J.D. Jobson, B.M. Korkie (1981), "Performance hypothesis testing with the Sharpe and Treynor measures", *Journal of Finance*, 36, 889–908.
[125] C. Memmel (2003), "Performance Hypothesis Testing with the Sharpe Ratio", *Finance Letters*, 1:21–23.
[126] O. Ledoit, M. Wolf (2008), "Robust performance hypothesis testing with the Sharpe ratio", *Journal of Empirical Finance*, 15, 850–859.
[127] Like the one proposed by Jobson and Korkie (1981).
[128] See Hall (1992).
[129] See Lahiri (2003).
[130] This term has been coined by Bradley Efron in 1979.
[131] For this reason bootstrap is sometimes called "resampling".
[132] The standard error is a measure of the variability of an estimators, and it is equal to $\frac{\hat{\sigma}}{\sqrt{n}}$ where $\hat{\sigma}$ is

$\hat{\theta}_b^*$ itself.

Studentized bootstrap confidence intervals (sometimes called "bootstrap t-intervals") are based on bootstrap-statistics, which make possible the computation of the intervals, with a formulation which resembles the one of the t-statistic[133]. Studentized bootstrap's static for the b-th sample, where $b = 1, \ldots, B$ is in fact equal to:

$$t_{boot,b} = \frac{\hat{\theta}_b^* - \hat{\theta}}{s_b(\hat{\theta})}. \tag{75}$$

After B values for $t_{boot,b}$ has been computed, the $\alpha/2$-lower and upper quantiles of the distribution of these $t_{boot,b}$ may calculated too and they are equal to t_L and t_U respectively. The boostrap confidence interval is the found as:

$$\left(\hat{\theta} + t_L \cdot s(\hat{\theta}),\ \hat{\theta} + t_U \cdot s(\hat{\theta})\right) \tag{76}$$

where $\hat{\theta}$ is the sample estimate of the unknown parameter θ.

What Ledoit and Wolf (2008) do, then, is to set a null hypothesis $H_0 : \Delta = 0$, where Δ is the difference of the two considered Sharpe ratios, using a two-sided confidence interval for Δ, with confidence level equal to $1-\alpha$ of the type described above. If, after choosing a specific nominal level α[134], the interval does not contain "0", H_0 is rejected at the nominal level α.

In the context of the present analysis, I set the bootstrap procedure to draw 5000 samples[135] from the original excess return one and, coherently with what is done in Kourtis (2015), to have a block size[136] of 10.[137]

the sample standard deviation and n is the sample size.

[133]Often used, for example, in Student's t-tests.

[134]usual values for α are 0.01, 0.05 and 0.1.

[135]It is the default value provided by Ledoit and Wolf (2008).

[136]It is the number of equal length blocks in which sample observations are divided when executing the resample. Resampling per blocks, is useful for allowing not to assume a constant data generator process during the sample period.

[137]The code used for this procedure by Ledoit and Wolf has been made freely-available online at http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9. The parts which have been modified for taking into account the different method that I used for computing the Sharpe ratios are reported in Appendix B.

4 An empirical analysis: results

In this section, the portfolio selection strategies listed in table 2 will be empirically compared in all the three different datasets described in section 3.1. As introduced in section 3.3, for each strategy I will report the average return during the out of sample period[138], the standard deviation, the Sharpe ratio, the turnover and, finally, the Sharpe ratio adjusted for transaction costs. In addition, for all the Sharpe ratios, both adjusted and unadjusted ones, I will signal the cases in which they are statistically different from the ones produced by the 1/N strategy, using Ledoit and Wolf's test described in section 3.3.4.
This chapter will be structured as follows: in section 4.1 I will discuss the results obtained using the first dataset ("SPL"), in section 4.2 I will study the results from the second dataset ("SPS"), while in section 4.3 the results from the third dataset will be analyzed. A comprehensive discussion about the effects of using shrinkage estimators, weights and turnover constraints will be provided in section 4.4, while, in section 4.5, I will finally address the issue of the identification of a superior portfolio selection strategy.

4.1 Results from the "SPL" dataset

In this section the results from the "SPL" dataset will be described. This dataset is composed by 40 stocks belonging to the S&P500, randomly chosen among the ones which have been listed since January 1995.

Coherently with what has been assumed in section 3.1, the large number of assets considered in this dataset, affects negatively the performances of traditional mean-variance solutions, which achieve, in particular for "mvT" and "mvU"[139], results that are strongly inferior than the ones reported for the 1/N benchmark. In general, poor risk adjusted returns are due to lower average returns and, when considering transaction costs, also to a level of average turnover which is usually quite larger than the one from the equally weighted strategy.

Instability in portfolio weights is particularly critical for tangency solutions and for the ones based on a direct maximization of the expected utility. For these combinations, in fact, average turnover levels are respectively 70 and 250 times larger than the one from "1/N".

As you may notice from table 3, for the "mvU" strategy both the mean, the Sharpe ratio and the Sharpe ratio adjusted have not been reported. This happens due to the very extreme portfolio returns achieved by the strategy that, at certain

[138] Coherently with what has been written in section 3.3.1, it will be computed in terms of geometric average.

[139] Results for "mvU" are not entirely reported for reasons that will be soon explained. However, as it will be cleared later, it is quite reasonable to assume very bad out of sample performances.

points in time, reach a values lower than -1, which represents a loss of the entire available equity. Given that it is not possible to compute the geometric average for negative values[140], it has not been feasible to obtain the mean and Sharpe ratios for this specific strategy. Such a strange result is probably related to an inaccurate estimation of the risk aversion coefficient. Empirical evidence provided by this analysis, in fact, seem suggesting the use of values for γ higher than the ones proposed in literature.[141] The utilization of higher risk aversion coefficients should actually lead to a less extreme asset structure and to a reduction in the portion of wealth allocated to the riskiest assets. This, in turn, should make portfolio returns more stable and predictable.

In general, when considering traditional mean-variance strategies, coherently with the hypotheses formulated in section 2, there is strong evidence in preferring global minimum variance portfolios ("min") or portfolio which set the target return to be equal to the maximum between the expected return from the global minimum variance portfolio and from the 1/N benchmark ("mvM"), at least when it is not possible to obtain precise estimates of the investor's risk aversion. The reason behind such preferences, lies in the greater capability of these models do deal with estimation error. Results from table 3 seem in fact confirming Merton's idea that most of the estimation error in mean-variance solutions is related to the estimates of the expected returns, which are completely ignored by "min" and only partially considered by "mvM".[142]

After analyzing classical mean-variance combinations, it is quite evident, however, that estimation error and instability in optimal solutions are strong enough to completely erase any advantage related to an optimal diversification, which, actually, has been proved to be all but truly optimal.

The introduction of shrinkage estimators, then, may be an appealing idea for reducing estimation error. Results provided in table 3 confirm this hypothesis. The performances of traditional mean-variance strategies, in fact, generally improve using either estimators discussed by Jorion or the ones proposed by Ledoit and Wolf. This positive effect is usually related to an increase in the average return, and to a reduction both in the standard deviation and in portfolios turnover. The only exceptions are "jorM" and "jorMin", whose profitability is the same of their classical counterparts. This finding may appear strange, but is easily explained recalling that the shrinkage target for Jorion's estimators is represented by the global minimum variance portfolio itself. It is quite natural, then, that no much shrinking is done in these two specific cases.

[140] And to continue investing after losing all the available capital.

[141] A further evidence will be provided in Appendix C when testing strategies based on the maximization of the expected utility for different risk aversion coefficients.

[142] In the case of "mvM", in fact, the choice of the desired return is restricted and, when it is set equal to the expected return of "min", results from "mvM" and from "min" obviously coincide.

If “min” and “mvM” have proved to be the most profitable mean-variance solutions among the traditional ones, the same is true when considering “shrinked” combinations, especially when applying the shrinkage technique proposed by Ledoit and Wolf: “lwM” and “lwMin”, in fact, are both characterized by a larger Sharpe ratio than the one provided by the equally weighted benchmark, when ignoring transaction costs. When such expenses are considered, however, even these two strategies are not able to replicate the performances achieved by “1/N”.

In order to obtain a further reduction in estimation error, and to start dealing with instability of mean-variance solution, we now consider the effect related to the application of nonnegativity constraints. From table 3, you may easily note that banning short sales has a positive effect in risk-adjusted terms both in presence and in absence of transaction costs. In particular, the more inefficient the considered strategy, the larger the benefits from constraining portfolio weights. These improvements appear to be related both to an increase in the average return and to a decrease in turnover[143], while the interpretation of the effect that nonnegativity constraints have on standard deviation remain trivial, and tend to vary with respect to the specific strategy considered.

We have 7 different constrained mean-variance strategies which outperform”1/N” in terms of adjusted Sharpe ratio: these are represented by all the combinations based on “min” and “mvM” methods for selecting optimal portfolios (“mvM-c”, “min-c”, “jorM-c”, “jorMin-c” and “lwMin-c”) and by “jorT-c”. The inclusion of the latter, which in the case of short-selling allowed has been shown to be quite inefficient[144], prove the effectiveness of nonnegativity constraints in fighting estimation error and instability in optimal mean-variance solutions, coherently with what has been explained in section 2.3.2.5.

Finally, we move to analyze the consequences of adding turnover constraints. From table 3 you may notice how, generally, they’re helpful in improving portfolio performances both when considering and when ignoring transaction costs. Comparing their effect to the one produced by the prohibition of short-sales, we find that, however, only in two cases (i.e. “lwMin-t” and “lwM-t”) the benefits of implementing turnover constraints are greater than the ones obtained banning short-selling. The reason for such result, lies probably in the limited exposition to estimation risk of the unconstrained version of these two strategies. If it were possible to obtain a precise estimation of return distribution moments, in fact, reducing the set of achievable combinations may even hurt the profitability of investment strategies. The only issue, in this case, would actually be to reduce the effect of transaction costs. We can conclude, then, that banning short sales is a more

[143] As explained in section 2.3.2.5, in fact, banning short-sales the asset structure of the resulting portfolios result less extreme.

[144] Look at the performances of “mvT”, “jorT”, “lwT” in table 3.

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
18	**1/N**	**0.0094**	**0.0430**	**0.2029**	**0.0569**	**0.1962**
Traditional mean-variance alternatives						
47	mvT	-0.0038	0.1398	-0.0322	3.9138	-0.1848
48	mvU	-	0.3139	-	14.3453	-
39	mvM	0.0060	0.0471	0.1134	1.0118	0.0053
41	min	0.0055	0.0483	0.1006	1.0168	-0.0052
Models based on shrinkage estimators						
44	jorT	0.0047	0.0778	0.0525	1.9216	-0.0719
45	jorU	0.0006	0.1107	-0.0008	3.2764	-0.1585
40	jorM	0.0060	0.0471	0.1134	1.0118	0.0053
42	jorMin	0.0055	0.0483	0.1006	1.0168	-0.0052
43	lwT	0.0017	0.0553	0.0182	0.7057	-0.0463
46	lwU	-0.0098	0.1299	-0.0802	2.2357	-0.1711
23	lwM	0.0073	0.0314	0.2127	0.2144	0.1784
21	lwMin	0.0078	0.0322	0.2209	0.1941	0.1906
Portfolios with nonnegativity constraints						
29	mvT-c	0.0073	0.0358	0.1865	0.2420	0.1527
36	mvU-c	0.0061	0.0456	0.1197	0.3158	0.0849
10	mvM-c	0.0089	0.0325	0.2523	0.1684	0.2264
7	min-c	0.0090	0.0330	0.2522	0.1603	0.2278
17	jorT-c	0.0085	0.0334	0.2343	0.1915	0.2056
22	jorU-c	0.0080	0.0341	0.2156	0.1887	0.1879
11	jorM-c	0.0089	0.0325	0.2523	0.1684	0.2264
8	jorMin-c	0.0090	0.0330	0.2522	0.1603	0.2278
28	lwT-c	0.0072	0.0348	0.1864	0.2138	0.1556
35	lwU-c	0.0066	0.0452	0.1314	0.2854	0.0997
15	lwM-c	0.0082	0.0314	0.2393	0.1377	0.2173
12	lwMin-c	0.0086	0.0324	0.2448	0.1263	0.2253
Stable portfolios with turnover constraints						
34	mvT-t	0.0056	0.0416	0.1194	0.0569	0.1125
37	mvU-t	0.0049	0.0645	0.0660	0.0569	0.0616
30	mvM-t	0.0055	0.0329	0.1466	0.0569	0.1379
32	min-t	0.0055	0.0335	0.1451	0.0569	0.1366
27	jorT-t	0.0060	0.0317	0.1690	0.0569	0.1600
26	jorU-t	0.0067	0.0343	0.1748	0.0569	0.1664
31	jorM-t	0.0055	0.0329	0.1466	0.0569	0.1379
33	jorMin-t	0.0055	0.0335	0.1451	0.0569	0.1366
38	lwT-t	0.0032	0.0385	0.0655	0.0569	0.0580
49	lwU-t	-	1.1010	-	0.0569	-
9	lwM-t	0.0075	0.0288	0.2373	0.0569	0.2273
1	lwMin-t	0.0080	0.0300	0.2453	0.0569	0.2358
19	mvT-ct	0.0079	0.0356	0.2038	0.0569	0.1958
25	mvU-ct	0.0082	0.0434	0.1741	0.0569	0.1675
5	mvM-ct	0.0085	0.0325	0.2412	0.0569	0.2324
3	min-ct	0.0086	0.0328	0.2415	0.0569	0.2328
14	jorT-ct	0.0083	0.0332	0.2284	0.0569	0.2198
16	jorU-ct	0.0081	0.0338	0.2191	0.0569	0.2107
6	jorM-ct	0.0085	0.0325	0.2412	0.0569	0.2324
4	jorMin-ct	0.0086	0.0328	0.2415	0.0569	0.2328
20	lwT-ct	0.0076	0.0346	0.1991	0.0569	0.1908
24	lwU-ct	0.0082	0.0434	0.1746	0.0569	0.1680
13	lwM-ct	0.0080	0.0314	0.2336	0.0569	0.2245
2	lwMin-ct	0.0085	0.0323	0.2437	0.0569	0.2348

Table 3: Results for the "SPL" dataset

appropriated answer to estimation error in mean-variance solution, while turnover constraints should be preferred when considering solutions which are known to be robust to estimation risk.

As it has been introduced in section 3.2, anyway, it is possible to combine nonnegativity and turnover constraints, and to apply this constraining-mix both to traditional mean-variance strategies and to "shrinked" ones. The effect of this double imposition is generally neutral in absence of transaction costs, while it is always positive when transaction costs are taken into account. For this latter kind of portfolio selection strategies, moreover, it is very likely to obtain performances superior or at least similar to the one achieved by the 1/N benchmark, even if we fail to obtain statistically significant differences using the Ledoit and Wolf's test described in section 3.3.4. It may be of interest to note that all the strategies whose performances result strictly superior to the "1/N" ones, refer either to the selection of global minimum variance portfolios or set the desired return as $\pi = \max(r_{1/N};\ r_{min})$.

4.2 Results from the "SPS" dataset

In this section results from the "SPS" dataset will be analyzed. Notice that this dataset is composed by 10 stocks listed on the S&P500, randomly chosen among the ones which have been excluded from the "SPL" dataset.

From table 4 you can easily note how traditional strategies (with no weights or turnover constraints) heavily underperform the 1/N benchmark both in absence and in presence of transaction costs. Generally, this is due both to a low average return level, which is even strongly negative in the "mvT" case, and to a very large level of turnover, which penalizes them with respect to the equally weighted benchmark when compared in adjusted Sharpe ratio terms. In particular "mvM" and "min" have a turnover that is 5 times the one produced by "1/N", while "mvU" and "mvT" reach a level which is respectively 27 and 183 times larger than the one associated to the equally weighted benchmark.

Looking at table 4, you may also notice that the mean, the Sharpe ratio and the Sharpe ratio adjusted for transaction costs have not been reported for the "mvU" strategy, while only the adjusted Sharpe ratio is missing for the "mvT" one. Similarly to what has been explained when describing results from the "SPL" dataset, the reason behind this shortcoming is the very extreme portfolio returns achieved by these strategy over time that, at certain points in time, reach values lower than -1. Given the impossibility of computing the geometric average for negative numbers, values for the mean and Sharpe ratios have not been reported for "mvU". What is different in the "mvT" case, is that while gross portfolio returns are always

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
13	**1/N**	**0.0059**	**0.0554**	**0.0946**	**0.0483**	**0.0902**
Traditional mean-variance alternatives						
47	mvT	-0.0261	0.2494	-0.1073	8.8118	-
48	mvU	-	0.1138	-	1.2916	-
39	mvM	0.0034	0.0485	0.0573	0.2286	0.0336
30	min	0.0041	0.0479	0.0724	0.1943	0.0521
Models based on shrinkage estimators						
49	jorT	-0.0169	0.2136	-0.0822	8.4370	-
43	jorU	-0.0004	0.0689	-0.0155	0.4531	-0.0484
40	jorM	0.0034	0.0485	0.0573	0.2286	0.0336
31	jorMin	0.0041	0.0479	0.0724	0.1943	0.0521
46	lwT	-0.0021	0.2295	-0.0119	3.4486	-0.1105
45	lwU	-0.0054	0.0927	-0.0653	0.8031	-0.1101
25	lwM	0.0044	0.0461	0.0813	0.1532	0.0646
15	lwMin	0.0052	0.0451	0.1005	0.1202	0.0871
Portfolios with nonnegativity constraints						
41	mvT-c	0.0038	0.0576	0.0538	0.2524	0.0319
35	mvU-c	0.0041	0.0534	0.0641	0.2189	0.0436
19	mvM-c	0.0050	0.0453	0.0953	0.1330	0.0806
6	min-c	0.0054	0.0448	0.1050	0.1025	0.0935
21	jorT-c	0.0062	0.0551	0.1002	0.2253	0.0797
26	jorU-c	0.0042	0.0471	0.0752	0.1413	0.0601
20	jorM-c	0.0050	0.0453	0.0953	0.1330	0.0806
7	jorMin-c	0.0054	0.0448	0.1050	0.1025	0.0935
37	lwT-c	0.0038	0.0568	0.0559	0.2165	0.0368
33	lwU-c	0.0041	0.0529	0.0647	0.1910	0.0466
18	lwM-c	0.0050	0.0451	0.0955	0.1162	0.0825
3	lwMin-c	0.0053	0.0445	0.1044	0.0822	0.0951
Stable portfolios with turnover constraints						
44	mvT-t	-0.0042	0.0772	-0.0628	0.0483	-0.0660
36	mvU-t	0.0032	0.0551	0.0468	0.0483	0.0424
16	mvM-t	0.0048	0.0452	0.0908	0.0483	0.0854
9	min-t	0.0049	0.0444	0.0960	0.0483	0.0906
27	jorT-t	0.0038	0.0525	0.0604	0.0483	0.0558
22	jorU-t	0.0047	0.0478	0.0846	0.0483	0.0796
17	jorM-t	0.0048	0.0452	0.0908	0.0483	0.0854
10	jorMin-t	0.0049	0.0444	0.0960	0.0483	0.0906
42	lwT-t	0.0020	0.0603	0.0216	0.0483	0.0175
28	lwU-t	0.0040	0.0566	0.0595	0.0483	0.0552
8	lwM-t	0.0052	0.0458	0.0985	0.0483	0.0932
1	lwMin-t	0.0055	0.0449	0.1069	0.0483	0.1015
38	mvT-ct	0.0029	0.0562	0.0403	0.0483	0.0360
32	mvU-ct	0.0034	0.0522	0.0515	0.0483	0.0468
11	mvM-ct	0.0050	0.0455	0.0960	0.0483	0.0906
4	min-ct	0.0051	0.0448	0.0994	0.0483	0.0940
24	jorT-ct	0.0044	0.0503	0.0737	0.0483	0.0689
23	jorU-ct	0.0045	0.0467	0.0815	0.0483	0.0763
12	jorM-ct	0.0050	0.0455	0.0960	0.0483	0.0906
5	jorMin-ct	0.0051	0.0448	0.0944	0.0483	0.0940
34	lwT-ct	0.0034	0.0546	0.0496	0.0483	0.0451
29	lwU-ct	0.0037	0.0518	0.0579	0.0483	0.0532
14	lwM-ct	0.0049	0.0452	0.0931	0.0483	0.0877
2	lwMin-ct	0.0052	0.0445	0.1019	0.0483	0.0965

Table 4: Results from the "SPS" dataset

greater than -1, net returns[145] are not. This, in turn, makes impossible only the computation of the Sharpe ratio adjusted for taking into account transaction costs.

While the causes for the extreme dynamic of the performances from "mvU" are the same which have been described in section 4.1, i.e. an unprecise estimation of the risk aversion coefficient, the reasons for the strong inefficiency of "mvT" lie in the highly risky mean-variance combination that it usually selects.[146] To achieve higher desired returns, in fact, tangency portfolios are usually characterized by an extreme asset structure[147] which, in addition, tend to be quite unstable over time; this, of course, causes large expenses in transaction costs and decrease profitability of the strategy especially when the magnitude of such expenses is taken into account.

Coherently with the findings for the "SPL" dataset, which have been described in section 4.1, also in the case of the "SPS" one there is strong evidence in preferring "min" and "mvM", when restricting the choice to traditional mean-variance solutions and it is not possible to obtain a precise estimate of the risk aversion coefficient.[148] None of the classical mean-variance solutions, however, is able to achieve comparable performances to the one provided by "1/N"[149], which is characterized by higher average mean and lower turnover.

Moving to analyze the effect of using shrinkage estimators, you may notice how, in general, they lead to an improvement in performances of mean-variance portfolios. In particular, empirical evidence suggests Jorion estimators to be more effective when applied to portfolio which contains particularly large estimation error, like "mvT" and "mvU"[150], while Ledoit and Wolf's ones result more efficient when considered altogether with "min" or "mvM" strategies. As it has been noted analyzing the results from the "SPL" dataset, moreover, "jorM" and "jorMin" achieve the same performances of their traditional counterparts, due to the nature of the shrinkage target used by Jorion.[151]
Even leading to an appreciable improvement in mean variance solutions' risk-adjusted returns, in absence of transaction costs only "lwMin" is able to reach performances superior to the ones provided by the naïve diversification rule. When transaction costs are considered, on the other hand, none of the strategies which has been described so far is able to outperform the 1/N benchmark.

The introduction of nonnegativity constraints, coherently with what has been

[145] Net of transaction costs.
[146] With respect for example to the more reasonable "mvM" and "min".
[147] Often due to large estimation error in the return distribution moments.
[148] Reasons for their greater efficiency are the same which have been reported in section 4.1.
[149] Both in absence and in presence of transaction costs.
[150] You may appreciate how it is possible to obtain a geometric mean for "mvU" when using Jorion estimators.
[151] It is in fact represented by the minimum variance portfolio itself.

stated in section 2.3.2.5, is quite helpful in improving portfolio performances, both in presence and in absence of transaction costs. This effect is especially evident when applied to mean-variance solutions which are strongly exposed to estimation error, like traditional ones, or even like "shrinkage strategies" based either on the maximization of the expected utility or to the selection of the tangency combination. Empirical evidence provided by this analysis suggests that nonnegativity constraints help in improving the performances of mean-variance solutions mainly by increasing average returns and reducing turnover[152], while the impact on standard deviation is usually less relevant.
From table 4 you may notice how, after the application of short-sale constraints, all the strategies based on the selection of minimum variance combinations ("min-c", "JorMin-c" and "lwMin-c") succeed in outperforming the 1/N benchmark even when transaction costs are taken into account.

All the strategies described until now, however, result more unstable than "1/N", and it may be useful to check whether, implementing turnover constraints, it would be possible to improve the performances of mean-variance solutions, especially in presence of transaction costs.

First of all, comparing the effect of adding nonnegativity and turnover constraints to previously unconstrained optimizations, you may notice how the consequences are actually quite similar. Limits to trading activity, in fact, generally increase average returns, reducing, at the same time, both standard deviation and turnover. When considering the application of turnover constraints on long-only mean-variance strategies, on the other hand, we are usually able to improve further their performances in presence of transaction costs. Among the turnover-constrained alternatives proposed, 8 succeed in outperforming the equally weighted benchmark in terms of Sharpe ratio adjusted and, once again, all the minimum variance combinations are included in this group.

In conclusion, results from the "SPS" dataset strongly suggest to use minimum variance solutions for selecting optimal portfolios. While there is not a significant preference in choosing between nonnegativity and turnover constraints, a combined application of both these restrictions is usually helpful in improving performances in presence of transaction costs. Among the 48 mean-variance alternatives considered, 12 are able to produce adjusted Sharpe ratio larger than the one provided by the 1/N benchmark but, however, this difference is never reported to be statistically significant.

[152]The reduction in turnover is of course related to the less extreme asset structure that characterizes short-sale constrained portfolios.

4.3 Results for the "IND" dataset

In this section the results from the "IND" dataset will be analyzed. As it has been stated in section 3.1, this dataset is composed by 10 industry portfolios of stocks listed on the S&P500.

As it is made clear in table 5, all traditional mean-variance strategies underperform the 1/N benchmark both when transaction costs are ignored and when they are considered. In general, the main fallacy of these solutions lies in showing a low average return, especially in the case of the tangency solution ("mvT"), and a very large level of turnover. For the tangency ("mvT") and the utility maximization rule ("mvU") in particular, the average turnover has been respectively 68 and 82 times greater than the one from the equally weighted benchmark. Such instability in portfolios' asset structure, heavily penalize traditional mean-variance strategies in terms of Sharpe ratio adjusted for transaction costs, which is negative for all the classical allocation rules considered.

Using Jorion shrinkage estimators, generally, we are not able to obtain consistent improvements in traditional strategies' performances. Better results come up with the application of Ledoit and Wolf's shrinkage solution, due to an increase in the average return, a decrease in standard deviation and, usually, also a drop in the average portfolios' turnover.[153] These improvements, even allowing "lwM" and "lwMin" to perform better than "1/N" in absence of transaction costs, are not sufficient to let any of the mean-variance strategies described so far to obtain superior performances with respect to the equally weighted benchmark, when these expenses are taken into account.

The introduction of nonnegativity constraints, on the other hand, has proved to be useful in increasing average returns and in decreasing turnover. In the case of portfolio selection strategies which are usually characterized by an extreme asset structure, with large return standard deviation and average turnover,[154] moreover, not only this effect is magnified, but also the standard deviation of portfolios' returns is significantly reduced.

As you may notice by looking at the results in table 5, simply restricting short sales, we may be able to obtain several combination with a Sharpe ratio adjusted for transaction costs larger than the one from the 1/N benchmark. These solutions usually refer to constrained mean-variance optimizations based either on global minimum variance portfolios ("min-c", "JorMin-c", "lwMin-c") or on the selection of a desired return equal to the maximum between the expected return from the

[153] The only exception is represented by "lwT", whose turnover increase when using Ledoit and Wolf's shrinkage estimator for the variance-covariance matrix. You may notice, however, that the same happens for "jorT". Shrinkage solutions, then, do not help at in improving performance of tangency solution in the case of the present dataset.

[154] As the one based either on the tangency rule or on the maximization of the expected utility.

global minimum variance portfolio and the one from the equally weighted benchmark ("mvM-c", "jorM-c", "lwM-c").[155] The improvement in performances, however, is significantly more effective when transaction costs are taken into account and this could mean that, at least for the most efficient traditional and "shrinked" mean-variance solutions, the benefits from imposing nonnegativity constraints are more related to an increase of stability than to a reduction in estimation error.

It is natural, then, to verify the consequence of adding turnover constraints. Their effect, at least in the of the present dataset, is generally quite more powerful than short-selling prohibition's one especially, of course, when considering transaction costs.[156] A possible explanation for this finding, lies in the particular structure of the dataset considered. The small number of available assets, which, in addition, are represented by large portfolios of stocks (instead of individual securities), probably allows a more precise estimation of both the efficient frontier and of the optimal mean-variance solutions, making estimation error a less dangerous enemy.[157] Moreover, when a precise estimation of the efficient frontier is feasible, short-sales prohibition may even hurt portfolio performances. A lower estimation error, however, solve only partially the instability problem in mean-variance solutions[158], for which a reduction in portfolio turnover is still required.
Another possible explanation for the greater impact of turnover constraints over nonnegativity ones, lies in the larger autocorrelation[159] that is often reported for portfolio returns, and which usually does not hold for individual securities.[160] Turnover constraints, then, leading to a more stable asset structure over time, increase the strategy's focus on momentum, allowing for larger returns when autocorrelation is positive.

When mixing short-sales prohibition and turnover constraint, as you may notice from table 5, we obtain better results than the ones coming from the usual short-selling ban. Simple turnover-constrained strategies[161], however, outperform turnover-constrained long-only solutions in all the analyzed cases.
For this dataset there are many mean-variance strategies which outperform the

[155] An exception is represented by "jorT-c" that, in spite of being based on the tangency rule, which it has been proved to be usually inefficient, is able to beat the 1/N benchmark. I think this result is more related to the low performances of "1/N" in this dataset than to particular properties of this specific portfolio selection strategy.

[156] The only exception are "lwT-t" and "lwU-t" whose performances are lower than the ones provided by "lwT-c" and "lwU-t" respectively, in particular when ignoring the effect of transaction costs; "lwT-c", moreover, shows a Sharpe ratio which smaller than "1/N" and this difference is statistically significant at the 0,1 confidence level.

[157] Further evidence to this hypothesis is given by that fact that, differently from the other dataset, this time has been possible to compute the geometric mean of both gross and net returns for all the analyzed strategies.

[158] As it has been explained in section 2.4, estimation risk and instability are positively correlated.

[159] It is defined as the degree of similarity between a given time series and a lagged version of itself over successive time intervals .

[160] See, e.g., Mech (1993).

[161] Meaning the ones for which short-selling is allowed.

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
24	**1/N**	**0.0046**	**0.0414**	**0.0942**	**0.0299**	**0.0906**
Traditional mean-variance alternatives						
48	mvT	0.0008	0.0703	0.0023	2.0694	-0.1465
47	mvU	0.0022	0.0812	0.0184	2.4624	-0.1337
43	mvM	0.0028	0.0357	0.0599	1.0173	-0.0827
41	min	0.0031	0.0346	0.0703	0.9261	-0.0638
Models based on shrinkage estimators						
49	jorT	-0.0018	0.0860	-0.0288	2.8988	-0.2073
45	jorU	0.0034	0.0506	0.0549	1.4945	-0.0931
44	jorM	0.0028	0.0599	0.0599	1.0173	-0.0827
42	jorMin	0.0031	0.0346	0.0703	0.9261	-0.0638
46	lwT	0.0027	0.0832	0.0239	2.2507	-0.1155
40	lwU	0.0048	0.0804	0.0512	1.8203	-0.0629
32	lwM	0.0044	0.0337	0.1115	0.3238	0.0634
28	lwMin	0.0044	0.0329	0.1143	0.2370	0.0783
Portfolios with nonnegativity constraints						
27	mvT-c	0.0050	0.0385	0.1116	0.2483	0.0794
35	mvU-c	0.0039	0.0434	0.0739	0.2157	0.0490
19	mvM-c	0.0049	0.0338	0.1241	0.1350	0.1041
17	min-c	0.0048	0.0338	0.1233	0.0949	0.1092
23	jorT-c	0.0052	0.0364	0.1250	0.2202	0.0948
31	jorU-c	0.0039	0.0366	0.0881	0.1645	0.0656
20	jorM-c	0.0049	0.0338	0.1241	0.1350	0.1041
18	jorMin-c	0.0048	0.0338	0.1233	0.0949	0.1092
26	lwT-c	0.0050	0.0381	0.1146	0.2371	0.0835
34	lwU-c	0.0039	0.0431	0.0749	0.2098	0.0506
16	lwM-c	0.0051	0.0331	0.1326	0.1189	0.1146
15	lwMin-c	0.0051	0.0332	0.1324	0.0809	0.1202
Stable portfolios with turnover constraints						
22	mvT-t	0.0024	0.0164	0.1064	0.0299	0.0973
33	mvU-t	0.0016	0.0149	0.0644	0.0299	0.0544
3	mvM-t	0.0056	0.0280	0.1752**	0.0299	0.1698**
1	min-t	0.0054	0.0259	0.1835**	0.0299	0.1777**
12	jorT-t	0.0036	0.0223	0.1333	0.0299	0.1265
6	jorU-t	0.0039	0.0219	0.1467	0.0299	0.1399
4	jorM-t	0.0056	0.0280	0.1752**	0.0299	0.1698**
2	jorMin-t	0.0054	0.0259	0.1835**	0.0299	0.1777**
36	lwT-t	0.0020	0.0275	0.0468	0.0299	0.0414
39	lwU-t	-0.0009	0.0259	-0.0533*	0.0299	-0.0550
8	lwM-t	0.0052	0.0322	0.1407	0.0299	0.1361
5	lwMin-t	0.0057	0.0318	0.1598	0.0299	0.1551
30	mvT-ct	0.0034	0.0381	0.0714	0.0299	0.0674
37	mvU-ct	0.0019	0.0433	0.0276	0.0299	0.0241
13	mvM-ct	0.0051	0.0343	0.1286	0.0299	0.1242
9	min-ct	0.0053	0.0343	0.1360	0.0299	0.1316
21	jorT-ct	0.0046	0.0366	0.1066	0.0299	0.1025
25	jorU-ct	0.0040	0.0369	0.0911	0.0299	0.0870
14	jorM-ct	0.0051	0.0343	0.1286	0.0299	0.1242
10	jorMin-ct	0.0053	0.0343	0.1360	0.0299	0.1316
29	lwT-ct	0.0036	0.0377	0.0774	0.0299	0.0734
38	lwU-ct	0.0018	0.0435	0.0256	0.0299	0.0221
11	lwM-ct	0.0052	0.0336	0.1355	0.0299	0.1310
7	lwMin-ct	0.0054	0.0337	0.1421	0.0299	0.1377

*,**,*** indicate a statistical divergence from "1/N" at 0.1, 0.05 and 0.01 confidence level

Table 5: Results from the "IND" dataset

1/N benchmark both in absence and in presence of transaction costs. Four of them, in particular, succeed in achieving Sharpe ratios (both adjusted and unadjusted) which are statistically different from the ones provided by "1/N"; these combinations are: "mvM-t", "min-t", "jorM-t" and "jorMin-t".[162]

4.4 Discussing the profitability of mean-variance alternatives

Nonetheless the three datasets considered were quite different for structure and composition, empirical evidences provided by their results appear similar and coherent. In all the datasets, in fact, it has been found that traditional mean-variance solutions heavily underperform the 1/N benchmark, both in presence and in absence of transaction costs, where no correction for estimation risk is implemented. This inefficiency is mainly due to low average returns and to a very high level of turnover, especially for solutions based either on the tangency rule or on a direct maximization of a quadratic utility function. The particularly poor out of sample performances of the latter strategy, moreover, may be explained by an erratic choice of the risk aversion parameter γ.[163]

We always find that focusing on minimum variance portfolios, or choosing a desired return equal to the maximum between the expected return from the minimum variance portfolio and the one from "1/N", are the most profitable choices for selecting optimal portfolios, even when not considering the classical version of mean-variance strategies.[164] This finding is probably related to lower importance that such selection methods attribute to the estimates of the expected return, which are known for being strongly exposed to estimation risk.[165]

Usually, using shrinkage estimators, we are able to improve the performances of traditional mean-variance strategies. While Jorion's estimators are generally more powerful when applied to the most inefficient combinations, Ledoit and Wolf's solution has always proved to be more useful than Jorion's one for increasing the out of sample profitability of strategies whose solutions are already more robust to estimation error.[166] Sometimes, simply using Ledoit and Wolf's estimators instead of usual sample ones, we are able to obtain Sharpe ratios larger than the one provided by "1/N" in absence of transaction costs;[167] when such expenses are taken into account, however, we always fail to outperform the equally weighted benchmark.

[162] You may notice how, once again, the results from the application of Jorion's shrinkage are completely absent when considering selection methods based on minimum variance portfolios, or to the specification of a desired return equal to the maximum between the expected return from the minimum variance portfolio and the one from the equally weighted benchmark.

[163] Results for different level of γ, which are reported in Appendix C, confirm this hypothesis.

[164] This remark, in fact, holds also when applying shrinkage estimators, and both nonnegativity and turnover constraints.

[165] See, on this topic, section 2.3.2.1.

[166] Like "min" or "mvM".

[167] It is always true for "lwMin", and always but in one dataset for "lwM".

The effect of adding nonnegativity constraints is positive in all the considered datasets. The benefits from their application are larger the more inefficient the unconstrained counterpart. The ways in which short sales prohibitions help in improving the out of sample performances of mean-variance solutions are mainly related to an increase in the average portfolio returns and to a reduction in turnover, while it is more trivial to provide a general interpretation of the effect that they cause on returns' standard deviation. In all the analyzed datasets, long-only minimum variance combinations are able to outperform "1/N", even in presence of transaction costs.

Also the effect of adding turnover constraints is usually positive and, of course, the benefits provided by their implementation increase when we allow for the presence of transaction costs. Comparing the effectiveness of nonnegativity and turnover constraints, what the present analysis reveals is that there is not a "superior technique" among these two for increasing the profitability of mean-variance solutions, but the best answer depends upon the structure of the dataset considered. For larger datasets composed by stocks, we usually have larger estimation risk and we should prefer nonnegativity constraints[168], while in smaller datasets, especially when the available assets themselves are represented by large portfolios, estimation risk is smaller, and we should focus more on reducing trading expenses by adding turnover constraints.

When applying short sales and turnover constraints together, we always obtain an improvement in risk-adjusted return with respect to usual long-only strategies, at least when transaction costs are considered. With the exception of the "IND" dataset, moreover, the constraint-mix usually outperform both "classical" turnover-constrained combinations and the 1/N benchmark. When estimation risk is small[169], however, the inclusion of a short-sales ban may hurt portfolio performances, if turnover constraints are already in place.

4.5 Is there an optimal portfolio selection strategy?

After performing such an analysis, and testing the profitability of so many different strategies, it is quite natural to ask ourselves whether a superior policy for selecting optimal mean-variance portfolios does exist, and whether such policy is actually capable to provide appealing performances.

Empirical evidence from the present study has revealed that there is not a single strategy which performs best in all the cases. On the other hand, it has been shown that there are several strategies whose performances are better than the

[168] The goal of turnover constraints, in fact, is not to reduce estimation risk.
[169] As it is true for the "IND" dataset.

ones provided by the 1/N benchmark in all the datasets considered.[170] These solutions are usually based on minimum variance portfolios that, ignoring expected return estimates, which are known for containing large estimation error, are able to select combinations which are truly close to optimal ones.

Another evidence provided by the present analysis is that, while the application of turnover constraints is always recommended, in order to reduce the effect of transaction costs, an additional prohibition of short-selling is suggested only when estimation error is expected to be large, that is when the number of available assets is large, the investment universe is only composed by individual securities, or the sample size is particularly small.[171]
When considering the solutions suggest above, finally, the difference in performances associated to the use of shrinkage estimators, instead of the usual sample ones, has been proved to be negligible.

[170] We fail, however, to obtain a difference which is statistical significance in all the datasets.
[171] From results for M=120 in appendix C, in fact, it is possible to assume a negative correlation between sample size and estimation error.

Conclusions

In this study the mean features and problems related to an application of the mean-variance approach to portfolio selection problems have been analyzed. In particular, I started by providing a detailed description of the Markowitz model, and a brief overview of some famous alternatives to the use of the mean and the variance as measures of central tendency and risk respectively.

I then proceeded by discussing the main reasons which explain the poor out of sample performances of mean-variance strategies, that essentially refer to estimation risk and to the instability in optimal portfolios' weights. Beside suggesting several methods for selecting the desired return, I have also proposed, in accordance with the recent literature on the topic, possible solutions for dealing with these fallacies and improving performances both in absence and in presence of transaction costs.

The core of this thesis, however, is represented by an empirical analysis which actually compares the profitability of different asset allocation strategies, and investigates about the existence of a superior portfolio selection model. More in detail, in addition to traditional mean-variance strategies based on sample estimates, I have also considered the use of shrinkage estimators and allowed for the introduction of both nonnegativity and turnover constraints.

Testing the performances provided by such models into 3 different datasets, I have proved that, as often reported in literature, traditional mean variance strategies strongly underperform the equally weighted benchmark, even in absence of transaction cost. When restricting our choice to usual sample-based mean-variance solutions, moreover, what I have found is that, in general, minimum variance combinations provide the largest return in risk-adjusted terms.

Using shrinkage estimators we are usually able to achieve better performances, and to approach the ones provided by the 1/N benchmark when transaction costs are not considered. When trading expenses are taken into account, on the other hand, 1/N still outperforms all these mean-variance strategies. A further step in the search for superior portfolio selection strategies is related to the introduction of nonnegativity constraints which, when applied to turnover-unconstrained models, are able to improve their performances in all the three dataset considered. Long-only minimum variance solutions, based either on sample or on shrinkage estimators, in particular, always succeed in outperforming the 1/N benchmark, even in presence of transaction costs.

Empirical evidence provided by the present analysis, moreover, shows that also adding turnover constraint to previously unconstrained solutions, has always a positive effect in terms of Sharpe ratio adjusted for transaction costs. Comparing

the effect of nonnegativity and turnover constraints, what I have found is that the first should usually be preferred when estimation error is expected to be large, while the latter works better when applied to strategies which are known already to be somehow robust to estimation risk. A simultaneous application of short sales prohibition and turnover constraint, finally, has been proved to be useful when a precise estimation of optimal mean-variance solutions is not feasible.

About the existence of a superior portfolio selection strategy, results from the present work seem suggesting that there is not a model which performs always best. On the other hand, formulating reasonable expectations about the magnitude of estimation error, it is often possible to identify a set of strategies that should provide appreciable performances.

Another contribute of this analysis lies in showing that minimum variance solutions usually outperform other methods proposed for selecting optimal mean-variance combinations. This finding, which is coherent with most of the recent literature, and that has proved to be robust to an increase in the sample size, is likely to be related both to the well known "low-volatility anomaly" and to the large estimation error contained in the estimates of the expected return, which are ignored by minimum variance strategies.

Ignoring information, clearly, isn't the most efficient way to deal with estimation error. It is quite obvious, therefore, that further efforts need to be made in order to obtain more precise forecasts of future returns and to achieve performances better than the ones provided by minimum variance solutions.

Appendix A: Proofs

Expected Utility and mean-variance optimization

We are going to prove that, in the case of a quadratic utility function, the maximization of the expected utility is coherent with the mean-variance dominance criterion.

Assume a generic quadratic utility function $U(X) = X - \frac{b}{2}X^2$ with $b > 0$. Being X the random wealth from a certain portfolio, we have that:

$$\begin{aligned} E[U(X)] &= E\left[X - \frac{b}{2}X^2\right] \\ &= E[X] - \frac{b}{2}E[X^2] \\ &= E[X] - \frac{b}{2}(E[X])^2 - \frac{b}{2}Var(X) \end{aligned} \tag{77}$$

You may appreciate how this expression depends only on the first two moments of X, and you should also remember that we can use this utility function only for $x < 1/b$, because only in this case we obtain an increasing function in x.

For a generic utility function $F(X)$ that is infinitely derivable in x, mean and variance are not sufficient to pursue the utility maximization, because its expected value may depend also upon other (higher) moments of the returns distribution. We can prove it by using Taylor expansions series, which allows for the approximation of a function as an infinite sum of terms, computed using the values of the function derivatives at a given point x.

$$\begin{aligned} E[F(X)] &= E\left[\sum_{i=0}^{\infty} \frac{F^{(i)}(x)}{i!}(X-x)^i\right] \\ &= \sum_{i=0}^{\infty} E\left[\frac{F^{(i)}(x)}{i!}(X-x)^i\right] \\ &= \sum_{i=0}^{\infty} \left[\frac{F^{(i)}(x)}{i!}E(X-x)^i\right] \end{aligned} \tag{78}$$

.

Computation of the weights for the tangency portfolio

As reported in Chapter 2, the tangency portfolio is defined as the portfolio of risky assets characterized by the highest Sharpe ratio. Starting from this definition we can set up the procedure for computing the weights of such a portfolio. Recovering the expression of the Sharpe ratio from (23), we maximize it under the constraint that the portfolio's weights sum to 1. In matrix form, we have:

$$\max_{w} \frac{w'r - r_f}{(w'\hat{V}w)^{1/2}} \; ; \text{ s.t. } w'e = 1 \tag{79}$$

For solving it, we set up the following Lagrangian:

$$L(w, \lambda) = (w'r - r_f)(w'\hat{V}w)^{-1/2} + \lambda(w'e - 1) \tag{80}$$

and, using the chain rule, we obtain the following first order conditions:

$$\begin{cases} \dfrac{\partial L}{\partial w} = r(w'\hat{V}w)^{-1/2} - (w'r - r_f)(w'\hat{V}w)^{-3/2}\hat{V}w + \lambda e = 0 \\ \dfrac{\partial L}{\partial \lambda} = w'e - 1 = 0 \end{cases} . \tag{81}$$

It's only a matter of algebra computations then for achieving the following expression:

$$w^T = \frac{\hat{V}^{-1}(r - r_f \cdot e)}{e'\hat{V}^{-1}(r - r_f \cdot e)}. \tag{24}$$

Jorion's shrinkage and intensity

As reported in section 2.3.2.3, Jorion, when formulating its shrinkage estimator, do not follow a strict Bayesian approach, but he allows for an estimation of the shrinkage target μ_0 and intensity $\hat{\delta}$ directly from the data. It will be now provided a brief description of how to derive expression (40) and (41) for calculating these coefficients. First of all, Jorion designs the following conjugate prior[172] for the sample means:

$$p(\mu \mid \eta, \hat{\lambda}) \propto \exp\left[-\frac{1}{2}(\mu - \eta e)'(\hat{\lambda}\hat{V})^{-1}(\mu - \eta e)\right] \tag{82}$$

with η and $\hat{\lambda}$, called hyper-parameters, representing the unknown grand mean and the prior precision.

The problem, then, is to find the predictive distribution of future returns r, conditional on the prior (82), on the available data y, on the covariance matrix $\hat{V}$ and on the hyper-parameter $\hat{\lambda}$, i.e.

$$p(r \mid y, \hat{V}, \hat{\lambda}) = \iint p(r, \mu, \eta \mid y, \hat{V}, \hat{\lambda}) d\mu \, d\eta. \tag{83}$$

[172] we have a conjugate prior when its distribution belongs to the same family of the related posterior distribution.

The joint density of r, μ and η is defined as:

$$\begin{aligned} p(r, \mu, \eta \mid y, \hat{V}, \hat{\lambda}) &= p(r \mid \mu, \eta, \hat{V}, \hat{\lambda}) \cdot p(\mu, \eta \mid y, \hat{V}, \hat{\lambda}) \\ &\propto p(r \mid \mu, \hat{V}) \cdot f(y \mid \mu, \hat{V}) p(\mu \mid \eta, \hat{\lambda}) p(\eta). \end{aligned} \tag{84}$$

As noted by Jorion (1985), with normality the likelihood function of y_m given μ and $\hat{V}$ is:

$$f(y \mid \mu, \hat{V}) \propto \exp\left[\left(-\frac{1}{2}\right)(y_m - \mu)'\hat{V}^{-1}(y_m - \mu)\right] \tag{85}$$

while the density function of μ given η and $\hat{\lambda}$, may written in terms of the following informative prior:

$$p(\mu \mid \eta, \hat{\lambda}, \hat{V}) \propto \exp\left[\left(-\frac{1}{2}\right)(\mu - \eta e)'\hat{\lambda}\hat{V}^{-1}(\mu - \eta e)\right]. \tag{86}$$

It is then possible to write the predictive density function as:

$$\begin{aligned} p(r, \mu, \eta \mid y, \hat{V}, \hat{\lambda}) \propto \exp\Bigg[&\left(-\frac{1}{2}\right)(r - \mu)'\hat{V}^{-1}(r - \mu) + \sum_{m=1}^{M}(y_m - \mu)'\hat{V}^{-1}(y_m - \mu) \\ &+(\mu - \eta e)'\hat{\lambda}\hat{V}^{-1}(\mu - \eta e)\Bigg]. \end{aligned} \tag{87}$$

After integrating over η and μ, this predictive density function can be shown to be normally distributed, with moments described by (39) and (43).

For estimating $\hat{\delta}$ directly from the data we recall that the probability density function $p(\lambda, \mid \mu, \eta, \hat{V})$ is a gamma distribution with mean equal to $(N+2)/d$ where d defined as $(\mu - e\eta)'\hat{V}^{-1}(\mu - e\eta)$, with μ and η that may be replaced by their sample counterparts $\hat{\mu}$ and $\hat{\mu}_0$. It is then possible to write the shrinkage coefficient $\hat{\delta}$ as in (41).

Estimation of Ledoit and Wolf's shrinkage intensity

For obtaining the optimal shrinkage intensity for the Ledoit and Wolf's variance-covariance matrix estimator we have to minimize the risk function defined in(50):

$$\begin{aligned}
\mathrm{R}(\delta) &= \mathrm{E}[\mathrm{L}(\delta)] = \|\delta F + (1-\delta)S - \Sigma)\| \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(\delta f_{ij} + (1-\delta)s_{ij}) + [\mathrm{E}(\delta f_{ij} + (1-\delta)s_{ij} - \sigma_{ij})]^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \delta^2\mathrm{Var}(f_{ij}) + (1-\delta)^2\mathrm{Var}(s_{ij}) + 2\delta(1-\delta)\mathrm{Cov}(f_{ij}, s_{ij}) + \delta^2(\phi_{ij} - \sigma_{ij})^2
\end{aligned} \tag{88}$$

where the first derivative is:

$$\mathrm{R}'(\delta) = 2\sum_{i=1}^{N}\sum_{j=1}^{N} \delta\mathrm{Var}(f_{ij}) - (1-\delta)\mathrm{Var}(s_{ij}) + (1-2\delta)\mathrm{Cov}(f_{ij}, s_{ij}) + \delta(\phi_{ij} - \sigma_{ij})^2 \tag{89}$$

and the second derivative is given by:

$$\mathrm{R}'' = 2\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2. \tag{90}$$

Since $\mathrm{R}(\delta)''$ is always positive, if we solve for (89) equal to zero, we are minimizing the function (50). In particular, when solving it for δ, we obtain the optimal shrinkage intensity(δ^*):

$$\delta^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(s_{ij}) - \mathrm{Cov}(f_{ij}, s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2} \tag{91}$$

where this equality may be rewritten as:

$$M\delta^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(\sqrt{M}s_{ij}) - \mathrm{Cov}(\sqrt{M}f_{ij}, \sqrt{M}s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2}. \tag{92}$$

As M gets larger, we define ξ, ρ and γ respectively as in (52), (53) and (54), while the first term in the denominator approaches 0:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(f_{ij} - s_{ij}) = O\left(\frac{1}{M}\right) \tag{93}$$

where $O(\frac{1}{M})$ indicate a term whose value is negligible as $\frac{1}{M}$ approaches 0 or, alternatively, when M tends to infinity. Now it is possible to write δ^* as:

$$\begin{aligned}\delta^* &= \frac{1}{M}\frac{\xi-\rho}{\gamma} + O\left(\frac{1}{M^2}\right) \\ &= \frac{k}{M} + O\left(\frac{1}{M^2}\right)\end{aligned} \tag{94}$$

which is equivalent to (51) as M gets large.

You should notice, however, that k and its components are not known and must be estimated. Here will be presented the consistent estimators that Ledoit and Wolf chose to use in their analysis.[173] Let:

$$\begin{aligned}\xi &= \sum_{i=1}^{N}\sum_{j=1}^{N}\xi_{ij}\ , \\ \rho &= \sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij}\ , \\ \gamma &= \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}\ .\end{aligned} \tag{95}$$

Ledoit and Wolf's consistent for ξ_{ij} is:

$$\hat{\xi}_{ij} = \frac{1}{M}\sum_{t=1}^{M}[(r_{it}-\bar{r}_{i.})(r_{jt}-\bar{r}_{j.})-s_{ij}]^2 \tag{96}$$

The consistent estimator for ρ_{ij} is:

$$\hat{\rho} = \sum_{i=1}^{N}\xi_{ii} + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\bar{\zeta}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\hat{\vartheta}_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}}\hat{\vartheta}_{jj,ij}\right) \tag{97}$$

where ζ is the average sample correlation:

$$\zeta = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \tag{98}$$

and $\vartheta_{xx,yx}$ is the consistent estimator for $\text{AsyCov}\left[\sqrt{M}s_{xx}, \sqrt{M}s_{yx}\right]$, which is given by:

$$\vartheta_{xx,yx} = \frac{1}{M}\sum_{i=1}^{M}\left[(r_{xt}-\bar{r}_{x.})^2 - s_{xx}\right]\left[(r_{yt}-\bar{r}_{y.})(r_{xt}-\bar{r}_{x.}) - s_{yx}\right]. \tag{99}$$

[173] see Ledoit and Wolf(2003a) for a detailed description of the consistency proofs.

Finally, the consistent estimator for γ_{ij} is defined as:

$$\gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} (f_{ij} - s_{ij})^2 \tag{100}$$

We are now able to compute a consistent estimator of k:

$$\hat{k} = \frac{\hat{\xi} - \hat{\rho}}{\hat{\gamma}}. \tag{101}$$

Kourtis' optimization equivalence for stable portfolios

The optimization procedure proposed by Kourtis for improving the stability of mean-variance solution may be rapresented in the following way:

$$\begin{aligned} &\min_{w} w'Vw + c(w - \tilde{w})'V(w - \tilde{w}) \\ &\text{subject to} \begin{cases} w'r = \pi \\ w'e = 1 \end{cases}. \end{aligned} \tag{102}$$

For solving it, we start by forming the following Lagrangian:

$$L = w'Vw + c(w - \tilde{w})'V(w - \tilde{w}) - \lambda_1(w'r - \pi) - \lambda_2(w'e - 1) \tag{103}$$

and then we write down the first order conditions:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 2(1+c)Vw - \lambda_1 r - \lambda_2 e - 2cV\tilde{w} = 0 \\ \dfrac{\partial L}{\partial \lambda_1} = -w'r + \pi = 0 \\ \dfrac{\partial L}{\partial \lambda_2} = -w'e + 1 = 0 \end{cases}. \tag{104}$$

Rearranging the first the first of the conditions defined above we obtain:

$$w = \frac{\lambda_1}{2(1+c)} V^{-1} r + \frac{\lambda_2}{2(1+c)} V^{-1} e + \frac{c}{1+c} \tilde{w} \tag{105}$$

and substituting this expression into the constraint ones we obtain:

$$\begin{cases} w = \dfrac{\lambda_1}{2(1+c)} V^{-1} r + \dfrac{\lambda_2}{2(1+c)} V^{-1} e + \dfrac{c}{1+c} \tilde{w} \\ \lambda_1 \dfrac{r' v^{-1} r}{2(1+c)} + \lambda_2 \dfrac{r' V^{-1} e}{2(1+c)} + \dfrac{c \tilde{w} r'}{2(1+c)} = \pi \\ \lambda_1 \dfrac{r' V^{-1} e}{2} + \lambda_2 \dfrac{e' V^{-1} e}{2} = 1 \end{cases} . \tag{106}$$

Solving for λ_1 and λ_2, and substituting these solutions in (105), we obtain the following:

$$\bar{w} = \frac{1}{1+\hat{c}} w + \frac{\hat{c}}{1+\hat{c}} \tilde{w}. \tag{68}$$

Appendix B: Matlab code

Main-Desimio-SPL

```
%%%%Main program%%%%
%%%Create mean-variance optimal portfolios and evaluate their performances


clear all

%%%INPUT
%%select the dataset
filename='datasets.xlsx';
sheet='stocks_spx_40';%as example, select the SPL dataset


%%set different choices for the TARGET RETURN
choice= menu('Choice of the Target Return','Tangency Portfolio','MAX(1/N, ...
    MIN(VAR))',...
    'UTILITY FUNCTION', 'Minimum Variance', '1/N');

c=0.005;% See DGU(2009)

%%extract the returns from the excel file
RetSeries=xlsread(filename,sheet)/100; %they are in x,xx format
[T,N]=size(RetSeries); %obtain n returns e n assets
M=60; %dimension of the rolling sample

%%extract the risk free rates
rf=xlsread(filename,'us_gov')/100;

ans=input('is short selling allowed? ("y"/"n")');

lb=zeros(1,N); %SS limit for each asset
if ans == 'y'
     minimo=-10;%a random low number, set anyway a limit
                 %to replicate a limited availability at a certain price
    for i=1:N
            lb(1,i)=minimo;
        i=i+1;
    end

end
%%choose whether to use shrinkage estimators or sample ones
shrink=menu('Choose the distribution moments estimators', 'Sample ...
    estimators', 'Jorion shrinkage', 'LW shrinkage');

%%choose whether to apply a turnover constraint or not
stable=menu('Do you want to obtain stable portfolios?', 'Yes, use ...
    turnover contraints', 'No');
```

```
if stable==1
%find eff ports, their E(r) and sd-->stable version
[ports, port_Em, ...
    port_Ecov]=extractFromSamplesStable(T,N,RetSeries,lb,choice,rf,shrink);

elseif stable==2
%unstable version
[ports, port_Em, ...
    port_Ecov]=extractFromSamplesUnstable(T,N,RetSeries,lb,choice,rf,shrink);
end

%%%Compute the out of sample monthly portfolio returns and the turnover

RetSeries_out=RetSeries(M+1:T,:);%select the out of the sample returns' ...
    matrix
RetSeries_inv=RetSeries_out';%compute the inverse matrix
act_r=ones(N,T-M); %asset return for each data
port_ret= zeros(T-M,1);%group returns per data

equity=zeros(T-M+1,1); %evolution of equity in time-->1   is time 0
transC_data=zeros(T-M,1);%how much money paid for trans costs at each data
net_returns=zeros(T-M,1);% returns net of taxes
equity(1,1)=1;%starting equity

avg_turn_data=zeros(T-M,1);%avg turnover per data

turnover_asset=zeros(N,T-M);%turnover for each asset at each data
turnover_data=zeros(T-M,1);%sum of asset turnovers per data


for i=1:T-M %for each data out of the sample
    for j=1:N %for each asset
        act_r(j,i)=ports(j,i)*RetSeries_inv(j,i);
        port_ret(i,1)=port_ret(i,1)+act_r(j,i); %sum returns per data, ...
            i.e. portfolio return
        j=j+1;
    end

    %%Compute the turnover

    if i==1 %in period 1 is simply the sum of the weights
        turnover_data(i)= sum(abs(ports(:,i)));
    else

         for j=1:N
            turnover_asset(j,i)=abs((ports(j,i)-ports(j,i-1))-...
                (ports(j,i-1)*(RetSeries_inv(j,i-1)-port_ret(i-1))/...
                (1+port_ret(i-1))));
            j=j+1;
         end
         %sum the turnover of all the assets for each data
         turnover_data(i)=sum(turnover_asset(:,i));
    end
```

```
        transC_data(i)= turnover_data(i)*equity(i)*c; %compute ...
            transaction costs per data
        equity(i+1)=equity(i)*(1+port_ret(i))- transC_data(i); %compute ...
            net equity
        net_returns(i)=(equity(i+1)/equity(i))-1; %compute net returns

    i=i+1;
end

%%PLOT THE EQUITY LINE
plot(equity)
hold on
Y=ones(size(equity));
plot(Y,'r--');
agg=input('scrivi il nome abbreviato');
agg1='Equity Line  -  ';
title_text=strcat(agg1,agg);
title(title_text);
xlabel('Months')
ylabel('Equity')
hold off

%%%RESULTS

%%compute the average turnover of the strategy
Turnover= (sum(turnover_data))/(T-M)

%%compute the standard deviation of portfolio returns
strat_sd= std(port_ret)

%strat_mean=(sum(port_ret))/(T-M)-->if you use the arithmetic avg
Z=port_ret+1; %preparare the data for the geometric avg
strat_gmean=geomean(Z)-1 %when computing the geometric avg remember to ...
    substract 1

rf_out=rf(M+1:T);%select the risk-free rate used in the out of sample period
Z=rf_out+1; %make the same thing for the risk free rate
rf_gavg= geomean(Z)-1;


%%Compute the Sharpe ratio

sharpe=(strat_gmean-rf_gavg)/strat_sd
sharpe_ann=sharpe*sqrt(12);%annualised sharpe

%compute excess returns
mu=port_ret-rf_out;

%bootstrap t-test for sharpe ratio (LW, 2008)
  temp=load('mu40.mat');%load the excess returns from 1/N
  ret_boot(:,1)=temp.mu40;
  ret_boot(:,2)=mu;
  [pValue]= bootInferenceRev(ret_boot)%obtain the p-value
```

```
%compute the average of the out of sample net returns
Z=net_returns+1;
strat_net_gmean=geomean(Z)-1;


%%compute the transaction costs-adjusted sharpe ratio
TransC_adj_sharpe= (strat_net_gmean-rf_gavg)/(strat_sd) %not annualised!

%net excess returns
mu_net=net_returns-rf_out;

 %%bootstrap t-test for the sharpe ratio adjusted
 temp_net=load('mu_net40.mat');%load the excess return from 1/N
 ret_boot_net(:,1)=temp_net.mu_net40;
 ret_boot_net(:,2)=mu_net;
 [pValue_net]= bootInferenceRev(ret_boot)

%remove some not useful variables
clear RetSeries_out Y act_r ans filename i j k minimo Z temp temp_net
```

extractFromSamplesUnstable function

```
function [ports, port_Em, port_Ecov] = ...
    extractFromSamplesUnstable(T,N,RetSeries,lb,choice,rf,shrink)
%%%It computes T-M efficient portfolios related to the choice of the target
%%%return. It does not consider turnover constraints


M=60;%I had too many input, so have to define it again

port_Em=zeros(T-M,1);%expected return from final portfolios
port_Ecov=ones(T-M,1);%sd of the final porfolios
ports=zeros(N,T-M); %matrix of portfolios' weights
x=1; %count the number of cycles (or created portfolios)
sample=zeros(M,N); %create the sample

for i=M+1:T
    z=i-M; %where to start for extracting returns

   for j=1:M
       sample(j,:)= RetSeries(z,:); %extract the row
       z=z+1;
       j=j+1;
   end

   [mean, cov] = ewstats(sample);%compute mean e cov matrices. ewstates ...
       requires MxN matrix (obsXassets)
                                %for the sample mean I use the arithmetic
                                %average (not geometric)
```

```
    p=Portfolio('Assetmean',mean,'Assetcovar',cov, ...
        'upperbudget',1,'lowerbudget',1,'lowerbound',lb, 'upperbound',10); ...
        %set portfolio
    %I set also an upperbound for avoiding crazy (and not always possible)
    %concentrantions in a specific asset



     if shrink==2 %JORION
         [mean_j,cov_j, delta_j]=shrinkageJorion(M,N,mean,cov,p);
         %delta_j
         p=Portfolio('Assetmean',mean_j,'Assetcovar',cov_j, ...
             'upperbudget',1,'lowerbudget',1,'lowerbound',lb, ...
             'upperbound',10);

     elseif shrink==3 %LW

         filename='datasets.xlsx';
         sheet='sp500';%use the data from the S&P500
         ret_mkt=xlsread(filename,sheet)/100;
         z=i-M; %where to start for extracting returns
        sampleM=zeros(M,1); %create the sample mkt return
        for j=1:M
            sampleM(j,:)= ret_mkt(z,:); %extract the row
            z=z+1;
            j=j+1;
        end

         [cov_lw]=shrinkageLWmkt(sample,sampleM);
         %I want to allow for portfolio selection also with other strategies,
         %withouth restricting the choice to global min-var portfolios
          p=Portfolio('Assetmean',mean,'Assetcovar',cov_lw, ...
              'upperbudget',1,'lowerbudget',1,'lowerbound',lb, ...
              'upperbound',10);
     end
        %%find the optimal portfolio, its exp ret and sd
        [ports(:,x), port_Em(x), ...
            port_Ecov(x)]=optimalMVshrink(p,choice,rf(i),N);
        i=i+1;
        x=x+1;
end
end
```

extractFromSamplesStable function

```
function [ports_def, port_Em, port_Ecov] = ...
    extractFromSamplesStable(T,N,RetSeries,lb,choice,rf,shrink)
%%%computes T-M efficient portfolios related to the choice of the target
```

```
%%%return. Stable version with turnover constraint
%%%(Turnover=Turnover 1/N)


M=60;%I had too many input, so I erase M and I define it again



port_Em=zeros(T-M,1);%exp return of final portfolios
port_Ecov=ones(T-M,1);%sd of the final portfolios
ports=zeros(N,T-M+1); %%matrix of portfolios' weights
x=1; %count the number of cycles (or created portfolios)
sample=zeros(M,N); %create the sample

ports(:,1)=zeros(N,1);

for i=M+1:T
    z=i-M; %where to start for extracting returns

   for j=1:M
       sample(j,:)= RetSeries(z,:); %extract the row
       z=z+1;
       j=j+1;
   end

   %extract the return for each data and each asset
    for j=1:N
    ret(j,x)= RetSeries(i-1,j);
    j=j+1;
    end

    %at the beginning we do not hold any portfolio, so the return is 0
    if x==1
        ret(:,x)=zeros(N,1);
    end


   [mean, cov] = ewstats(sample);%compute mean e cov matrices. ewstates ...
       requires MxN matrix (obsXassets)
                                %for the sample mean I use the arithmetic
                                %average (not geometric)

  p=Portfolio('Assetmean',mean,'Assetcovar',cov, ...
      'upperbudget',1,'lowerbudget',1,'lowerbound',lb, 'upperbound', 10);
   %I set also an upperbound for avoiding crazy (and not always possible)
   %concentrantions in a specific asset
    if shrink==2 %JORION
        [mean,cov, delta]=shrinkageJorion(M,N,mean,cov,p);

        p=Portfolio('Assetmean',mean,'Assetcovar',cov, ...
            'upperbudget',1,'lowerbudget',1,'lowerbound',lb, ...
            'upperbound', 10);

    elseif shrink==3 %LW
```

```
        filename='datasets.xlsx';
        sheet='sp500';%use the data from S&P500
        ret_mkt=xlsread(filename,sheet)/100;
        z=i-M; %where to start for extracting returns
       sampleM=zeros(M,1); %create the sample mkt return
       for j=1:M
           sampleM(j,:)= ret_mkt(z,:); %extract the row
           z=z+1;
           j=j+1;
       end

        [cov]=shrinkageLWmkt(sample,sampleM);
        %I want to allow for portfolio selection also with other strategies,
        %withouth restricting the choice to global min-var portfolios
         p=Portfolio('Assetmean',mean,'Assetcovar',cov, ...
             'upperbudget',1,'lowerbudget',1,'lowerbound',lb, ...
             'upperbound',10);
    end

   %find efficient portfolio, exp mean and exp sd--->before STABILITY
   [ports(:,x+1), port_Em(x), ...
      port_Ecov(x)]=optimalMVshrink(p,choice,rf(i),N);


%%%%%%%%%%%%%Compute TURNOVER 1/N

    act_r_EW(:,x)=1/N*(ret(:,x)); %return of the asset in 1/N is mean of ...
        the returns for the period
    port_ret_EW(x)=sum(act_r_EW(:,x));

    if x==1
     turnover_EW=1;%by definition
    else

          for k=1:N %for each asset
            turnover_asset_EW(k,x)=abs((1/N)*(ret(k,x)-port_ret_EW(x))/...
            (1+port_ret_EW(x)));
            k=k+1;
          end
    turnover_EW=sum(turnover_asset_EW(:,x));
    end

    %%%%%%%%%%%%%%Compute TURNOVER MV

       for j=1:N
        act_r_MV(j,x)=ports(j,x)*(ret(j,x)); %return of the asset in 1/N ...
            is mean of the returns for the period
        j=j+1;
      end

        port_ret_MV(x)=sum(act_r_MV(:,x));

      for k=1:N %for each asset
        turnover_asset_MV(k,x)=abs((ports(k,x+1)-ports(k,x))-...
```

```
            (ports(k,x)*(ret(k,x)-port_ret_MV(x))/(1+port_ret_MV(x))));
        k=k+1;
      end
      turnover_MV=sum(turnover_asset_MV(:,x));

    %%%%%%%%%%%%%%%%Compute STABLE Portfolio
       %instability penalty
      c=(turnover_MV-turnover_EW)/turnover_EW;

      %w tilde
      for j=1:N
       ports_withR(j,x)=ports(j,x)*(1+ret(j,x))/(1+port_ret_MV(x));
       j=j+1;
      end

     target_ret=port_Em(x);%fix the target return for the new MV port
     target_before=estimatePortReturn(p,ports_withR(:,x));% set target ...
         ret for the old port


     p=Portfolio('Assetmean',mean,'Assetcovar',cov, ...
         'upperbudget',1,'lowerbudget',1,'lowerbound',lb, 'upperbound',10);
     portaf_new(:,x)=estimateFrontierByReturn(p,target_ret);%compute the ...
         new efficient port

     ports(:,x+1)=(1/(1+c))*portaf_new(:,x)+(c/(1+c))*ports_withR(:,x);
    %compute the weights of the stable port


   i=i+1;
   x=x+1;
end

%%%erase the first column which has an empty portfolio
ports_def=ports(:,2:(T-M+1));

end
```

shrinkageJorion function

```
function [r_jorion,v_jorion, delta] = shrinkageJorion(M,N,mean,v,p)
%%%%Compute Jorion's estimators for the mean and the variance

%%compute the return of the minimum variance portfolio
[~,ret_limiti] = estimatePortMoments(p, estimateFrontierLimits(p));%find ...
    the returns form extreme ports
r_min= ret_limiti(1,1);%compute the minimum variance return

r=mean';
e = ones(N,1);
```

```
%%shrinkage
v_hat=(M-1)/(M-N-2)*v;
v1_hat=inv(v_hat);

delta_pre=(N+2)/((N+2)+M*(r-r_min)'*v1_hat*(r-r_min));

delta=max(0,min(delta_pre,1));%check delta lies in 0,1
r_jorion_trans=r_min*delta+r*(1-delta);
r_jorion=r_jorion_trans';
lambda=(M*delta)/(1-delta);

v_jorion=(1+(1/(M+lambda)))*v_hat+(lambda/(M*(M+1+lambda)))*...
                            ((e*e')/(e'*v1_hat*e));
end
```

shrinkageLWmkt function

```
function [cov_lw] = shrinkageLWmkt(x,xmkt)
%%compute the shrinkage estimator from Ledoit and Wolf(2003)
%%it uses 1 factor model as shrinkage target, where the factor is the
%%market. Differently from the original code from Ledoit and Wolf, which
%%is freely available at
%%http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9,
%%the mkt return is given as input and not computed from the
%%available assets (index may be different from the the set of available
%%assets).



t=size(x,1);
n=size(x,2);
meanx=mean(x);
x=x-meanx(ones(t,1),:);


sample=cov([x xmkt])*(t-1)/t;
covmkt=sample(1:n,n+1);
varmkt=sample(n+1,n+1);
sample(:,n+1)=[];
sample(n+1,:)=[];
prior=covmkt*covmkt'./varmkt;
prior(logical(eye(n)))=diag(sample);

  % compute shrinkage parameters
  c=norm(sample-prior,'fro')^2;
  y=x.^2;
  p=1/t*sum(sum(y'*y))-sum(sum(sample.^2));

  rdiag=1/t*sum(sum(y.^2))-sum(diag(sample).^2);
```

```
  z=x.*xmkt(:,ones(1,n));
  v1=1/t*y'*z-covmkt(:,ones(1,n)).*sample;
  roff1=sum(sum(v1.*covmkt(:,ones(1,n))'))/varmkt...
      -sum(diag(v1).*covmkt)/varmkt;
  v3=1/t*z'*z-varmkt*sample;
  roff3=sum(sum(v3.*(covmkt*covmkt')))/varmkt^2 ...
      -sum(diag(v3).*covmkt.^2)/varmkt^2;
  roff=2*roff1-roff3;
  r=rdiag+roff;
  % compute shrinkage constant
  k=(p-r)/c;
  shrinkage=max(0,min(1,k/t))


% compute the estimator
cov_lw=shrinkage*prior+(1-shrinkage)*sample;

end
```

optimalMVshrink function

```
function [portaf, portaf_m, portaf_cov] = optimalMVshrink(p,choice,rf,N)
%%%find the optimal portfolio and its expected mean and sd,
%%%allowing for the choice of the target return


if (choice==1)%%tangency
   p.RiskFreeRate=rf;
   portaf=estimateMaxSharpeRatio(p);

elseif (choice==2)
    %%target return for optimization is max(1/N, minVar)

    %a)compute E(r) for a 1/N strategy
    sample_mean=p.AssetMean;
    ret_EW=mean(sample_mean);

    %b)compute return of the global minimum var portfolio
    [~,ret_limiti] = estimatePortMoments(p, estimateFrontierLimits(p));
    ret_MinVar= ret_limiti(1,1);
    TargetReturn= max(ret_EW,ret_MinVar); %choose the higher between the ...
        previous 2

    portaf= estimateFrontierByReturn (p,TargetReturn); %compute the ...
        efficient portfolio

elseif (choice==3) %%My solution for the utility maximization problem
                   %%with known risk aversion parameter
    PortWts=estimateFrontier(p,100);
    [PortRisk, PortReturn]=estimatePortMoments(p,PortWts);
```

```
    A=5; %risk aversion parameter
    utility=zeros(100,1);

    for i=1:100
    r_p=PortReturn(i);
    sd_p=PortRisk(i);
    utility(i,1)= r_p- 0.5*A*((sd_p)^2+(r_p)^2); %quadratic utility function
    i=i+1;
    end

    %find the maximum utility and the index number
    [maxU,index]= max(utility);
    %extract the weights of the optimal port
    portaf=(PortWts(:,index));
    % portaf=(portaf_inv)'

elseif (choice==4)
    %%find the minimum variance portfolio
    [~,ret_limiti] = estimatePortMoments(p, ...
        estimateFrontierLimits(p));%find the returns form extreme ports
    ret_MinVar= ret_limiti(1,1);%compute the minimum variance return
    portaf= estimateFrontierByReturn (p,ret_MinVar); %compute the minVar ...
        portfolio from its e(r)

elseif (choice==5)
    %%find the equally weighted portfolio
    portaf=ones(N,1)*(1/N); %all the weights=1/N




end

[portaf_cov,portaf_m]=estimatePortMoments(p,portaf);%estimate e(r) and e(sd)
end
```

bootInferenceRev

```
function [pValue] = bootInferenceRev(ret,b,M,seType,pw,DeltaNull)
%% Carries out bootstrap test for equality of Sharpe ratios
%%Most of the code here reported is freely available at
%%http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9
%%This one differs from the origninal only for a different way of computing
%%the Sharpe ratio (geom-mean instead of arithmetic one)

% Inputs:
    % ret = [T,2] matrix of returns (in excess of the risk-free rate)
    % b = block size of circular bootstrap;
    %     of not specified by user, `optimal' block size will be
```

```
    %     computed by the routine blockSizeCalibrate
    % M = number of bootstrap repetitions; the default is M = 4999
    % seType = type of HAC standard error for 'original' test statistic
    %          use 'G' for Parzen-Gallant or 'QS' for Quadratic-Spectral;
    %          the default is seType = 'G'
    % pw = logical variable of whether to use prewhitended HAC standard
    %      error or not; the default is pw = 1
    % DeltaNull = the hypothesized value for Delta;
    %             the default is DeltaNull = 0
% Outputs:
    % pValue = bootstrap p-value for H_0: Delta = DeltaNull
    % DeltaHat = observed difference in Sharpe ratios
    % d = 'original' test statistic
% Note:
    %
    if (nargin < 6)
        DeltaNull = 0;
    end
    if (nargin < 5)
        pw = 1;
    end
    if (nargin < 4)
        seType = 'G';
    end
    if (nargin < 3)
        M = 3000;
    end
    if (nargin < 2)
        %b = blockSizeCalibrate(ret);
        b=10;%Kourtis (2015)
    end
    % compute observed difference in Sharpe ratios
    DeltaHat = sharpeRatioDiff_rev(ret);
    % compute HAC standard error (prewhitended if desired)
    [se,pval,sePw,pvalPw] = sharpeHACnoOut(ret,seType);
    if (pw)
        se = sePw;
    end
    % compute 'original' test statistic
    d = abs(DeltaHat-DeltaNull)/se;
    bRoot = b^0.5;
    [T,N] = size(ret);
    l = floor(T/b);
    % adjusted sample size for block bootstrap (using a multiple of the ...
        block size)
    Tadj = l*b;
    pValue = 1;
    for (m = 1:M)
        % bootstrap pseudo data and various bootstrap statistics
        retStar = ret(cbbSequence(Tadj,b),:);
        DeltaHatStar = sharpeRatioDiffRev(retStar);
        ret1Star = retStar(:,1);
        ret2Star = retStar(:,2);
        mu1HatStar = mean(ret1Star);
```

```
        mu2HatStar = mean(ret2Star);
        gamma1HatStar = mean(ret1Star.^2);
        gamma2HatStar = mean(ret2Star.^2);
        gradient = zeros(4,1);
        gradient(1) = gamma1HatStar/(gamma1HatStar-mu1HatStar^2)^1.5;
        gradient(2) = -gamma2HatStar/(gamma2HatStar-mu2HatStar^2)^1.5;
        gradient(3) = -0.5*mu1HatStar/(gamma1HatStar-mu1HatStar^2)^1.5;
        gradient(4) = 0.5*mu2HatStar/(gamma2HatStar-mu2HatStar^2)^1.5;
        yStar = [ret1Star-mu1HatStar,ret2Star-mu2HatStar,ret1Star.^2-...
                gamma1HatStar,ret2Star.^2-gamma2HatStar];
        % compute bootstrap standard error
        PsiHatStar = zeros(4,4);
        for (j = 1:l)
            zetaStar = bRoot*mean(yStar(((j-1)*b+1):(j*b),:));
            PsiHatStar = PsiHatStar+zetaStar'*zetaStar;
        end
        PsiHatStar = PsiHatStar/l;
        seStar = sqrt(gradient'*PsiHatStar*gradient/Tadj);
        % compute bootstrap test statistic (and update p-value accordingly)
        dStar = abs(DeltaHatStar-DeltaHat)/seStar;
        if (dStar >= d)
            pValue = pValue+1;
        end
    end
    pValue = pValue/(M+1);
end
```

sharpeRatioDiffRev

```
function [diff] = sharpeRatioDiffRev(ret)
% Computes the difference betweeen two Sharpe ratios
% most of this code has been made freely available by Ledoit and Wolf at
%http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9
% Inputs:
    % ret = T*2 matrix of returns (type double)
% Outputs:
    % diff = difference of the two Sharpe ratios
% Note:
    % returns are assumed to be in excess of the risk-free rate already
    ret1 = ret(:,1);
    ret2 = ret(:,2);
    %%%different from LW approach
    mu1hat = geomean(ret1+1)-1;
    mu2hat = mean(ret2+1)-1;
    sig1hat = var(ret1)^0.5;
    sig2hat = var(ret2)^0.5;
    SR1hat = mu1hat/sig1hat;
    SR2hat = mu2hat/sig2hat;
    diff = SR1hat - SR2hat;
end
```

Appendix C: Robustness checks

This appendix is useful to check whether the results provided in section 4 are robust to changes in the assumptions made in 3. In the context of the present analysis two different modifications will be tested:

- A change in the sample size, from M=60 (the usual) to M=120;
- Changes in the risk aversion coefficient (γ) used for the computation of optimal portfolios based on a direct maximization of a quadratic utility function.

I will start by addressing the first issue, i.e. testing all the 49 strategies[174] defined in section 3.2 assuming a sample size equal M=120. The out of sample periods, then, will be reduced from 192 to 132. Results for different risk aversion coefficients will be discussed in the second subsection of the present appendix. Given the evidence provided in section 4, the values of γ that will be tested here will all be larger than 5, which is the parameter's value assumed in section 3. More in detail, I will check for changes in the ranking of mean-variance strategies when transaction costs are taken into account, using the following values of γ: $10, 20, 50$.

Results for M=120

As you may notice making a comparison between the results provided in section 4 and the ones which are displayed by the table 6, 7 and 8, empirical evidence provided by the analysis remain almost intact. In spite of small changes in how mean-variance solutions are ranked, in fact, the two different studies lead to quite according results.

The small differences are usually related to an improvement in the precision of the sample estimates, caused by the larger amount of data used for their calculation. This effect is strongest when considering extreme solutions, like the ones based either on the tangency rule or on the maximization of the utility function, whose performances sometimes approach the ones offered my "min", "mvM" and their shrinked, or constrained, counterparts. The equally weighted porfolio, moreover, always ranks worse than how it did for M=60, and it becomes easier for mean-variance strategies to outperform it, even in presence of transaction costs.

Hereafter will be reported the results obtained by running portfolio selection strategies using a sample size of 120 monthly observations, in all the three datasets defined in section 3.1.

[174] 48 mean-variance alternatives and the equally weighted benchmark.

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
34	**1/N**	**0.0085**	**0.0422**	**0.1887**	**0.0543**	**0.1822**
Traditional mean-variance alternatives						
47	mvT	0.0033	0.0525	0.0530	0.7694	-0.0207
49	mvU	-0.0023	0.1354	-0.0205	2.5722	-0.1177
41	mvM	0.0039	0.0310	0.1105	0.3117	0.0600
43	min	0.0037	0.0315	0.1015	0.3086	0.0522
Models based on shrinkage estimators						
45	jorT	0.0039	0.0352	0.0971	0.4147	0.0379
46	jorU	0.0041	0.0588	0.0610	0.9128	-0.0169
42	jorM	0.0039	0.0310	0.1105	0.3117	0.0600
44	jorMin	0.0037	0.0315	0.1015	0.3086	0.0522
40	lwT	0.0063	0.0428	0.1358	0.3756	0.0916
48	lwU	0.0039	0.0893	0.0387	1.1625	-0.0277
32	lwM	0.0067	0.0282	0.2189	0.1399	0.1939
33	lwMin	0.0065	0.0282	0.2149	0.1349	0.1908
Portfolios with nonnegativity constraints						
19	mvT-c	0.0089	0.0344	0.2438	0.1423	0.2231
20	mvU-c	0.0104	0.0412	0.2412	0.1622	0.2215
15	mvM-c	0.0076	0.0292	0.2442	0.0939	0.2281
21	min-c	0.0073	0.0293	0.2330	0.0906	0.2175
6	jorT-c	0.0082	0.0301	0.2572	0.1004	0.2405
7	jorU-c	0.0087	0.0322	0.2565	0.1276	0.2366
16	jorM-c	0.0076	0.0292	0.2442	0.0939	0.2281
22	jorMin-c	0.0073	0.0293	0.2330	0.0906	0.2175
13	lwT-c	0.0089	0.0338	0.2499	0.1324	0.2303
23	lwU-c	0.0103	0.0416	0.2366	0.1611	0.2173
5	lwM-c	0.0078	0.0288	0.2549	0.0819	0.2407
11	lwMin-c	0.0076	0.0287	0.2462	0.0764	0.2328
Stable portfolios with turnover constraints						
26	mvT-t	0.0075	0.0313	0.2226	0.0543	0.2139
31	mvU-t	0.0078	0.0359	0.2052	0.0543	0.1976
36	mvM-t	0.0047	0.0286	0.1465	0.0543	0.1370
38	min-t	0.0044	0.0290	0.1355	0.0543	0.1261
35	jorT-t	0.0060	0.0289	0.1892	0.0543	0.1798
28	jorU-t	0.0076	0.0322	0.2207	0.0543	0.2122
37	jorM-t	0.0047	0.0286	0.1465	0.0543	0.1370
39	jorMin-t	0.0044	0.0290	0.1355	0.0543	0.1261
14	lwT-t	0.0082	0.0326	0.2366	0.0543	0.2283
30	lwU-t	0.0083	0.0378	0.2074	0.0543	0.2002
27	lwM-t	0.0067	0.0278	0.2223	0.0543	0.2125
29	lwMin-t	0.0065	0.0279	0.2164	0.0543	0.2066
2	mvT-ct	0.0091	0.0339	0.2532	0.0543	0.2452
9	mvU-ct	0.0104	0.0408	0.2420	0.0543	0.2353
17	mvM-ct	0.0074	0.0293	0.2370	0.0543	0.2277
24	min-ct	0.0071	0.0295	0.2247	0.0543	0.2154
8	jorT-ct	0.0078	0.0299	0.2446	0.0543	0.2355
4	jorU-ct	0.0084	0.0315	0.2510	0.0543	0.2423
18	jorM-ct	0.0074	0.0293	0.2370	0.0543	0.2277
25	jorMin-ct	0.0071	0.0295	0.2247	0.0543	0.2154
1	lwT-ct	0.0092	0.0334	0.2612	0.0543	0.2531
10	lwU-ct	0.0104	0.0411	0.2416	0.0543	0.2350
3	lwM-ct	0.0078	0.0288	0.2527	0.0543	0.2433
12	lwMin-ct	0.0074	0.0288	0.2411	0.0543	0.2317

Table 6: Results for the "SPL" dataset when M=120

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
35	**1/N**	**0.0071**	**0.0545**	**0.1204**	**0.0474**	**0.1160**
Traditional mean-variance alternatives						
48	mvT	-	0.1478	-	0.9964	-
45	mvU	0.0068	0.0736	0.0861	0.3921	0.0594
9	mvM	0.0091	0.0429	0.1998	0.1209	0.1857
3	min	0.0093	0.0414	0.2120	0.1007	0.1998
Models based on shrinkage estimators						
47	jorT	0.0047	0.0649	0.0655	0.2875	0.0429
27	jorU	0.0089	0.0472	0.1790	0.1618	0.1620
10	jorM	0.0091	0.0429	0.1998	0.1209	0.1857
4	jorMin	0.0093	0.0414	0.2120	0.1007	0.1998
49	lwT	-	0.1202	-	1.1266	-
43	lwU	0.0071	0.0648	0.1016	0.3070	0.0779
12	lwM	0.0089	0.0431	0.1943	0.0940	0.1833
5	lwMin	0.0090	0.0417	0.2050	0.0751	0.1959
Portfolios with nonnegativity constraints						
40	mvT-c	0.0069	0.0526	0.1225	0.1548	0.1077
33	mvU-c	0.0071	0.0490	0.1342	0.1480	0.1191
22	mvM-c	0.0081	0.0422	0.1800	0.1039	0.1676
13	min-c	0.0080	0.0406	0.1854	0.0555	0.1786
31	jorT-c	0.0071	0.0458	0.1442	0.1254	0.1304
25	jorU-c	0.0079	0.0419	0.1770	0.0839	0.1670
23	jorM-c	0.0081	0.0422	0.1800	0.1039	0.1676
14	jorMin-c	0.0080	0.0406	0.1854	0.0555	0.1786
37	lwT-c	0.0070	0.0530	0.1221	0.1353	0.1093
32	lwU-c	0.0071	0.0498	0.1324	0.1287	0.1195
28	lwM-c	0.0078	0.0422	0.1730	0.0952	0.1616
18	lwMin-c	0.0077	0.0408	0.1774	0.0505	0.1712
Stable portfolios with turnover constraints						
46	mvT-t	0.0038	0.0638	0.0519	0.0474	0.0482
39	mvU-t	0.0072	0.0593	0.1126	0.0474	0.1086
7	mvM-t	0.0089	0.0432	0.1948	0.0474	0.1893
1	min-t	0.0090	0.0410	0.2076	0.0474	0.2018
29	jorT-t	0.0079	0.0494	0.1502	0.0474	0.1454
17	jorU-t	0.0087	0.0459	0.1783	0.0474	0.1731
8	jorM-t	0.0089	0.0432	0.1948	0.0474	0.1893
2	jorMin-t	0.0090	0.0410	0.2076	0.0474	0.2018
44	lwT-t	0.0047	0.0633	0.0670	0.0474	0.0632
38	lwU-t	0.0070	0.0571	0.1135	0.0474	0.1093
11	lwM-t	0.0088	0.0435	0.1911	0.0474	0.1856
6	lwMin-t	0.0089	0.0416	0.2015	0.0474	0.1958
42	mvT-ct	0.0061	0.0536	0.1045	0.0474	0.1000
34	mvU-ct	0.0065	0.0489	0.1220	0.0474	0.1171
19	mvM-ct	0.0079	0.0421	0.1765	0.0474	0.1708
15	min-ct	0.0079	0.0407	0.1819	0.0474	0.1761
30	jorT-ct	0.0070	0.0454	0.1444	0.0474	0.1391
26	jorU-ct	0.0076	0.0419	0.1696	0.0474	0.1639
20	jorM-ct	0.0079	0.0421	0.1765	0.0474	0.1708
16	jorMin-ct	0.0079	0.0407	0.1819	0.0474	0.1761
41	lwT-ct	0.0062	0.0539	0.1060	0.0474	0.1015
36	lwU-ct	0.0063	0.0499	0.1174	0.0474	0.1126
24	lwM-ct	0.0078	0.0422	0.1729	0.0474	0.1672
21	lwMin-ct	0.0077	0.0408	0.1757	0.0474	0.1698

Table 7: Results for the "SPS" dataset when M=120

Rank	Model	Mean	St.Dev	Sharpe ratio	Turnover	Sharpe r.adj.
43	**1/N**	**0.0062**	**0.0404**	**0.1420**	**0.0289**	**0.1384**
Traditional mean-variance alternatives						
47	mvT	0.0074	0.0502	0.1372	1.3314	0.0033
46	mvU	0.0071	0.0480	0.1375	1.1531	0.0170
27	mvM	0.0082	0.0325	0.2366	0.3386	0.1841
22	min	0.0083	0.0316	0.2459	0.2721	0.2026
Models based on shrinkage estimators						
49	jorT	0.0085	0.0420	0.1909	1.7220	-0.0167
44	jorU	0.0075	0.0391	0.1799	0.8442	0.0721
28	jorM	0.0082	0.0325	0.2366	0.3386	0.1841
23	jorMin	0.0083	0.0316	0.2459	0.2721	0.2026
48	lwT	0.0072	0.0527	0.1273	1.4264	-0.0085
45	lwU	0.0065	0.0481	0.1249	0.8896	0.0322
17	lwM	0.0082	0.0320	0.2419	0.1919	0.2118
7	lwMin	0.0084	0.0312	0.2519*	0.1493	0.2279*
Portfolios with nonnegativity constraints						
34	mvT-c	0.0071	0.0406	0.1624	0.1234	0.1472
42	mvU-c	0.0068	0.0400	0.1566	0.1264	0.1408
20	mvM-c	0.0076	0.0323	0.2211	0.0666	0.2107
18	min-c	0.0075	0.0321	0.2198	0.0536	0.2114
30	jorT-c	0.0072	0.0358	0.1883	0.0850	0.1764
29	jorU-c	0.0072	0.0351	0.1906	0.0873	0.1781
21	jorM-c	0.0076	0.0323	0.2211*	0.0666	0.2107*
19	jorMin-c	0.0075	0.0321	0.2198*	0.0536	0.2114*
33	lwT-c	0.0071	0.0403	0.1652	0.1225	0.1499
40	lwU-c	0.0068	0.0400	0.1584	0.1239	0.1428
16	lwM-c	0.0077	0.0321	0.2235*	0.0629	0.2136*
15	lwMin-c	0.0076	0.0319	0.2228*	0.0490	0.2150*
Stable portfolios with turnover constraints						
36	mvT-t	0.0020	0.0094	0.1611	0.0289	0.1456
31	mvU-t	0.0025	0.0109	0.1820	0.0289	0.1687
4	mvM-t	0.0079	0.0299	0.2476**	0.0289	0.2427**
2	min-t	0.0065	0.0230	0.2631**	0.0289	0.2568**
24	jorT-t	0.0036	0.0150	0.2053	0.0289	0.1956
8	jorU-t	0.0058	0.0227	0.2323*	0.0289	0.2260*
5	jorM-t	0.0079	0.0299	0.2476**	0.0289	0.2427**
3	jorMin-t	0.0065	0.0230	0.2631**	0.0289	0.2568**
35	lwT-t	0.0029	0.0155	0.1554	0.0289	0.1461
32	lwU-t	0.0031	0.0161	0.1622	0.0289	0.1532
6	lwM-t	0.0082	0.0315	0.2448**	0.0289	0.2402**
1	lwMin-t	0.0075	0.0264	0.2641**	0.0289	0.2586**
38	mvT-ct	0.0065	0.0412	0.1469	0.0289	0.1434
39	mvU-ct	0.0063	0.0399	0.1468	0.0289	0.1432
13	mvM-ct	0.0078	0.0326	0.2244*	0.0289	0.2199*
11	min-ct	0.0077	0.0321	0.2251*	0.0289	0.2206*
26	jorT-ct	0.0074	0.0357	0.1934	0.0289	0.1893
25	jorU-ct	0.0074	0.0348	0.1978	0.0289	0.1936
14	jorM-ct	0.0078	0.0326	0.2244*	0.0289	0.2199*
12	jorMin-ct	0.0077	0.0321	0.2251*	0.0289	0.2206*
37	lwT-ct	0.0065	0.0408	0.1484	0.0289	0.1448
41	lwU-ct	0.0065	0.0398	0.1463	0.0289	0.1427
10	lwM-ct	0.0078	0.0323	0.2262*	0.0289	0.2217*
9	lwMin-ct	0.0078	0.0318	0.2282**	0.0289	0.2237**

*,**,*** indicate a statistical divergence from "1/N" at 0.1, 0.05 and 0.01 confidence level

Table 8: Results for the "IND" dataset when M=120

Results for different risk-aversion coefficient values

Here will be reported the results obtained by testing the mean-variance strategies based on a direct maximization of a quadratic utility function, for different values of the risk aversion coefficient γ. Results described in section 4 have shown how these portfolio selection models tend to be particularly inefficient with respect to other selection rules, as for example the minimum variance one.

My hypothesis is that such poor performances are mainly related to a choice of the risk aversion parameter which is not fully reasonable. For $\gamma = 5$, in fact, we usually select mean-variance combinations characterized by an extremely large expected return. The drawbacks of choosing a very low risk aversion coefficient, which corresponds to a very high desired return (π), are that the associated optimal solutions are often erroneously estimated, quite extreme in terms of portfolio's weights, and particularly unstable over time. All these features, with the exception of portfolios' weights which are not displayed, may be easily observed from the results provided in section 4.

Hereafter I will report the average returns, the standard deviations, the Sharpe ratios, the average turnover and the adjusted Sharpe ratios associated to mean-variance strategies with different level of γ. In particular, given that empirical evidence seem suggesting $\gamma = 5$ to be too small, I have tested the profitability of mean-variance solutions based on risk aversion coefficients equal to 10, 20 and 50. What I have found is that, in general, an increase in the risk aversion parameter lead to a significant improvement in all the quantities cited above. It is important to notice, however, that superior performances do not automatically imply greater reasonability. In particular, it is possible for these improvements to be related to an asset structure of the optimal portfolios which gets closer to the one of minimum variance solutions, which have proved to perform quite well, as the risk aversion increases.

Model	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
Results for the SPL dataset				
mvU	-	-0.0134	0.0008	0.0045
jorU	0.0006	0.0033	0.0050	0.0054
lwU	-0.0098	0.0027	0.0056	0.0075
mvU-c	0.0061	0.0067	0.0078	0.0087
jorU-c	0.0080	0.0085	0.0088	0.0090
lwU-c	0.0066	0.0069	0.0075	0.0083
mvU-t	0.0049	0.0058	0.0064	0.0061
jorU-t	0.0067	0.0065	0.0059	0.0056
lwU-t	-	0.0055	0.0075	0.0080
mvU-ct	0.0082	0.0075	0.0081	0.0085
jorU-ct	0.0081	0.0082	0.0085	0.0086
lwU-ct	0.0082	0.0075	0.0079	0.0084
Results for the SPS dataset				
mvU	-	0.0001	0.0029	0.0039
jorU	-0.0004	0.0027	0.0036	0.0041
lwU	-0.0054	0.0022	0.0042	0.0049
mvU-c	0.0041	0.0040	0.0047	0.0052
jorU-c	0.0042	0.0048	0.0051	0.0053
lwU-c	0.0041	0.0040	0.0046	0.0052
mvU-t	0.0032	0.0046	0.0050	0.0050
jorU-t	0.0047	0.0049	0.0049	0.0049
lwU-t	0.0040	0.0050	0.0055	0.0054
mvU-ct	0.0034	0.0042	0.0047	0.0050
jorU-ct	0.0045	0.0048	0.0050	0.0051
lwU-ct	0.0037	0.0044	0.0048	0.0052
Results for the IND dataset				
mvU	0.0022	0.0035	0.0030	0.0031
jorU	0.0032	0.0032	0.0030	0.0032
lwU	0.0045	0.0051	0.0048	0.0045
mvU-c	0.0039	0.0041	0.0045	0.0049
jorU-c	0.0039	0.0043	0.0047	0.0048
lwU-c	0.0039	0.0043	0.0047	0.0050
mvU-t	0.0016	0.0037	0.0049	0.0054
jorU-t	0.0039	0.0046	0.0051	0.0054
lwU-t	-0.0009	0.0028	0.0046	0.0055
mvU-ct	0.0019	0.0038	0.0050	0.0053
jorU-ct	0.0040	0.0049	0.0052	0.0053
lwU-ct	0.0018	0.0041	0.0052	0.0055

Table 9: Average return for different values of γ

Model	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
Results for the SPL dataset				
mvU	0.3139	0.1521	0.0799	0.0504
jorU	0.1107	0.0693	0.0532	0.0489
lwU	0.1299	0.0682	0.0407	0.0337
mvU-c	0.0456	0.0379	0.0348	0.0328
jorU-c	0.0341	0.0331	0.0328	0.0329
lwU-c	0.0452	0.0373	0.0338	0.0320
mvU-t	0.0645	0.0478	0.0344	0.0322
jorU-t	0.0343	0.0311	0.0307	0.0327
lwU-t	1.1010	0.0378	0.0317	0.0298
mvU-ct	0.0434	0.0371	0.0343	0.0328
jorU-ct	0.0338	0.0332	0.0328	0.0328
lwU-ct	0.0434	0.0366	0.0335	0.0320
Results for the SPS dataset				
mvU	0.1138	0.0726	0.0562	0.0496
jorU	0.0689	0.0551	0.0503	0.0485
lwU	0.0927	0.0622	0.0504	0.0461
mvU-c	0.0534	0.0488	0.0462	0.0451
jorU-c	0.0471	0.0456	0.0451	0.0449
lwU-c	0.0529	0.0486	0.0458	0.0447
mvU-t	0.0551	0.0502	0.0470	0.0449
jorU-t	0.0478	0.0461	0.0452	0.0445
lwU-t	0.0566	0.0502	0.0474	0.0453
mvU-ct	0.0522	0.0485	0.0460	0.0450
jorU-ct	0.0467	0.0454	0.0449	0.0448
lwU-ct	0.0518	0.0487	0.0459	0.0447
Results for the IND dataset				
mvU	0.0812	0.0498	0.0398	0.0354
jorU	0.0501	0.0405	0.0365	0.0349
lwU	0.0807	0.0489	0.0373	0.0334
mvU-c	0.0434	0.0371	0.0343	0.0336
jorU-c	0.0366	0.0344	0.0338	0.0337
lwU-c	0.0431	0.0368	0.0338	0.0330
mvU-t	0.0149	0.0223	0.0267	0.0272
jorU-t	0.0219	0.0250	0.0265	0.0265
lwU-t	0.0258	0.0299	0.0327	0.0329
mvU-ct	0.0433	0.0373	0.0352	0.0343
jorU-ct	0.0369	0.0352	0.0344	0.0343
lwU-ct	0.0435	0.0367	0.0345	0.0336

Table 10: Returns' standard deviation for different values of γ

Model	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
Results for the SPL dataset				
mvU	-	-0.0134	0.0013	0.0762
jorU	-0.0008	0.0383	0.0811	0.0967
lwU	-0.0802	0.0298	0.1213	0.2028
mvU-c	0.1197	0.1580	0.2050	0.2460
jorU-c	0.2156	0.2362	0.2476	0.2529
lwU-c	0.1314	0.1678	0.2023	0.2404
mvU-t	0.0660	0.1080	0.1669	0.1694
jorU-t	0.1748	0.1871	0.1717	0.1523
lwU-t	-	0.1273	0.2152	0.2445
mvU-ct	0.1741	0.1841	0.2165	0.2390
jorU-ct	0.2191	0.2284	0.2385	0.2417
lwU-ct	0.1746	0.1876	0.2161	0.2421
Results for the SPS dataset				
mvU	-	-0.0083	0.0392	0.0662
jorU	-0.0155	0.0371	0.0575	0.0705
lwU	-0.0653	0.0242	0.0703	0.0928
mvU-c	0.0641	0.0684	0.0866	0.1012
jorU-c	0.0752	0.0906	0.0994	0.1038
lwU-c	0.0647	0.0693	0.0860	0.1016
mvU-t	0.0468	0.0792	0.0913	0.0972
jorU-t	0.0846	0.0922	0.0940	0.0958
lwU-t	0.0595	0.0860	0.1011	0.1044
mvU-ct	0.0515	0.0735	0.0874	0.0973
jorU-ct	0.0815	0.0902	0.0958	0.0983
lwU-ct	0.0579	0.0777	0.0900	0.1009
Results for the IND dataset				
mvU	0.0184	0.0578	0.0592	0.0687
jorU	0.0501	0.0624	0.0628	0.0715
lwU	0.0481	0.0913	0.1113	0.1160
mvU-c	0.0739	0.0917	0.1118	0.1256
jorU-c	0.0881	0.1070	0.1182	0.1240
lwU-c	0.0749	0.0979	0.1189	0.1327
mvU-t	0.0644	0.1355	0.1583	0.1741
jorU-t	0.1467	0.1564	0.1667	0.1788
lwU-t	-0.0590	0.0718	0.1189	0.1462
mvU-ct	0.0276	0.0848	0.1232	0.1363
jorU-ct	0.0911	0.1219	0.1315	0.1360
lwU-ct	0.0256	0.0938	0.1321	0.1428

Table 11: Sharpe ratios for different values of γ

Model	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
Results for the SPL dataset				
mvU	14.3453	4.6365	2.0617	1.1731
jorU	3.2764	1.7028	1.2198	1.0543
lwU	2.2357	0.9401	0.4691	0.2423
mvU-c	0.3158	0.2575	0.2053	0.1552
jorU-c	0.1887	0.1628	0.1587	0.1594
lwU-c	0.2854	0.2296	0.1692	0.1219
mvU-t	0.0569	0.0569	0.0569	0.0569
jorU-t	0.0569	0.0569	0.0569	0.0569
lwU-t	0.0569	0.0569	0.0569	0.0569
mvU-ct	0.0569	0.0569	0.0569	0.0569
jorU-ct	0.0569	0.0569	0.0569	0.0569
lwU-ct	0.0569	0.0569	0.0569	0.0569
Results for the SPS dataset				
mvU	1.2916	0.4585	0.2688	0.2096
jorU	0.4531	0.2531	0.2147	0.1987
lwU	0.8031	0.3116	0.1737	0.1310
mvU-c	0.2189	0.1662	0.1205	0.1026
jorU-c	0.1413	0.1124	0.1058	0.1022
lwU-c	0.1910	0.1418	0.0967	0.0815
mvU-t	0.0483	0.0483	0.0483	0.0483
jorU-t	0.0483	0.0483	0.0483	0.0483
lwU-t	0.0483	0.0483	0.0483	0.0483
mvU-ct	0.0483	0.0483	0.0483	0.0483
jorU-ct	0.0483	0.0483	0.0483	0.0483
lwU-ct	0.0483	0.0483	0.0483	0.0483
Results for the IND dataset				
mvU	2.4624	1.3659	1.3439	0.9990
jorU	1.4450	1.4532	1.2490	0.9215
lwU	1.7556	0.8337	0.4470	0.2721
mvU-c	0.2157	0.1714	0.1277	0.0936
jorU-c	0.1645	0.1321	0.1016	0.0930
lwU-c	0.2098	0.1675	0.1126	0.0803
mvU-t	0.0299	0.0299	0.0299	0.0299
jorU-t	0.0299	0.0299	0.0299	0.0299
lwU-t	0.0299	0.0299	0.0299	0.0299
mvU-ct	0.0299	0.0299	0.0299	0.0299
jorU-ct	0.0299	0.0299	0.0299	0.0299
lwU-ct	0.0299	0.0299	0.0299	0.0299

Table 12: Average turnover levels for different values of γ

Model	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
Results for the SPL dataset				
mvU	-	-0.2781	-0.1301	-0.0411
jorU	-0.1585	-0.0874	-0.0347	-0.0116
lwU	-0.1711	-0.0402	0.0633	0.1667
mvU-c	0.0849	0.1240	0.1754	0.2223
jorU-c	0.1879	0.2116	0.2234	0.2287
lwU-c	0.0997	0.1369	0.1772	0.2213
mvU-t	0.0616	0.1020	0.1585	0.1605
jorU-t	0.1664	0.1779	0.1623	0.1435
lwU-t	-	0.1197	0.2061	0.2349
mvU-ct	0.1675	0.1764	0.2081	0.2303
jorU-ct	0.2107	0.2198	0.2297	0.2329
lwU-ct	0.1680	0.1798	0.2076	0.2331
Results for the SPS dataset				
mvU	-	-0.0400	0.0153	0.0450
jorU	-0.0484	0.0141	0.0361	0.0500
lwU	-0.1101	-0.0009	0.0531	0.0786
mvU-c	0.0436	0.0514	0.0735	0.0897
jorU-c	0.0601	0.0782	0.0876	0.0923
lwU-c	0.0466	0.0546	0.0754	0.0924
mvU-t	0.0424	0.0743	0.0861	0.0918
jorU-t	0.0796	0.0870	0.0886	0.0903
lwU-t	0.0552	0.0811	0.0960	0.0991
mvU-ct	0.0468	0.0685	0.0822	0.0919
jorU-ct	0.0763	0.0849	0.0904	0.0929
lwU-ct	0.0532	0.0727	0.0847	0.0955
Results for the IND dataset				
mvU	-0.1337	-0.0793	-0.1102	-0.0727
jorU	-0.0944	-0.1178	-0.1094	-0.0607
lwU	-0.0614	0.0061	0.0514	0.0753
mvU-c	0.0490	0.0685	0.0932	0.1116
jorU-c	0.0656	0.0878	0.1031	0.1102
lwU-c	0.0506	0.0751	0.1022	0.1205
mvU-t	0.0544	0.1287	0.1527	0.1686
jorU-t	0.1399	0.1505	0.1610	0.1732
lwU-t	-0.0648	0.0668	0.1144	0.1416
mvU-ct	0.0241	0.0807	0.1190	0.1319
jorU-ct	0.0870	0.1176	0.1272	0.1317
lwU-ct	0.0221	0.0897	0.1277	0.1384

Table 13: Sharpe ratios adjusted for transaction costs for different values of γ

Appendix D: Additional figures

Equity lines for the "SPL" dataset, M=60

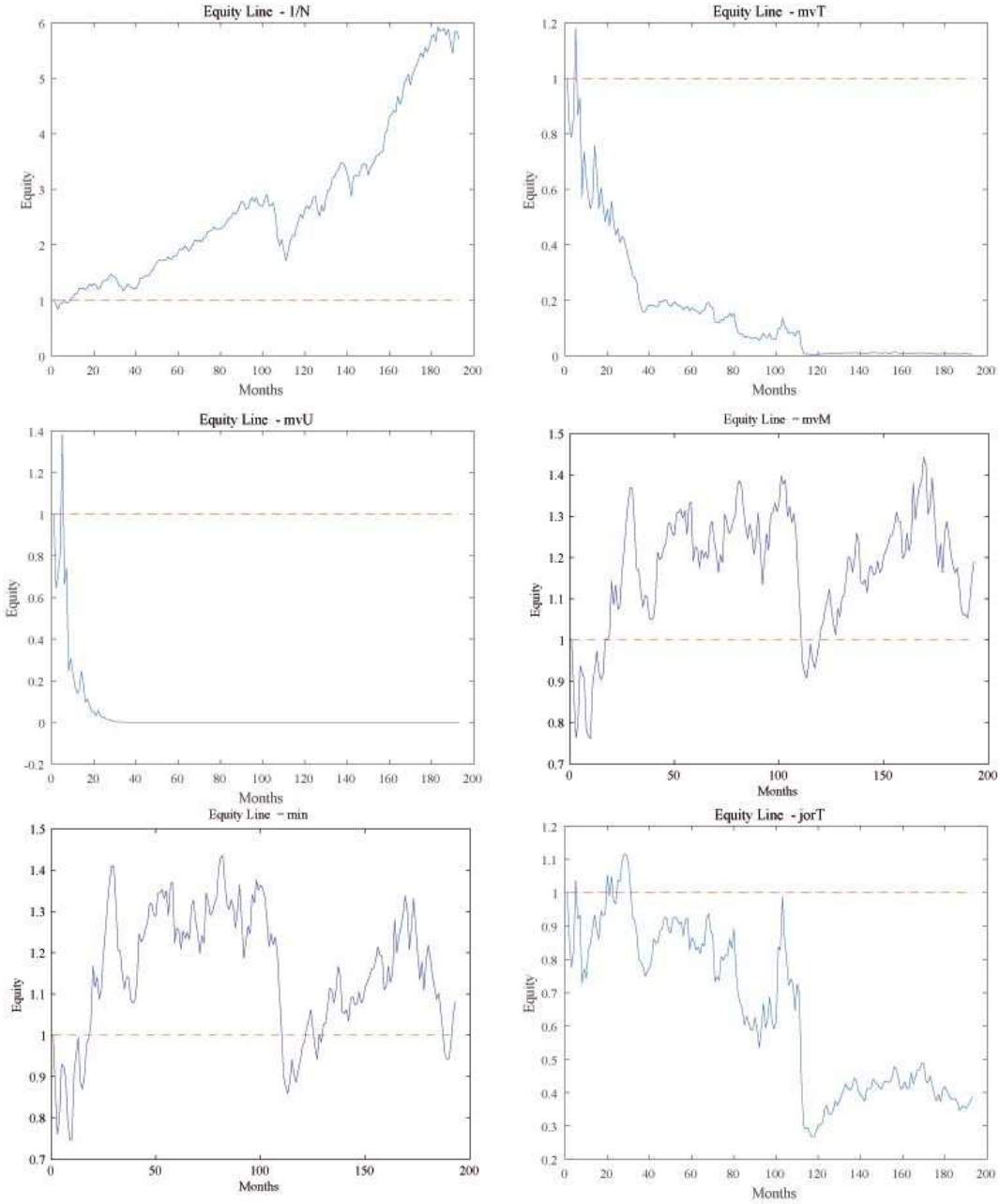

Figure 6: Equity lines for the "SPL" datasets (1)

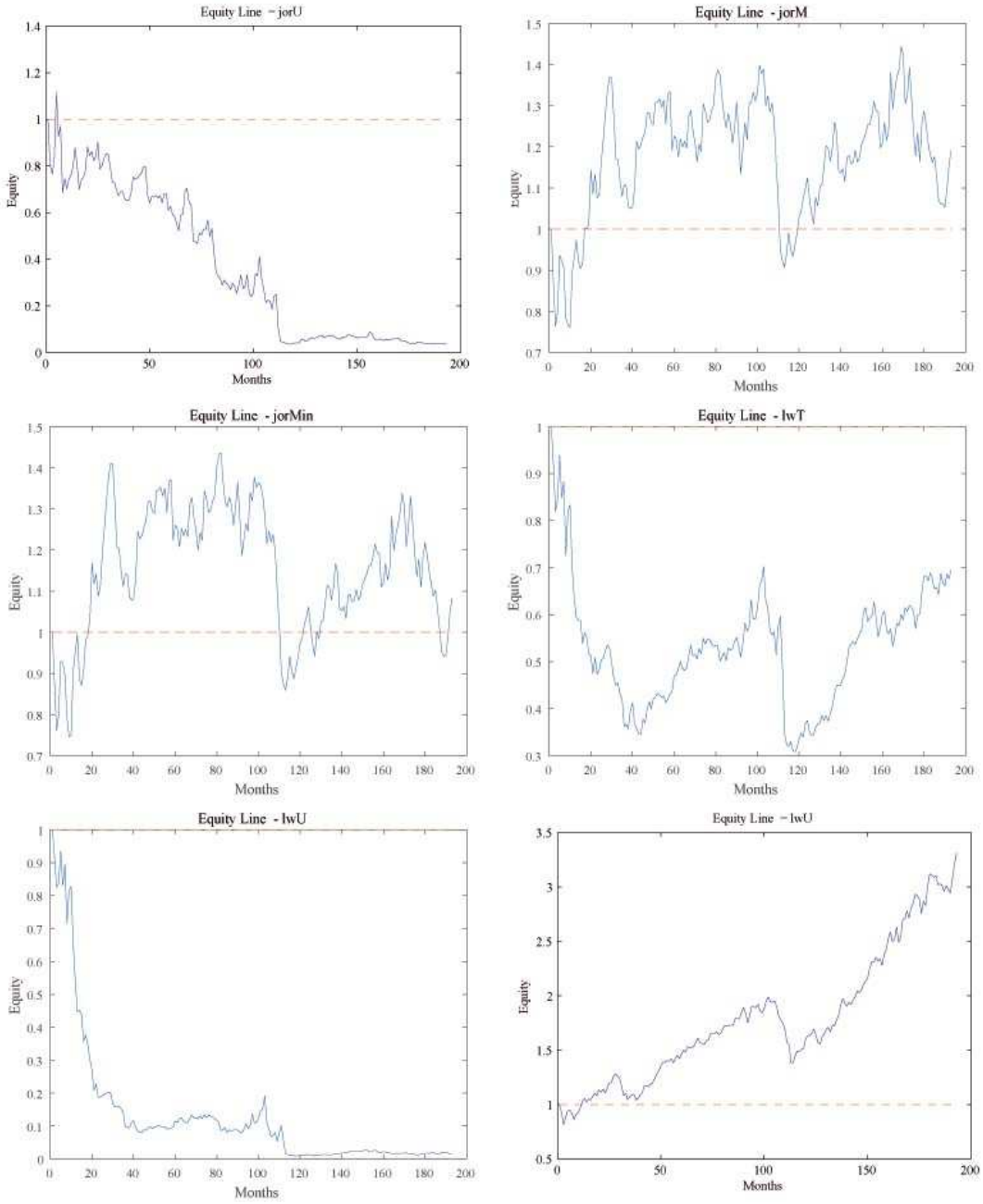

Figure 7: Equity lines for the "SPL" datasets (2)

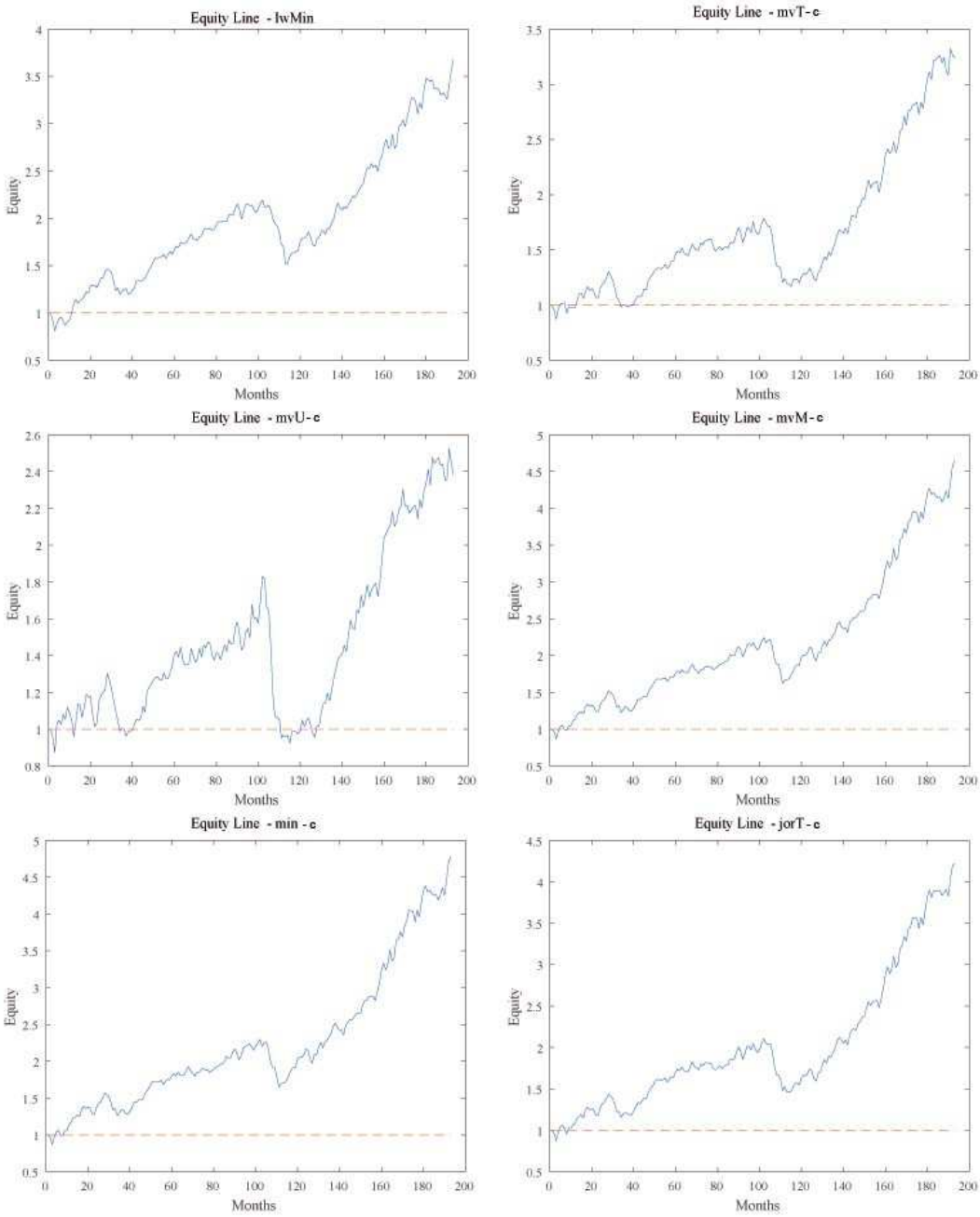

Figure 8: Equity lines for the "SPL" datasets (3)

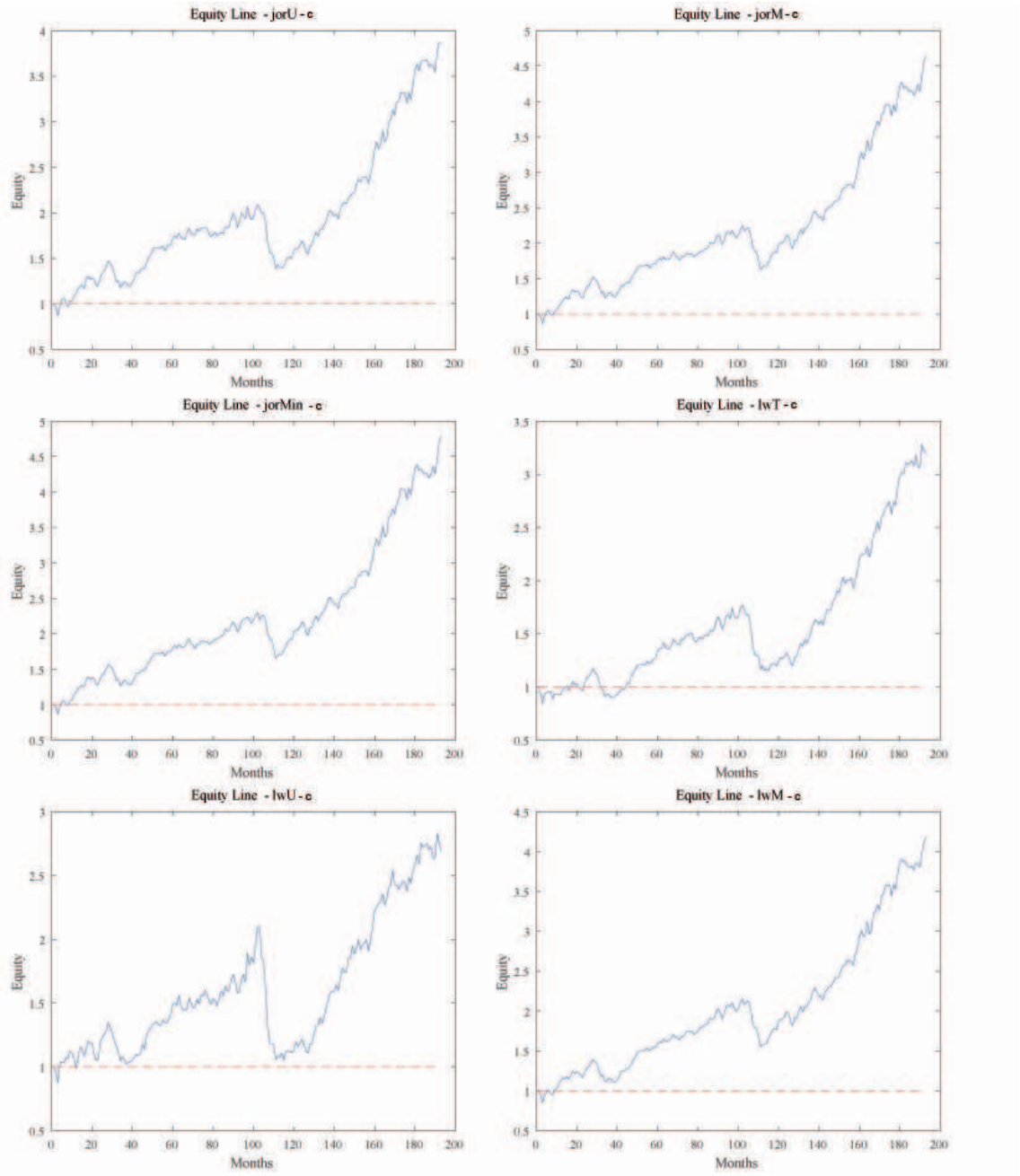

Figure 9: Equity lines for the "SPL" datasets (4)

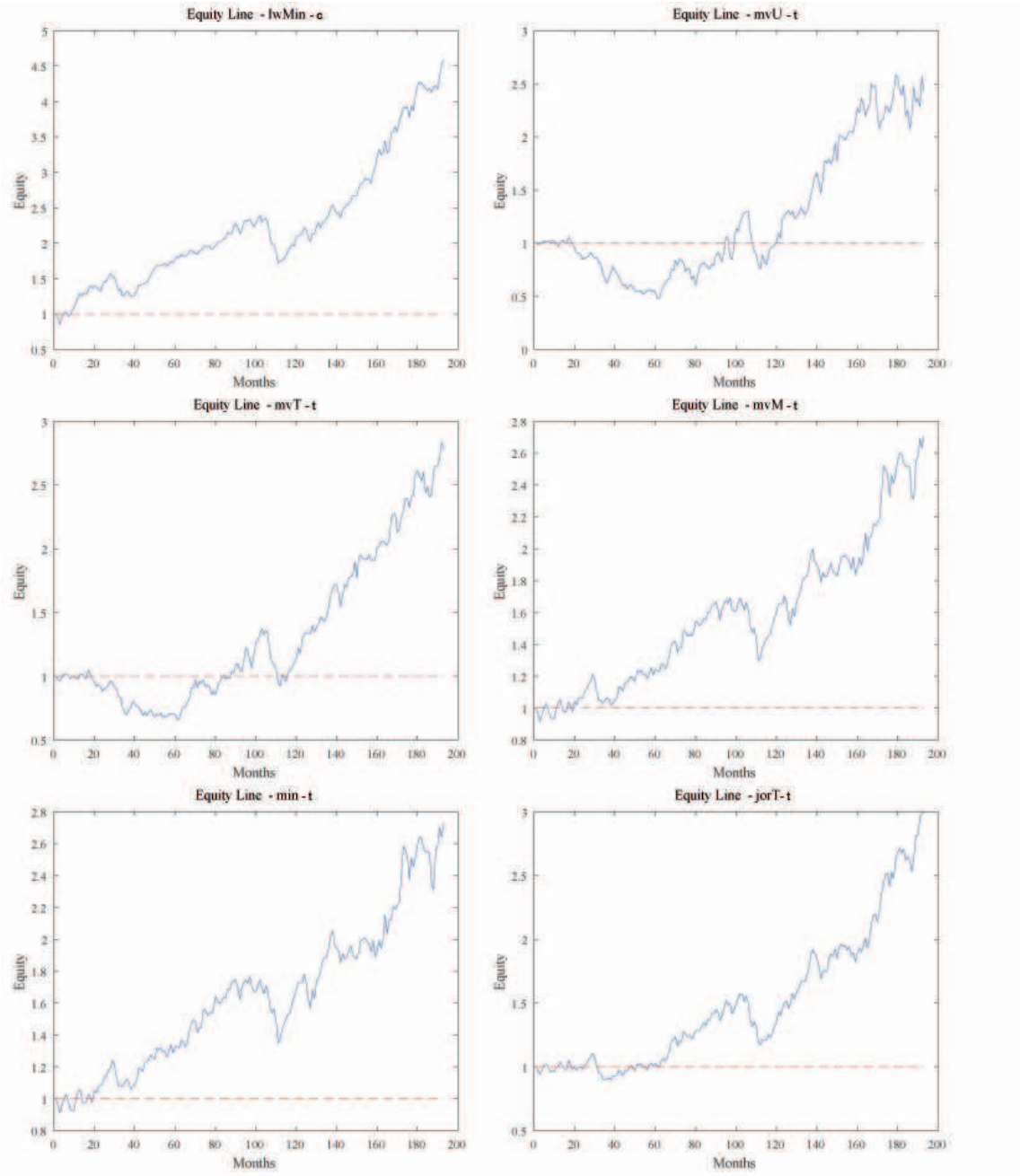

Figure 10: Equity lines for the "SPL" datasets (5)

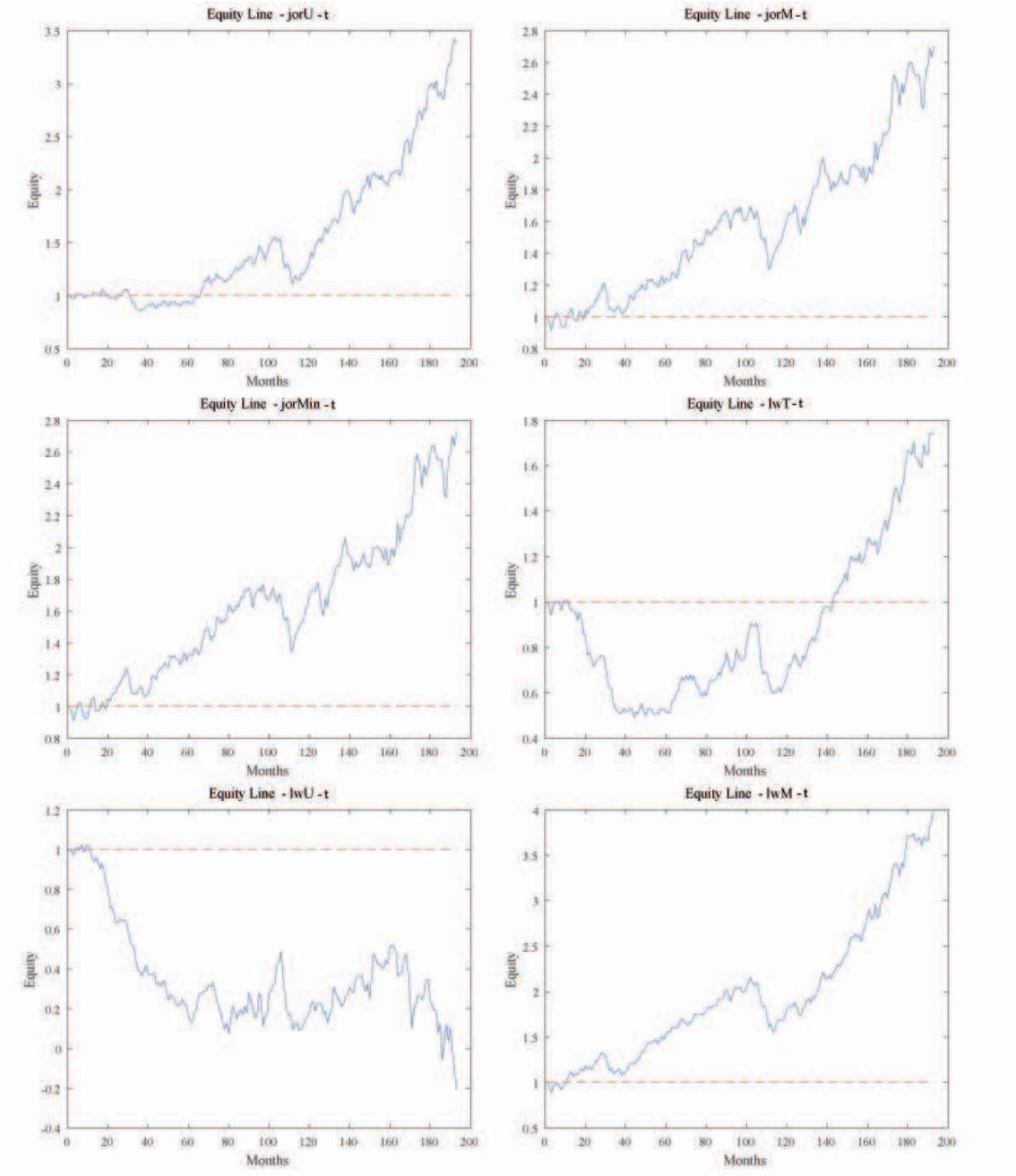

Figure 11: Equity lines for the "SPL" datasets (6)

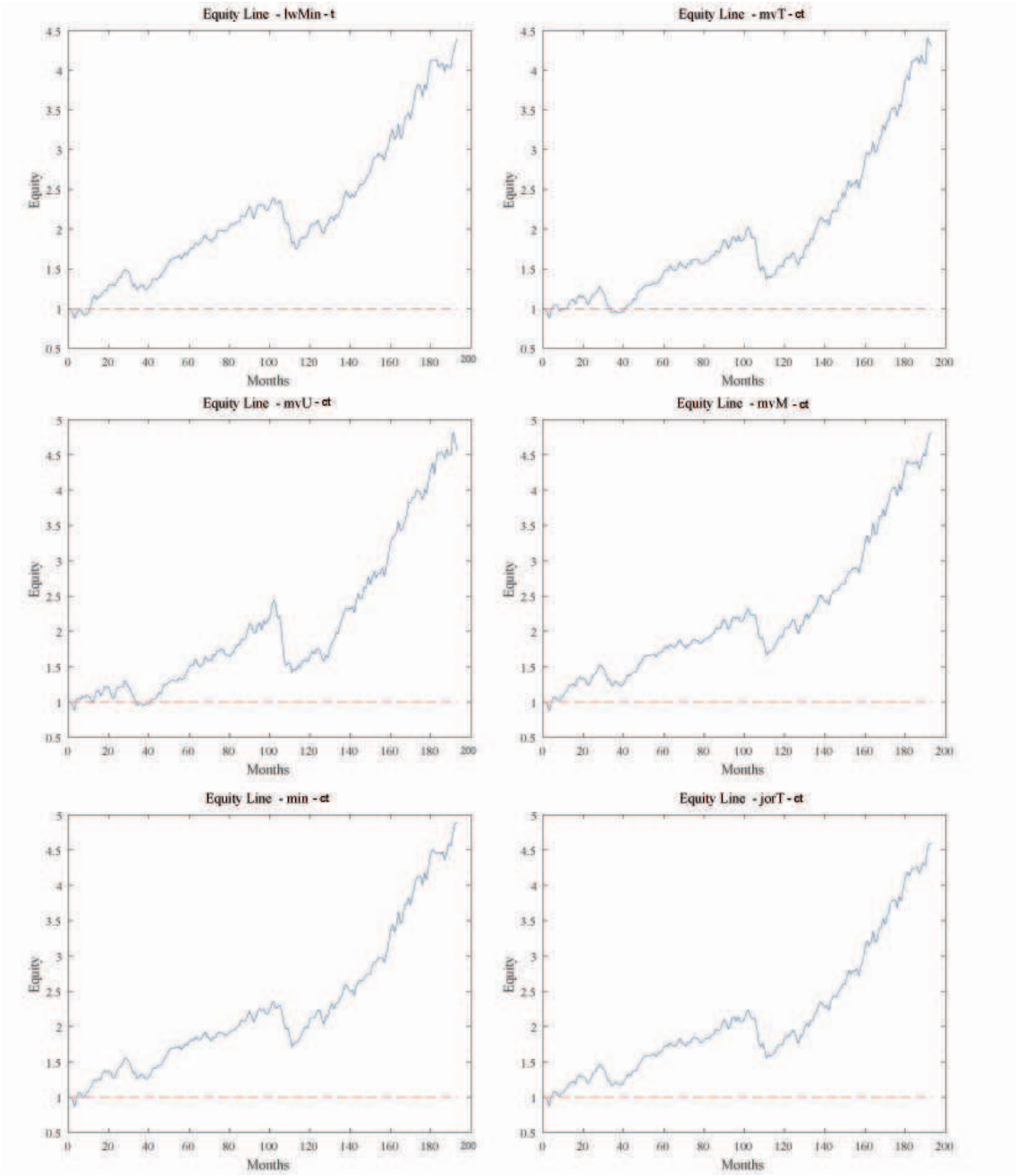

Figure 12: Equity lines for the "SPL" datasets (7)

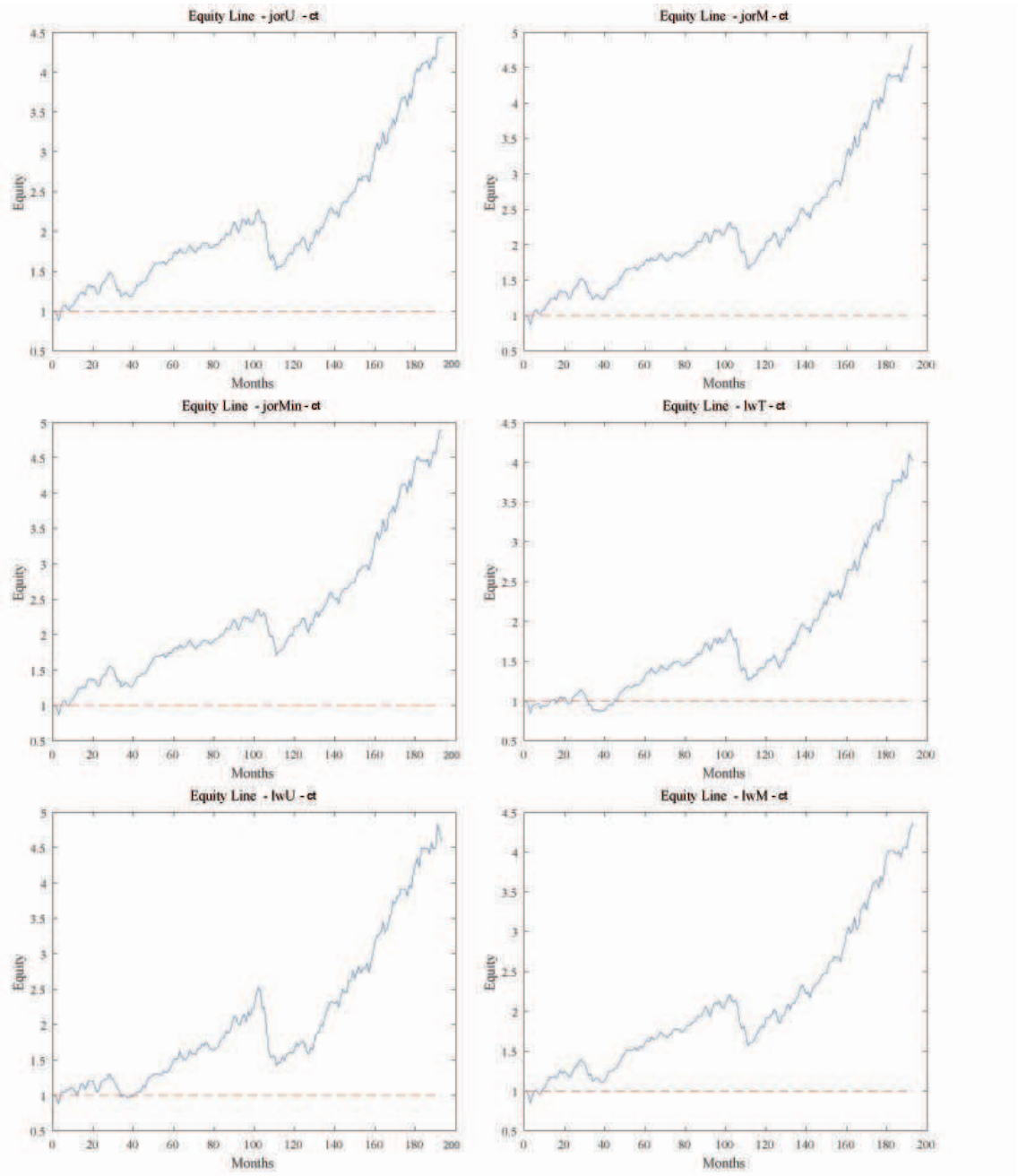

Figure 13: Equity lines for the "SPL" datasets (8)

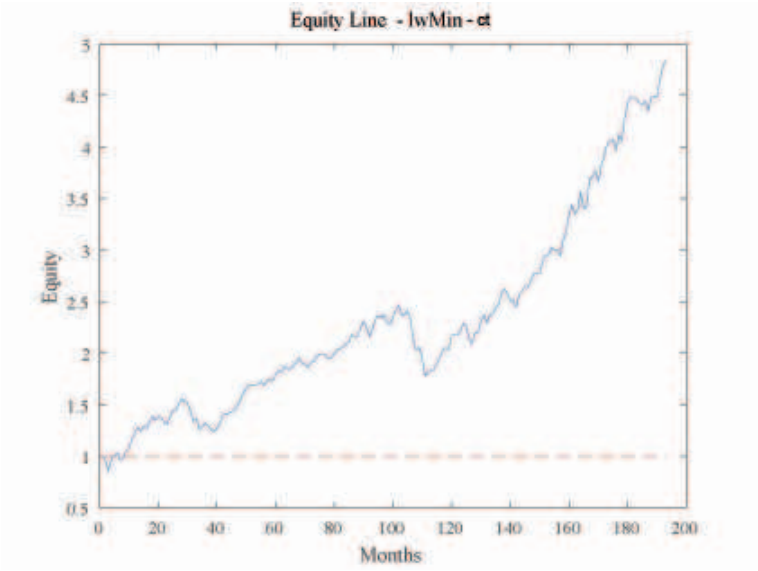

Figure 14: Equity lines for the "SPL" datasets (9)

Equity lines for the "SPS" dataset, M=60

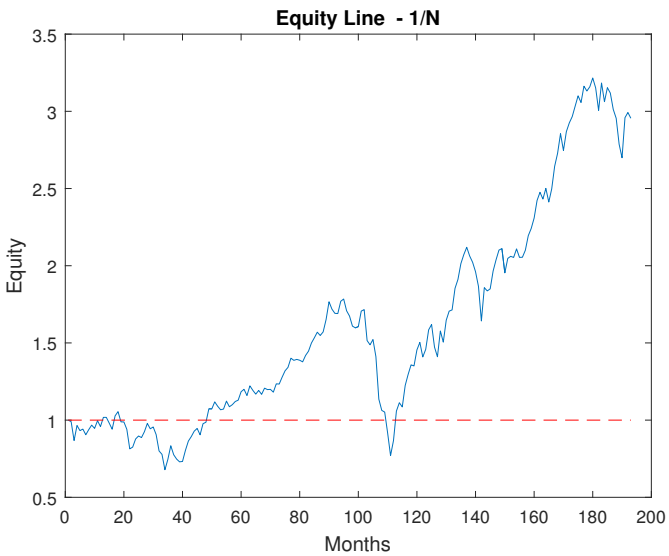

Figure 15: Equity lines for the "SPS" datasets (1)

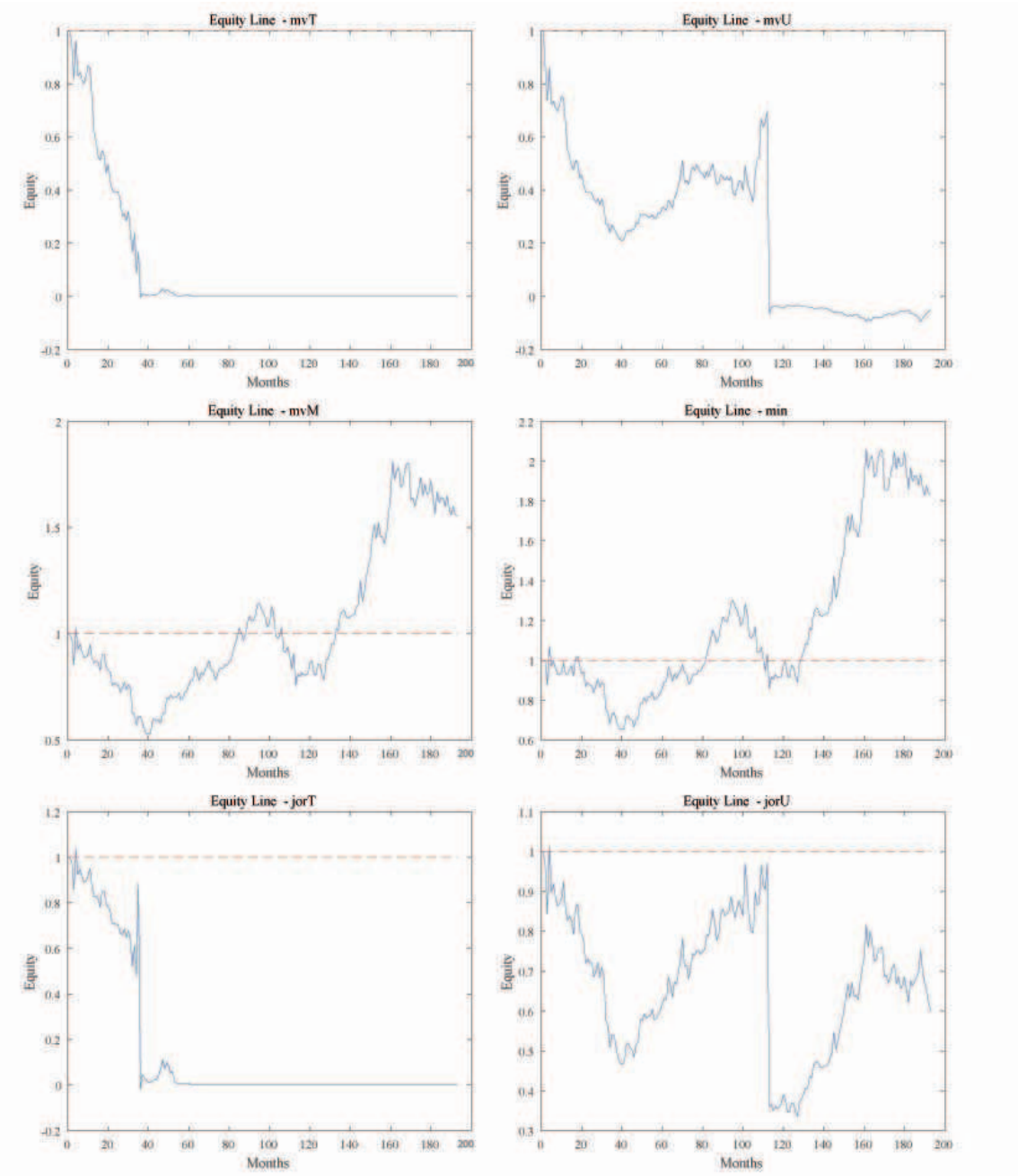

Figure 16: Equity lines for the "SPS" datasets (2)

Equity lines for the "IND" dataset, M=60

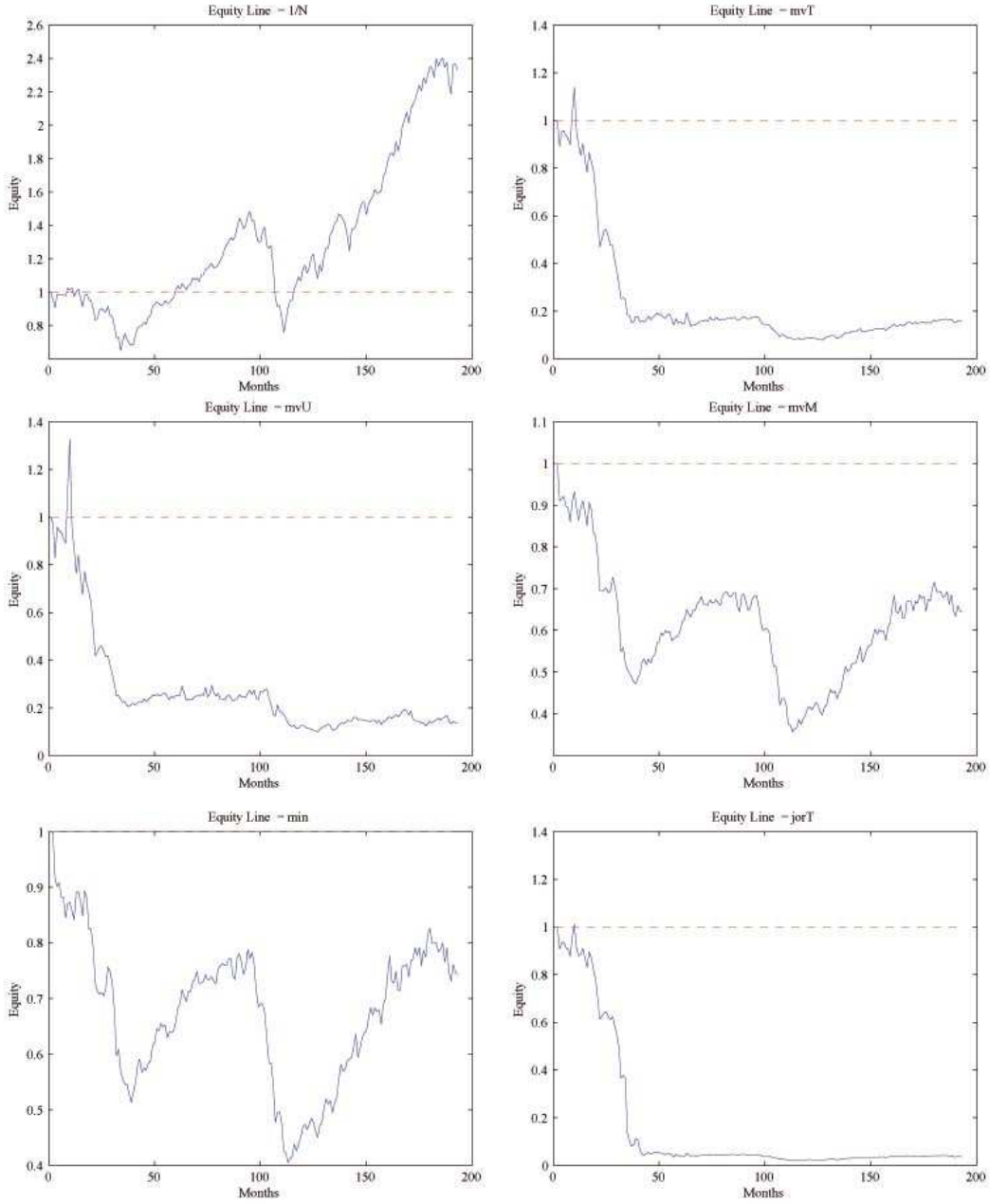

Figure 24: Equity lines for the "IND" datasets (1)

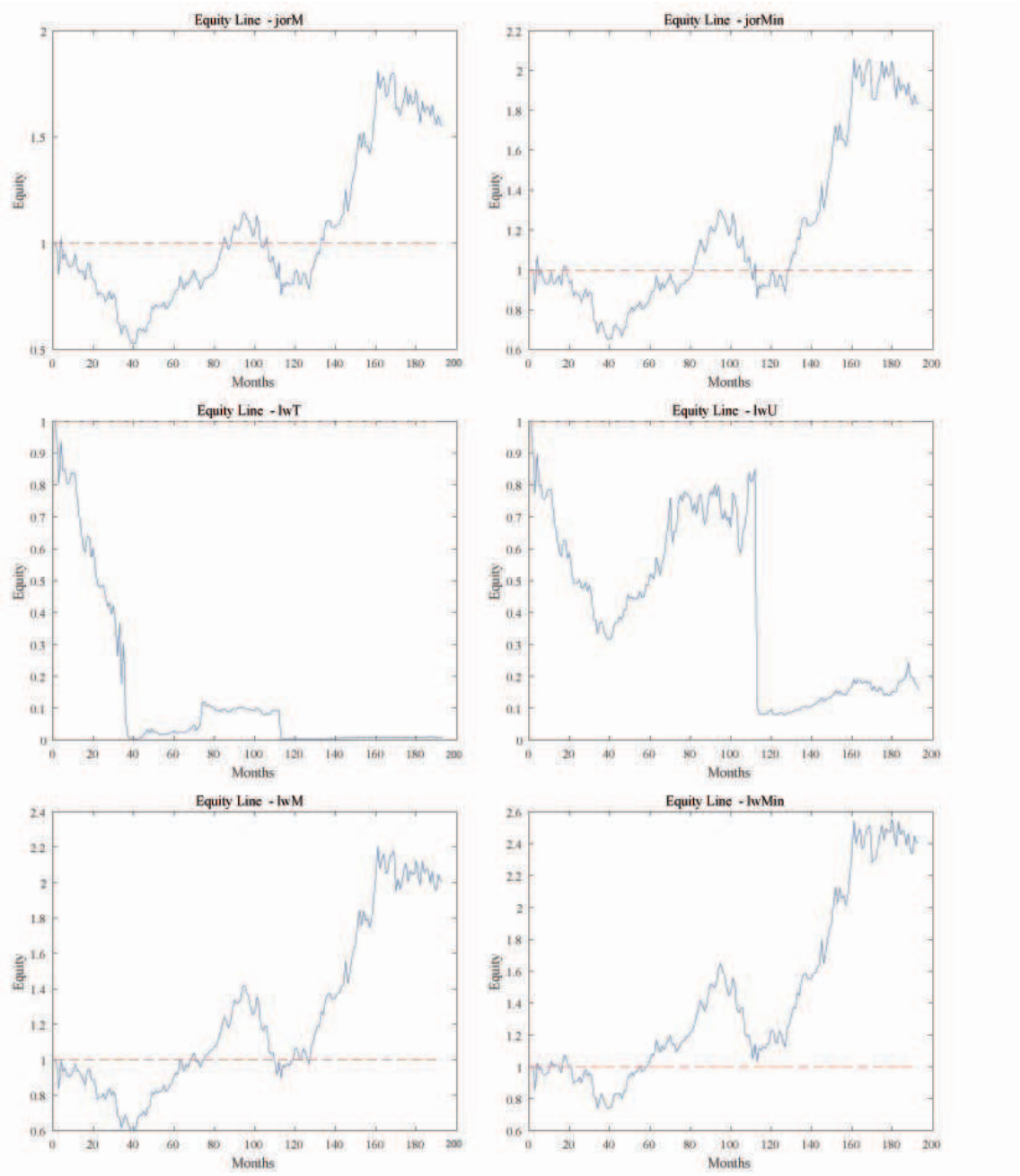

Figure 17: Equity lines for the "SPS" datasets (3)

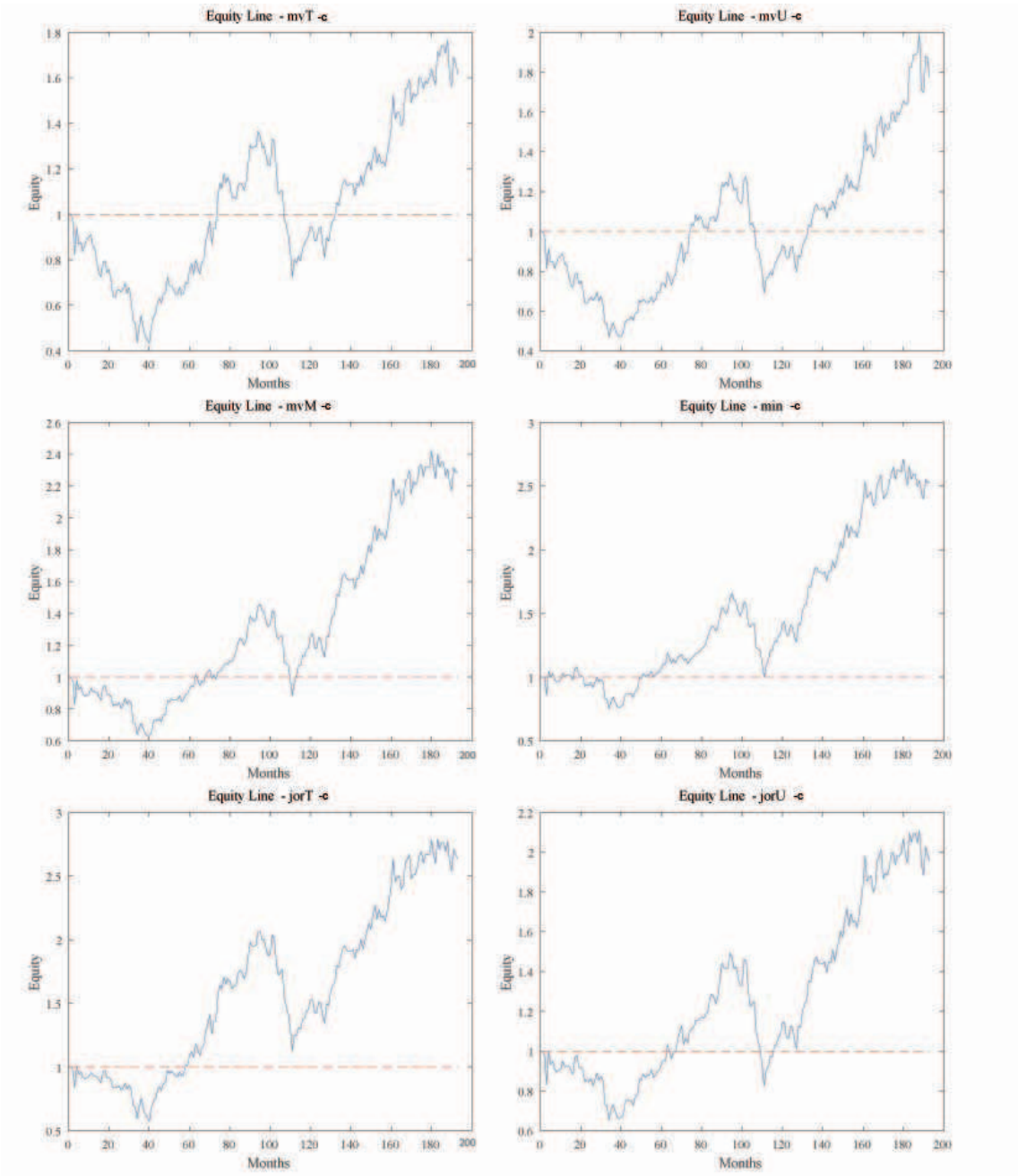

Figure 18: Equity lines for the "SPS" datasets (4)

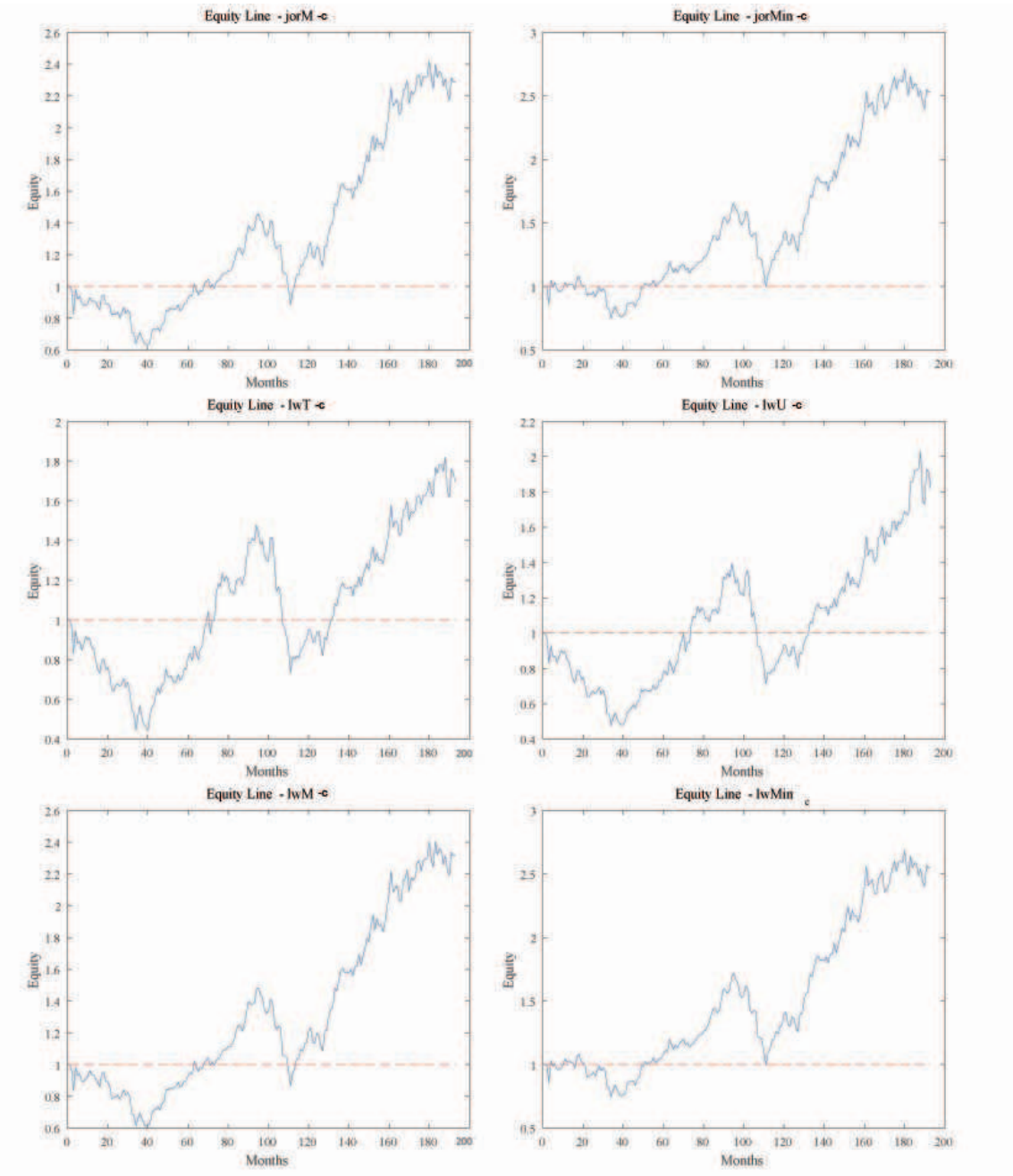

Figure 19: Equity lines for the "SPS" datasets (5)

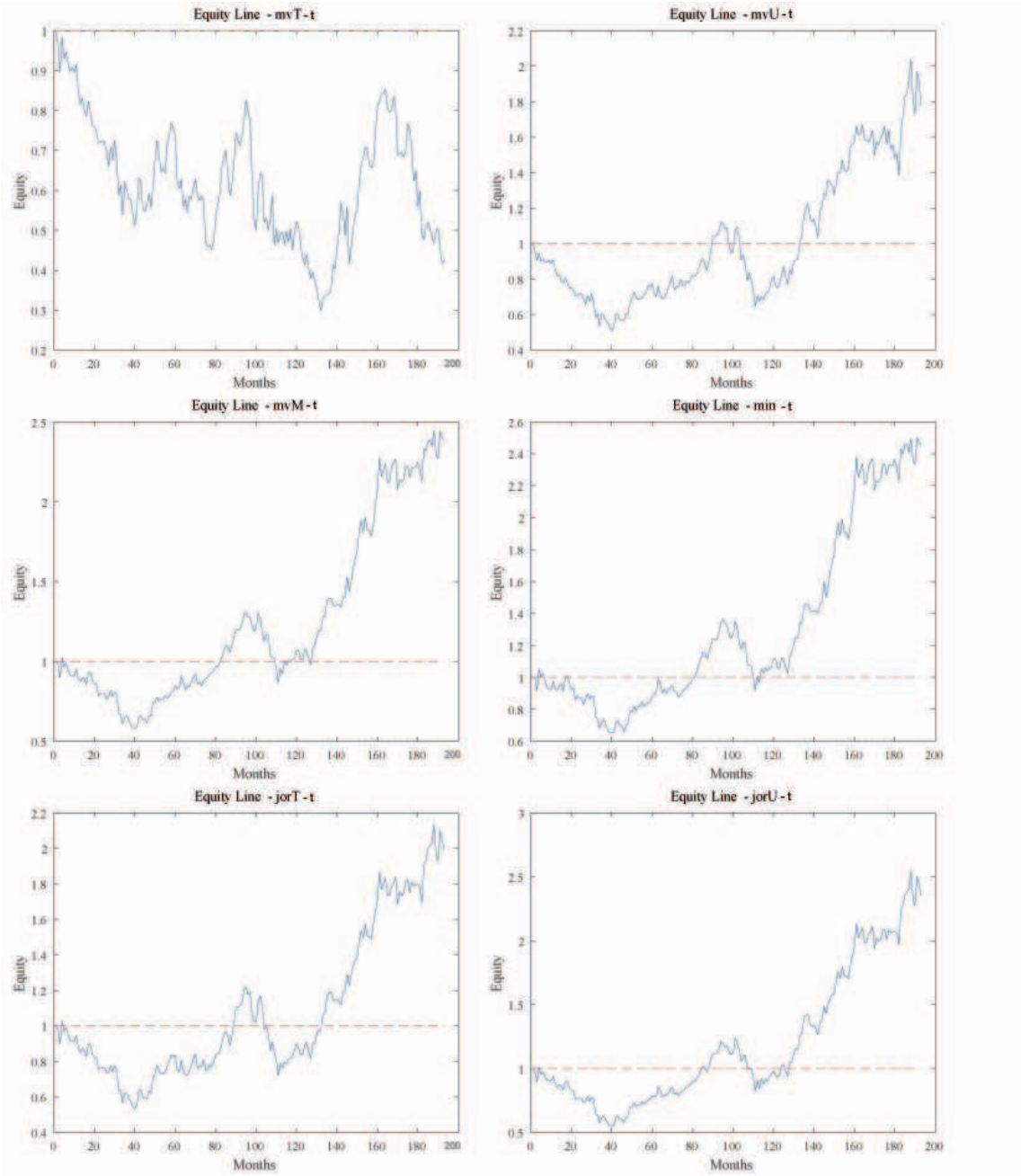

Figure 20: Equity lines for the "SPS" datasets (6)

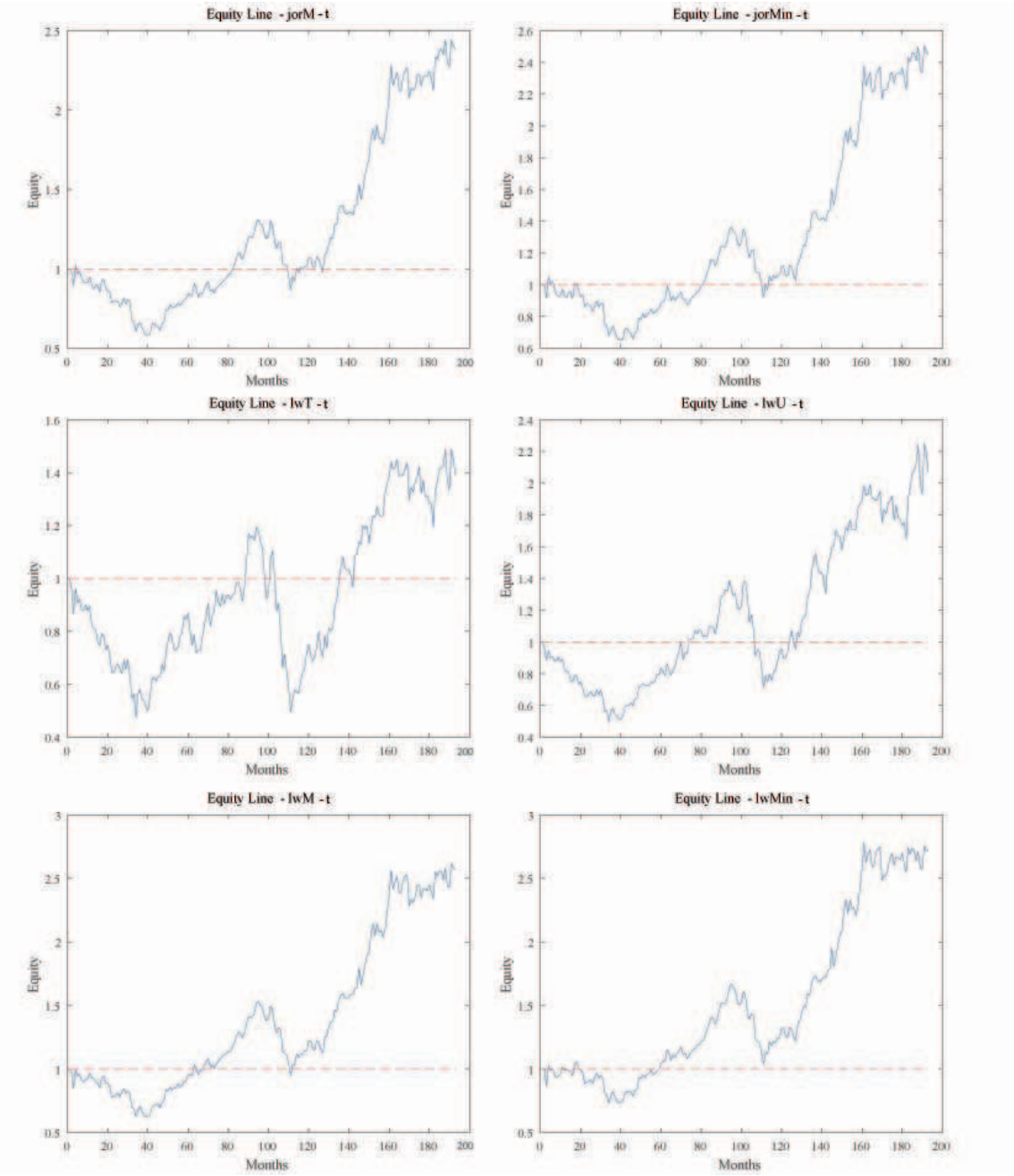

Figure 21: Equity lines for the "SPS" datasets (7)

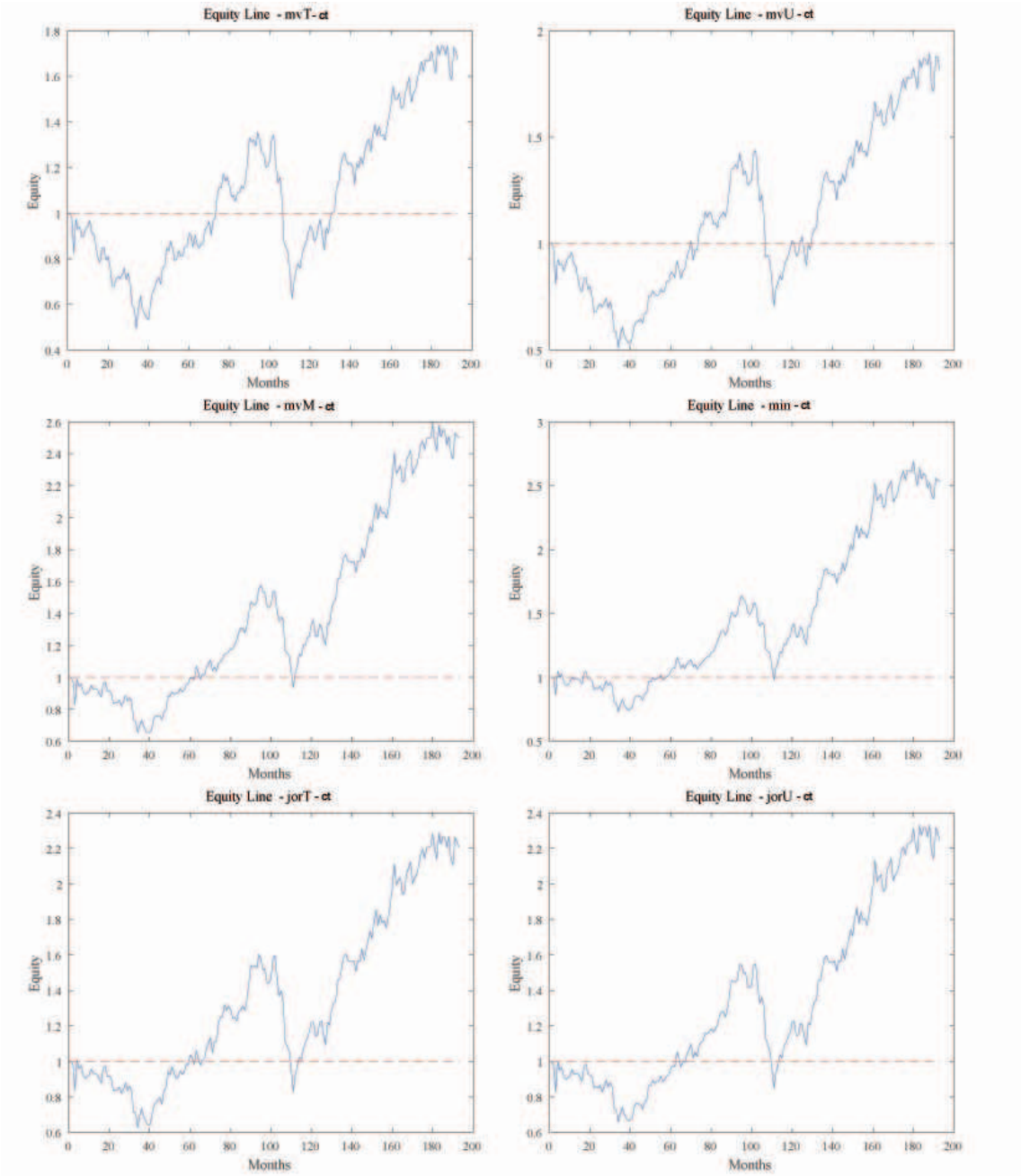

Figure 22: Equity lines for the "SPS" datasets (8)

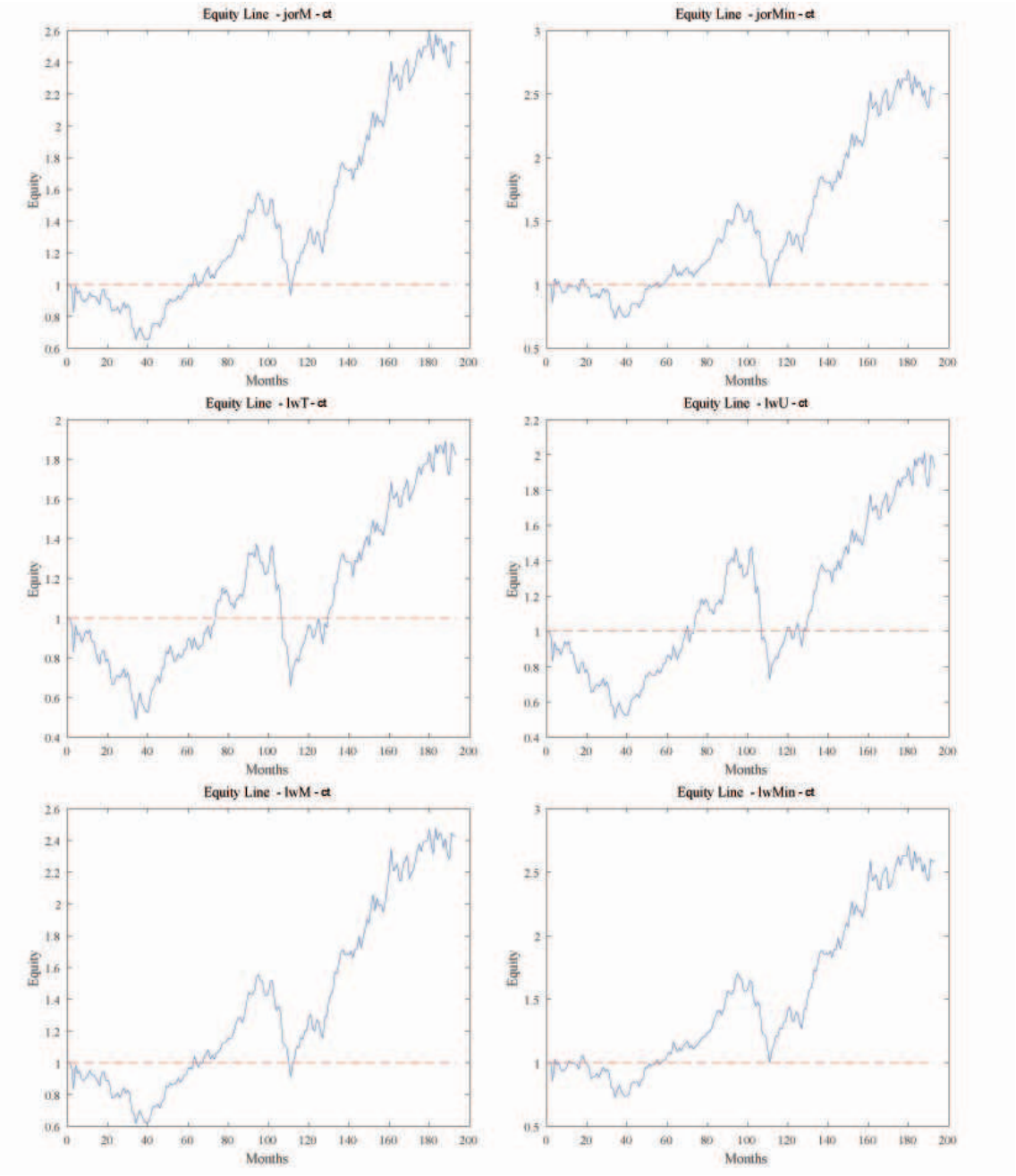

Figure 23: Equity lines for the "SPS" datasets (9)

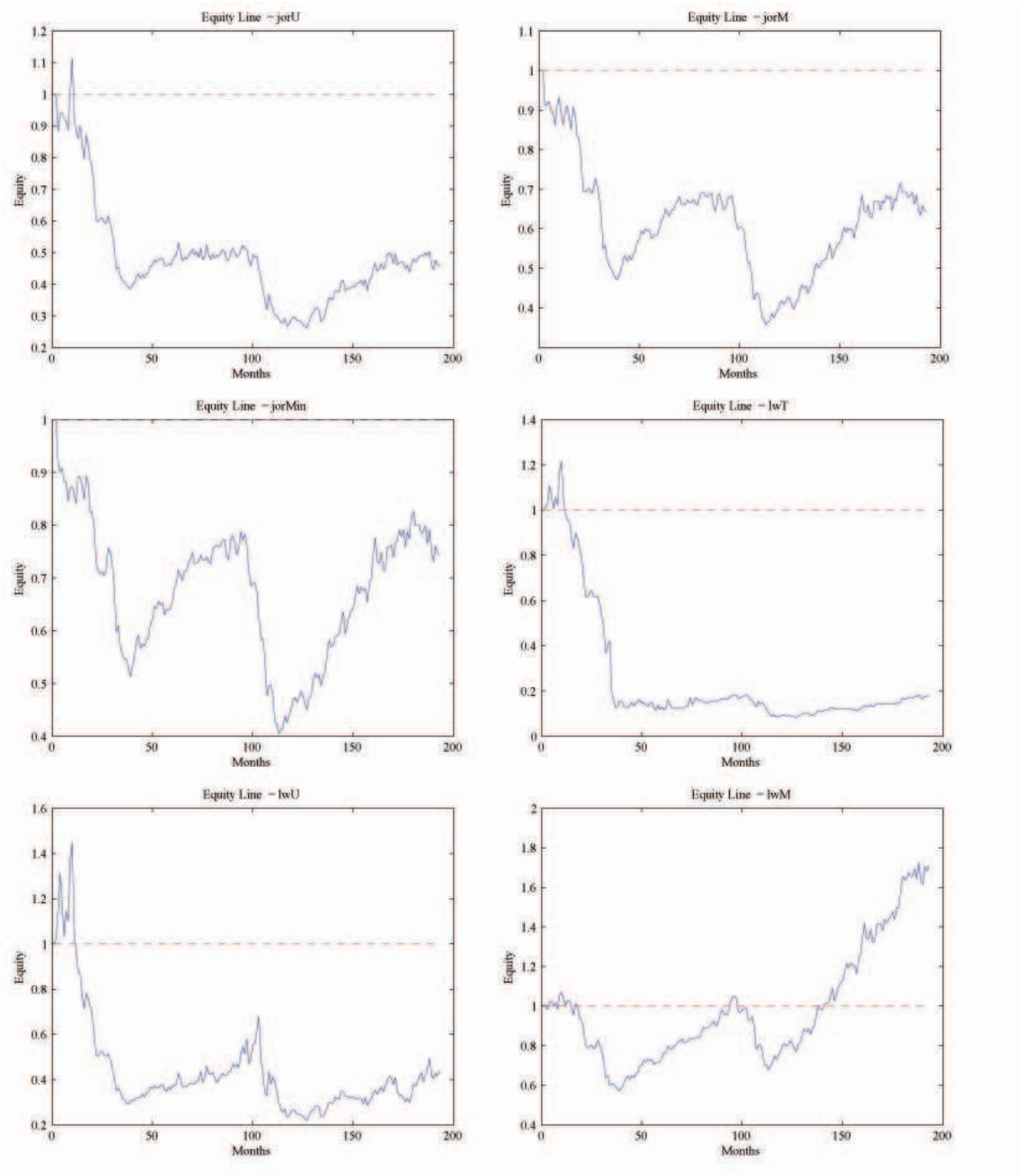

Figure 25: Equity lines for the "IND" datasets (2)

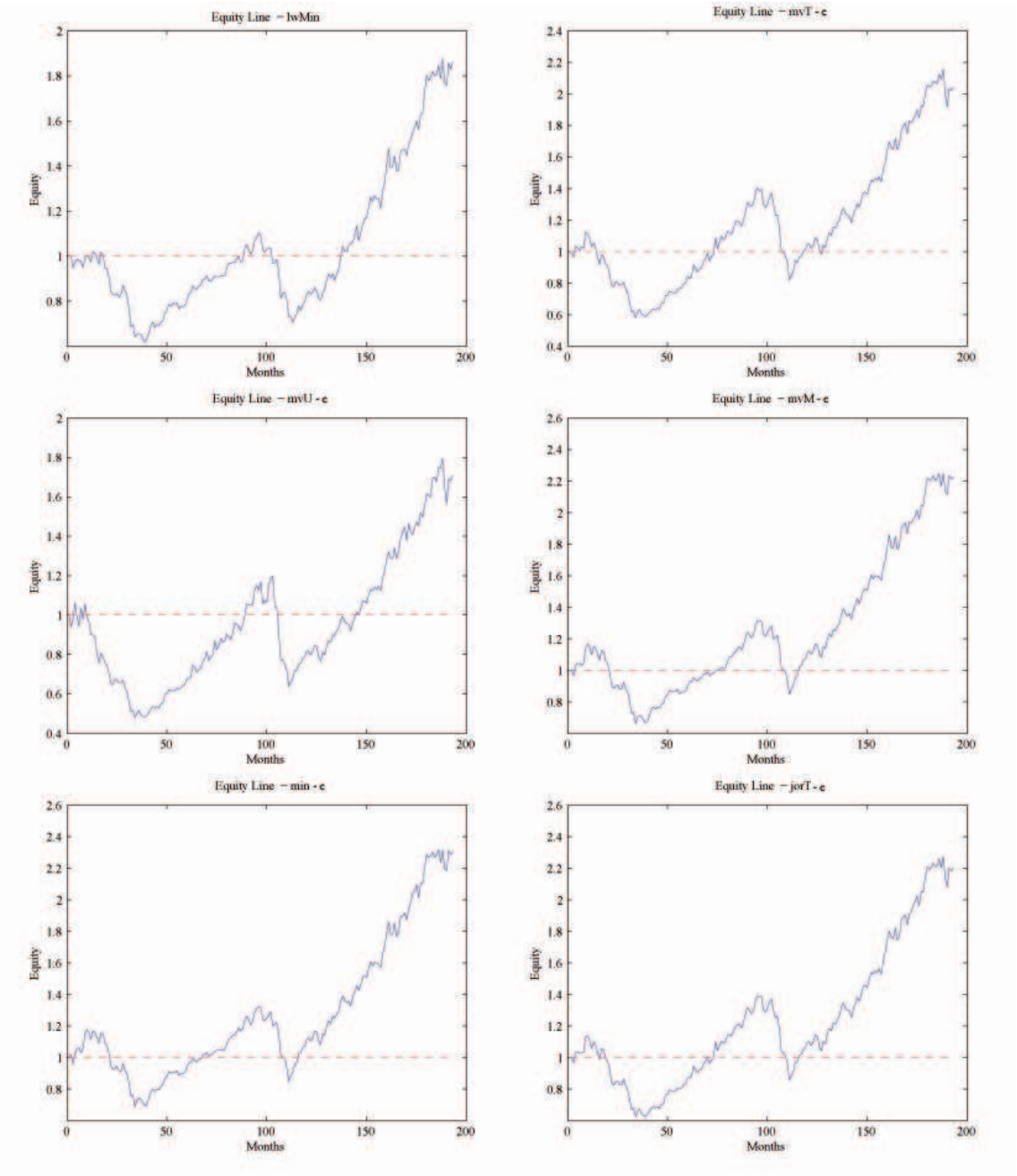

Figure 26: Equity lines for the "IND" datasets (3)

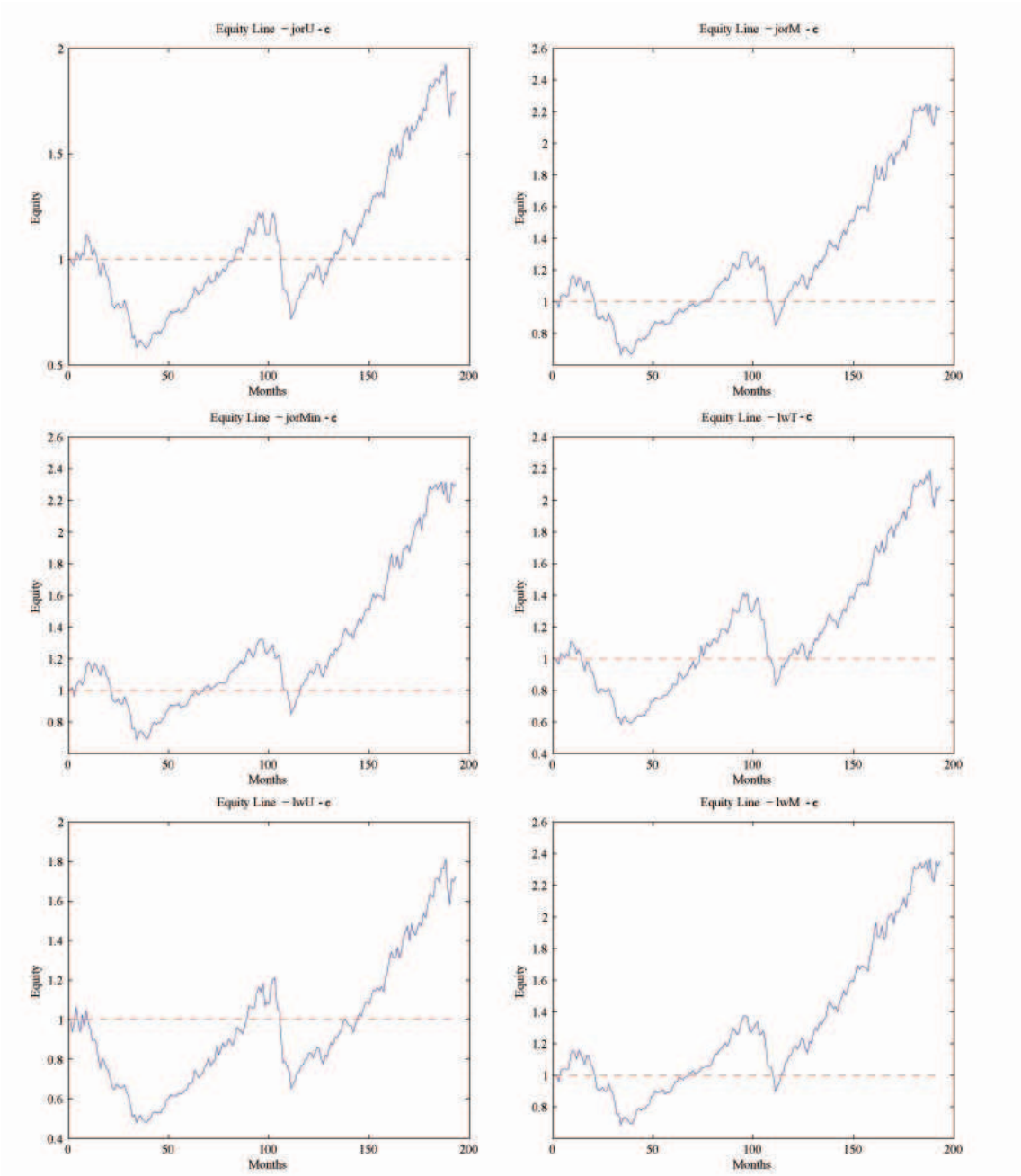

Figure 27: Equity lines for the "IND" datasets (4)

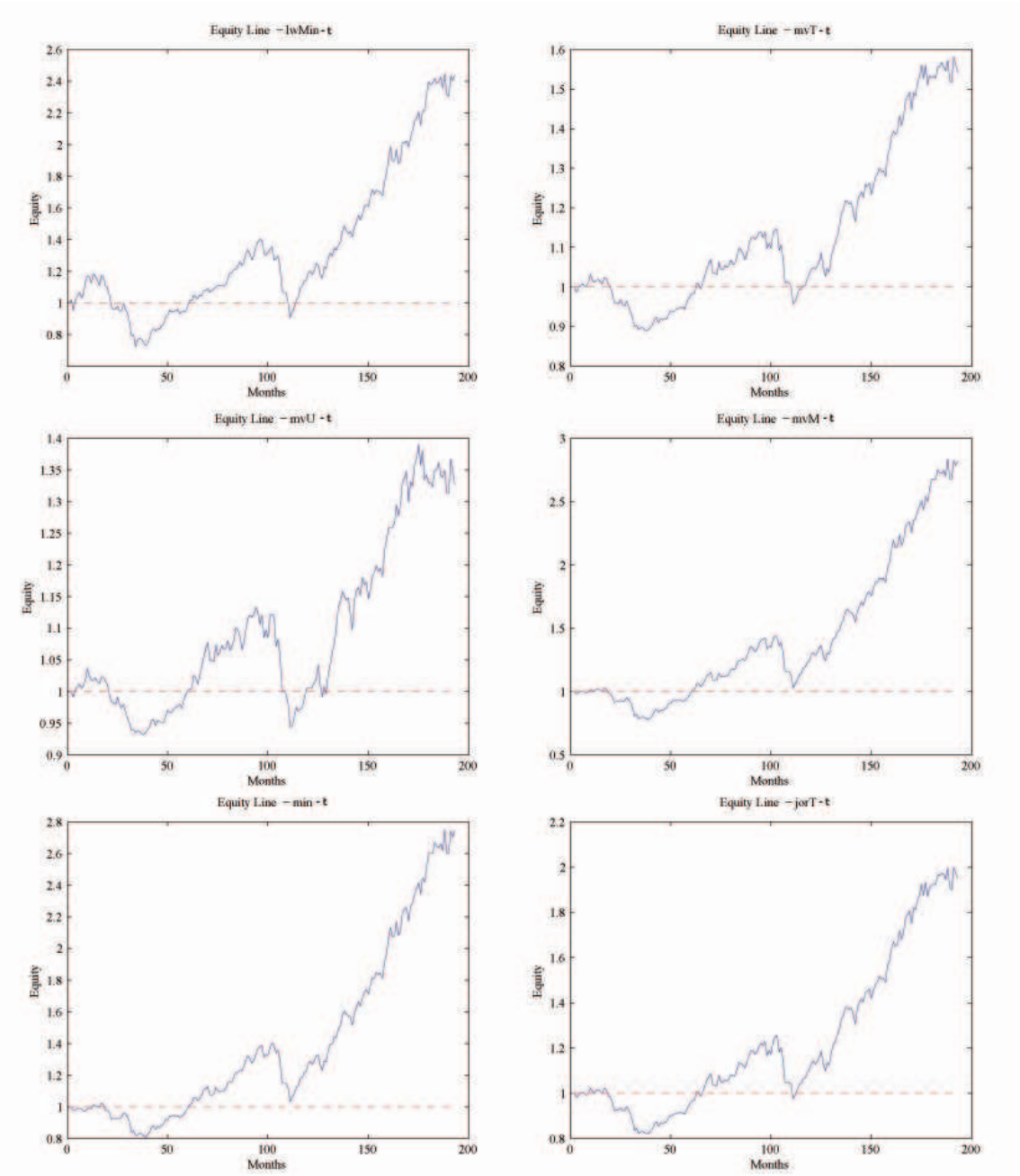

Figure 28: Equity lines for the "IND" datasets (5)

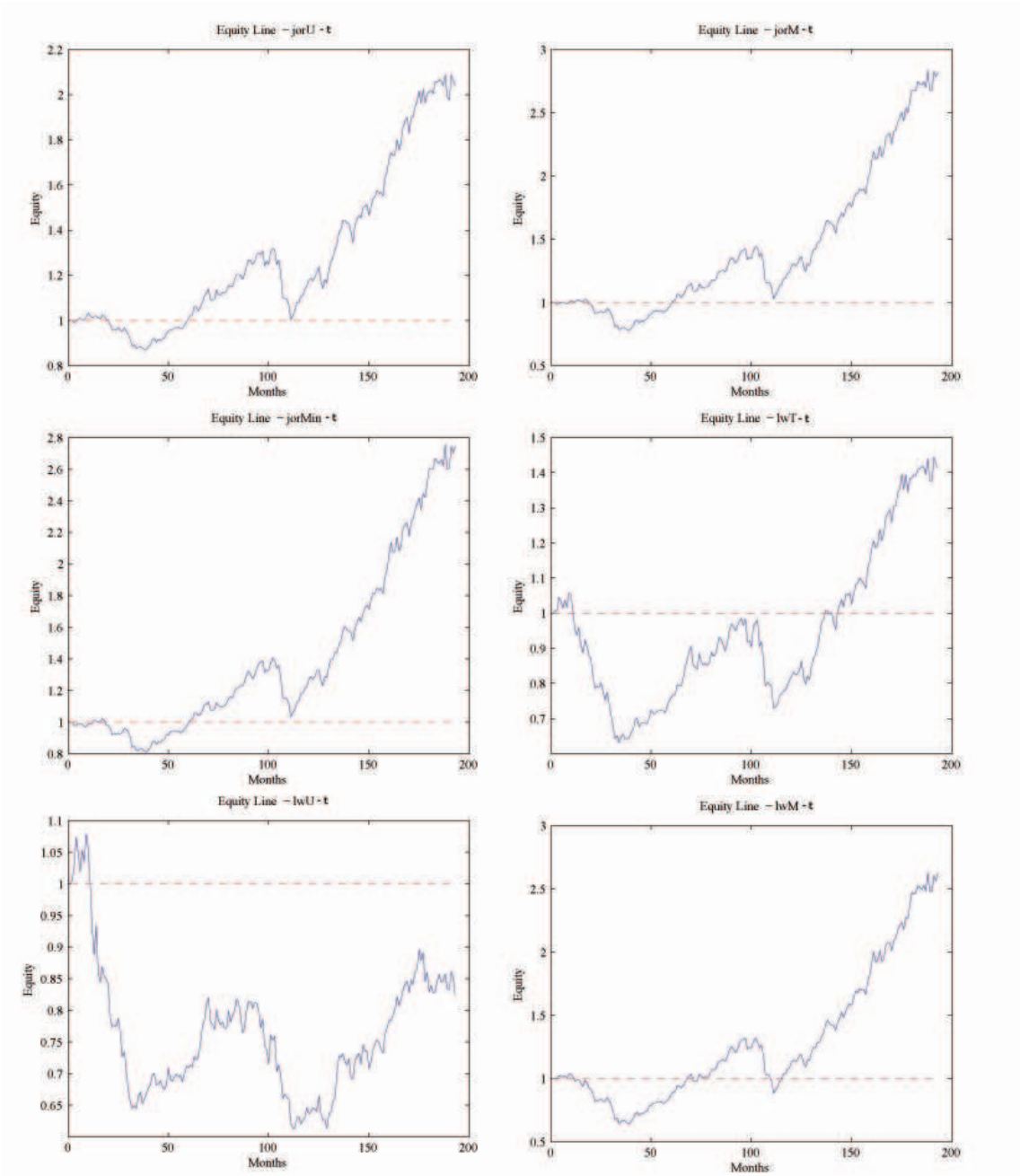

Figure 29: Equity lines for the "IND" datasets (6)

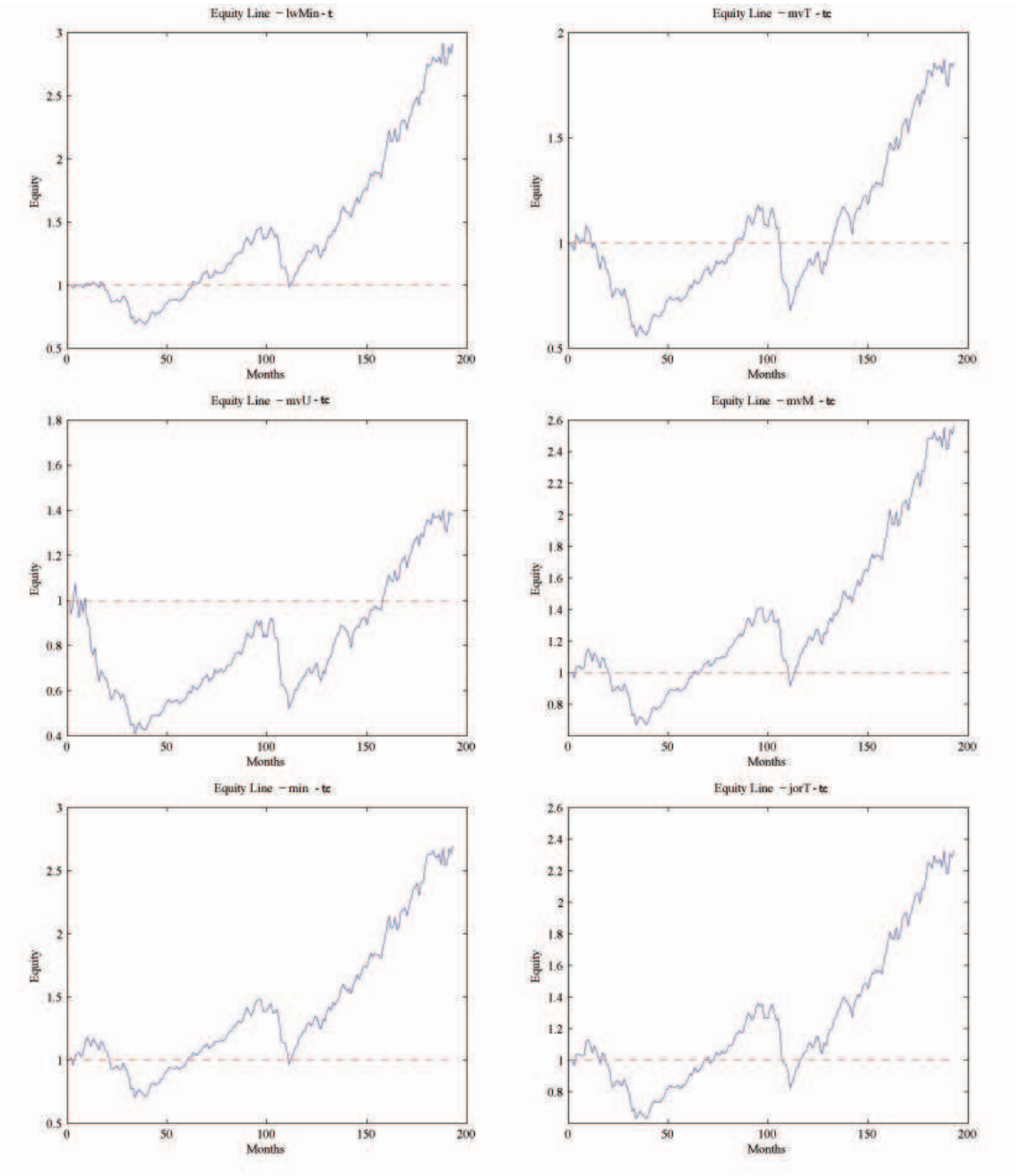

Figure 30: Equity lines for the "IND" datasets (7)

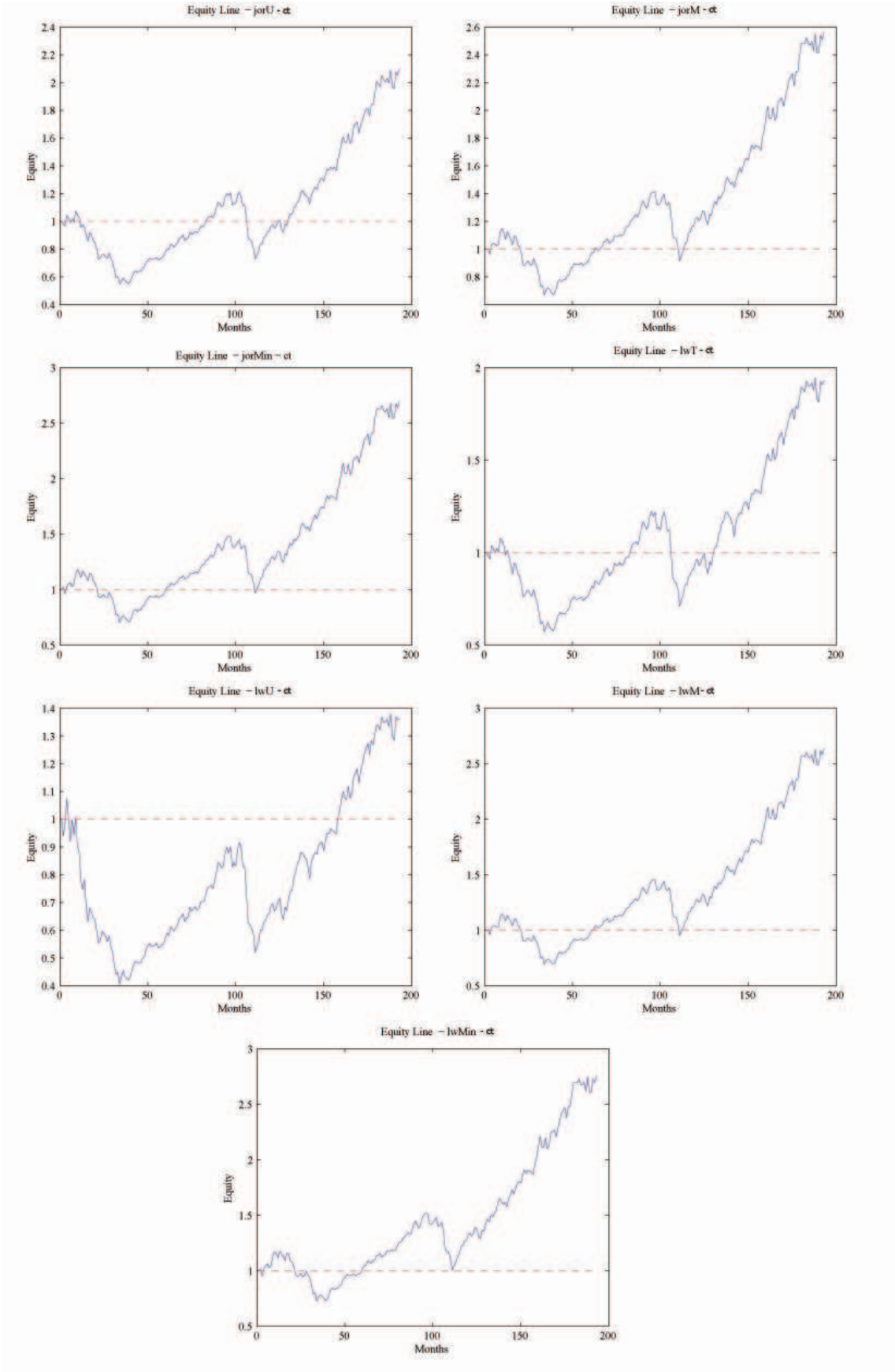

Figure 31: Equity lines for the "IND" datasets (8)

References

[1] Acerbi C., D. Tasche (2002), “On the Coherence of Expected Shortfall”, Journal of Banking and Finance, 26, pp. 1487–1503.

[2] Balduzzi P., A.W. Lynch (1999), “Transaction Costs and Predictability: Some Utility Cost Calculations”, *Journal of Financial Economics*, 52, pp. 47–78.

[3] Barone L. (2006), “Bruno de Finetti and the Case of the Critical Line’s Last Segment”, *Bruno de Finetti Centenary Conference*, Accademia dei Lincei, Rome.

[4] Benartzi S., R. H. Thaler (2001), "Naïve Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review*, 91, pp. 79-98.

[5] Bertsimas D., G.J. Lauprete, A. Samarov (2004), “Shortfall as a risk measure:properties, optimization and applications, *Journal of Economic Dynamics & Control*, No. 28, pp. 1353-1381.

[6] Bhardwaj R., L. Brooks (1992), “The January Anomaly: Effects of Low Share Price, Transaction Costs, and Bid-Ask Bias”, *Journal of Finance*, 47, pp. 553–574.

[7] Bucciol A., R. Miniaci (2007), “Risk aversion and portfolio choice”, *Working paper.*

[8] Cambanis S., S. Huang, G. Simons (1981), "On the Theory of Elliptically Contoured Distributions", *Journal of Multivariate Analysis*, 11, pp. 368–385.

[9] Chapados N., Y. Bengio (2001), “Cost functions and model combination for VaR-based asset allocation using neural networks”,*IEEE Transactions on Neural Networks*, Vol. 12, pp. 890-906.

[10] Chen W., H. Chung, K. Ho, T. Hsu (2010), “Portfolio optimization models and mean-variance spanning tests”, *Handbook of Quantitative Finance and Risk Management*, ch.10, pp. 165-184.

[11] Cho D.D. (2003), “Uncertainty in Second Moments: Implications for Portfolio Allocation”, *Working Paper*, SUNY at Buffalo.

[12] Chopra V.K. (1993), “Improving optimization”, *Journal of investing*, 8, pp. 51-59.

[13] Coqueret G. (2015), “Diversified minimum-variance portfolios”, *Annals of Finance*, Volume 11, Issue 2, pp. 221-241.

[14] De Giorgi E. (2002) “A Note on Portfolio Selection under Various Risk Measures”, *Working paper No. 9*, National Centre of Competence in Research, Financial Valutation and Risk Management, Zurich.

[15] DeMiguel V., L. Garlappi, R. Uppal (2009a) "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy", *Review of Financial Studies*, 22, pp. 1915–1953.

[16] DeMiguel V., L. Garlappi, F.J. Nogales, R. Uppal (2009b) "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portoflio Norms", *Management Science*, Vol. 55, No. 5, pp. 798-812.

[17] Farber H.S. (1978), "Individual Preferences and Union Wage Determination: the Case of the United Mine Workers", *Journal of Political Economy*, Vol.86(5), pp. 923-942.

[18] Efron B., C. N. Morris (1977), "Stein's Paradox in Statistics", *Scientific American*, 236(5), pp. 119-127.

[19] Fama E.F., K.R. French (1992), "The Cross-Section of Expected Stock Returns", *The Journal of Finance*, 47, pp. 427–465.

[20] Frankfurter G.M., H.E. Phillips, J.P. Seagle (1971), "Portfolio Selection: The Effects of Uncertain Means, Variances, and Covariances", *The Journal of Financial and Quantitative Analysis*, Vol.6, No.5, pp. 1251-1262.

[21] Frazzini A., L.H. Pedersen (2011), "Betting Against Beta", *Swiss Finance Institute Research Paper*, No. 12-17.

[22] Frost P.A. and J.E. Savarino (1986), "An Empirical Bayes Approach to Efficient Portfolio Selection", *Journal of Financial and Quantitative Analysis*, Vol. 21, No. 3.

[23] Frost P.A., J. E. Savarino (1988), "For Better Performance: Constrain Portfolio Weights", *Journal of Portfolio Management*, Vol. 15, No. 1, pp. 29-34.

[24] Gasbarro D., W.K. Wong, J.K. Zumwalt (2007), "Stochastic dominance of iShares", *European Journal of Finance*, 13(1), pp. 89–101.

[25] Hall P.(1992), "The Bootstrap and Edgeworth Expansion", Springer, New York.

[26] Hansen L.P., K.J. Singleton (1983), "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Stock Returns", *Journal of Political Economy*, Vol. 91(2), pp. 249-265.

[27] Hsu J., F. Li (2013), "Low-Volatility Investing", *Journal of Index Investing*, vol. 4, no. 2 (Fall) pp. 67–72.

[28] Huberman G., W. Jiang (2006) "Offering vs. Choice in 401(k) Plans: Equity Exposure and Number of Funds", *Journal of Finance*, 61, pp. 763–801.

[29] Hullirmann W. (2002), "An Alternative Approach to Portfolio Selection", in Proceedings of the 12th International AFIR Colloquium, Cancun, Mexico, Value and Risk Management, Winther thur Life and Pensions.

[30] James W., C. Stein (1961), "Estimation with quadratic loss", *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, J.Neyman ed., Vol. 1, pp. 361–379.

[31] Jannagathan R., T. Ma (2002), "Risk Reduction In Large Portfolios: Why Imposing The Wrong Constraints Helps", *Journal of Finance*, vol. 58(4), pp. 1651-1684.

[32] Jobson J.D., B.M. Korkie (1980), "Estimation for Markowitz Efficient Portfolios", *Journal of the American Statistical Association*, Vol. 75, No. 371, pp. 544-554.

[33] Jobson J.D., B.M. Korkie (1981), "Performance hypothesis testing with the Sharpe and Treynor measures", *Journal of Finance*, 36, pp. 889–908.

[34] Jorion P. (1985), "International Portfolio Diversification with Estimation Risk", *The Journal of Business*, Vol. 58, No. 3, pp. 259-278.

[35] Jorion P. (1986), "Bayes-Stein Estimation for Portfolio Analysis", *The Journal of Financial and Quantitative Analysis*, Vol. 21, No. 3, pp. 279-292.

[36] Karacabey A.A. (2005), "Is mean variance more efficient than MAD in Istanbul?", *Working Paper*, Ankara University.

[37] Kirby C., B. Ostdiek (2012), "It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification", *Journal of Financial and Quantitative Analysis*, Vol.47, pp. 437-467.

[38] Klein R., V. S. Bawa (1976), "The effect of estimation risk on optimal portfolio choice", *Journal of Financial Economics*, vol. 3, issue 3, pp. 215-231.

[39] Kolusheva D. (2008), "Out-of-sample Performances of Asset Allocation Strategies", *Working Paper*.

[40] Konno H., H. Yamazaki (1991), "Mean-Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market", *Management Science*, Vol.37, No.5, 519-531.

[41] Kourtis A. (2015), "A Stability Approach to Mean-Variance Optimization", *The Financial Review*, 50, pp. 301-330.

[42] Kuttler K. (2015), "Linear Algebra, Theory And Applications", *Online e-book*, Brigham Young University.

[43] Lahiri S.N. (2003), "Resampling Methods for Dependent Data", Springer, New York.

[44] Ledoit O., M. Wolf (2003), "Improved Estimation of the Covariance Matrix of Stock Returns", *Journal of Empirical Finance*, Vol. 10, Issue 5, pp. 110-119.

[45] Ledoit O., M. Wolf (2004), “Honey, I shrunk the Sample Covariance Matrix”, *Journal of Portfolio Management*, Vol. 30, No.4, pp. 603-621.

[46] Ledoit O., M. Wolf (2008), “Robust performance hypothesis testing with the Sharpe ratio”, *Journal of Empirical Finance*, 15, pp. 850–859.

[47] Lesmond D.A., J.P. Ogden, C.A. Trzcinka (1999), “A New Estimate of Transaction Costs”, *The Review of Financial Studies*, 12, pp. 1113–1141.

[48] Levy H., H.M. Markowitz (1979), *The American Economic Review*, Vol. 69, No. 3, pp. 308-317.

[49] Markowitz H.M. (1952)"Portfolio Selection"*The Journal of Finance*, Vol. 7, No. 1, pp. 77-91.

[50] Markowitz H.M. (1959), “Portfolio Selection: Efficient Diversification of Investments”, pp. 14-19.

[51] Markowitz H.M. (2006), “De Finetti Scoops Markowitz”, *Journal of Investment Management*, Vol. 4, No. 3.

[52] Merton R.C. (1972), “An Analytic Derivation of the Efficient Portfolio Frontier”, *The Journal of Finance and Quantitative Analysis*, Vol. 7, No. 4, pp. 1851-1872.

[53] Memmel C. (2003), “Performance Hypothesis Testing with the Sharpe Ratio”, *Finance Letters*, 1, pp. 21–23.

[54] Merton R.C. (1980), “On estimating the expected return on the market: An exploratory investigation”, *Journal of Financial Economics*, Vol. 8, Issue 4, pp. 323–361.

[55] Michaud R.O. (1989), “The Markowitz optimization enigma: is ‘optimized’ optimal?”, *Financial Analysts Journal*.

[56] Owen J., R. Rabinovitch (1983), "On the class of elliptical distributions and their applications to the theory of portfolio choice", Journal of Finance, 38, pp. 745-752.

[57] Rockafellar R.T., S. Uryasev (2000), “Optimization of Conditional Value- at-Risk”, *Journal of Risk*, 2(3), pp. 21–41.

[58] Roncalli T. (2010), “Understanding the Impact of Weights Constraints in Portfolio Theory”, *Working Paper*, Available at www.ssrn.com/abstract=1761625.

[59] Rubinstein M. (2002), “Markowitz’s Portfolio Selection: a 50-year retrospective”, *The Journal of Finance*, Vol. 57, N. 3.

[60] Rubinstein M. (2006), “Bruno de Finetti and Mean-Variance Portfolio Selection Problem”, *Journal of Investment Management*, Vol. 4, No. 3.

[61] Shakespeare W. (1596-1598), “The Merchant of Venice”, Act I, Scene 1, p. 2.

[62] Sharpe W.F. (1966), "Mutual Fund Performance", *The Journal of Business*, 39, No. 1, pp. 119–138.

[63] Stein C. (1956), “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution”, *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability, University of California, Berkeley*, J. Neyman ed., Volume I, pp. 197–206.

[64] Stevenson R.L. (1883), "Treasure Island", Part II, Chapter 11, p. 102.

[65] Stoll H., R. Whaley (1983), “Transaction Costs and the Small Firm Effect”, *Journal of Financial Economics*, 12, pp. 57–79.

[66] Szegö G. (2005), Measures of risk, *European Journal of Operational Research*, 163, pp. 5–19.

[67] Towle M.M. (2009) “Ideas and Innovation across Multiple Disciplines: A Discussion with Nobel Laureate Harry M. Markowitz, PhD”, *The Journal of Investment consulting*, Vol. 10, N. 1.

[68] van Vliet P., D. Blitz, B. van der Grient, “Is the Relation Between Volatility and Expected Stock Returns Positive, Flat or Negative?”, *Working paper*.

[69] Woodside-Oriakhi M., C. Lucas, J.E. Beasley (2013), “Portfolio rebalancing with an investment horizon and transaction costs”, *Omega*, 41, pp. 406–420.

[70] Zelner A., V.K. Chetty (1965), “Prediction and Decision Problems in Regression Models from a Bayesian Point of View”, *Journal of the American Statistical Association*,vol. 60, issue 310, pp. 608-661.