Three Essays on Robust Optimization of Efficient Portfolios

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)
am Fachbereich Wirtschaftswissenschaften
der Universität Konstanz

vorgelegt von:
Hao Liu
Wallgutst. 47
78462 Konstanz

Tag der mündlichen Prüfung: 29. Juli 2013

1. Referent: Prof. Dr. Winfried Pohlmeier
2. Referent: Prof. Dr. Jens Jackwerth

To my family

Acknowledgement

The completion of this thesis would not have been possible without the help and support of many people, to whom I would like to express my sincere gratitude with these first words.

First of all, I would like to thank my supervisor, Prof. Dr. Winfried Pohlmeier for his supervision, advice and help at any time in every aspect of my study. I would also like to thank Prof. Dr. Jens C. Jackwerth, Prof. Dr. Ralf Brüggemann, Prof. Dr. Günter Franke, and other members of the faculty who have generously given their time and expertise to better my work.

I would also like to thank my friends and colleagues: Fabian Krüger, Peter Schanbacher, Lidan Großmaß, Ruben Seiberlich, Derya Usyal, Laura Wichert, Dr. Roxana Halbleib, Frieder Mokinski, Zhen Guo, Jing Zeng, Minhui Han, Carlos Fernandez Noya, Fangyi Jin and many others who have helped me in different situations and made life in Konstanz so pleasant.

I want to express my deepest gratitude to my parents for all their love and support over the years, and also to my wife, Zhihua for her continuous encouragement. Finally, special thanks must go to our daughter, Siqi, who has enriched our life beyond measure.

Contents

List of Tables

List of Figures

Summary

The mean-variance approach was first proposed by Markowitz (1952), and laid the foundation of the modern portfolio theory. Despite its theoretical appeal, the practical implementation of optimized portfolios is strongly restricted by the fact that the two inputs, the means and the covariance matrix of asset returns, are unknown and have to be estimated by available historical information. Due to the estimation risk inherited from inputs, desired properties of estimated optimal portfolios are dramatically degraded. This problem has been addressed by empirical research and is well known by both practitioners and academics for many years. However, only quite recently, some studies such as Kan & Zhou (2007) and Frahm & Memmel (2010) tried to provide analytical insights into the real-world portfolio choice problems which help us to understand key aspects of the empirical portfolios and to find the possible way to improve the portfolio performance. This dissertation is a collection of three stand-alone papers and contributes to the recent literature by taking some important issues into account such as the investor's risk preference, non-negativity constraints as well as the presence of structural breaks.

The first chapter analyzes the estimation risk of efficient portfolio selection. We use the concept of certainty equivalent as the basis for a well-defined statistical loss function and a monetary measure to assess estimation risk. For given risk preferences we provide analytical results for different sources of estimation risk such as sample size, dimension of the portfolio choice problem and correlation structure of the return process. Our results show that theoretically sub-optimal portfolio choice strategies turn out to be superior once estimation risk is taken into account. Since estimation risk crucially depends on risk preferences, the choice of the estimator for a given portfolio strategy becomes endogenous depending on sample size, number of assets and properties of the return process. We show that a shrinkage approach accounting for estimation risk is generally superior to simple theoretically suboptimal strategies. Moreover, focusing on just one source of estimation risk, e.g. risk reduction in covariance estimation, can lead to suboptimal portfolios.

Imposing portfolio constraints is one of the most effective ways to improve plug-in estimates of mean-variance portfolios. Jagannathan & Ma (2003) show that the non-negativity constraint in construction of the global minimum variance portfolio has a shrinkage interpretation and could improve the portfolio performance even if the constraint is wrong in population. The second chapter generalizes the theoretical result of Jagannathan & Ma (2003) to the efficient portfolio case where the investor's risk preference plays a crucial role in portfolio construction. We show that imposing the non-negativity constraint on efficient portfolios is equivalent to using a modified covariance matrix which depends on asset expected returns and the risk preferences of investors. We conduct a simulation study with realistic inputs to demonstrate the trade-off between the theoretical and empirical losses of the constrained portfolio with respect to the investor's risk preferences. In addition, different constrained and unconstrained portfolio strategies are compared in both simulation and empirical studies. We find that conservative but unconstrained portfolio strategies proposed by recent studies could outperform constrained portfolios even in the small sample case where the mean and the covariance matrix are estimated with large estimation errors.

The third chapter of the thesis takes possible structural breaks into account and analyzes the estimation risk of different mean-variance portfolio strategies with and without the adding-up constraint. Building upon as idea from Pesaran & Timmermann (2007), we provide an analytical comparison of empirical portfolios estimated by including pre-break data with pure post-break portfolio strategies. It is shown that portfolios incorporating pre-break information can be dominating with respect to their certainty equivalents and the dominance relationship is consistent for different risk aversion levels. Although the theoretical result is obtained under the assumption that there is only a unique structural break whose date is known, our approach combining portfolios estimated from pre- and post-break data can be easily generalized to the multiple break case with unknown break points. In addition, under the normality assumption, we provide an unbiased way to estimate the difference of certainty equivalents between combined portfolios and pure post-break portfolios which allows us to identify the benefit of using pre-break information in portfolio construction.

Bibliography

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

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

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Zusammenfassung

Der Mean-Variance-Ansatz wurde als erstes von Markowitz (1952) vorgeschlagen und legte das Fundament für die Moderne Portfoliotheorie. Trotz seinen theoretischen Vorteile ist die praktische Implementierung optimierter Portfolios durch die Tatsache beschränkt, dass die beiden Inputs, der Erwartungswert und die Kovarianzmatrix der Renditen, unbekannt sind und aus den verfügbaren historischen Daten geschätzt werden müssen. Durch die von den Inputs herrührenden Schätzrisiken werden die gewünschten Eigenschaften der geschätzten optimalen Portfolios drastisch herabgesetzt. Dieses Problem ist sowohl praktischen wie akademischen Kreisen seit vielen Jahren wohl bekannt. Allerdings haben erst vor kurzem einige Studien, wie z.B. Kan & Zhou (2007) und Frahm & Memmel (2010), versucht analytische Einblicke in die realistische Portfolio-Auswahlprobleme zu schaffen, die es ermöglichen Schlüsselaspekte der empirischen Portfolios zu verstehen und einen möglichen Weg zu finden um die Portfolio-Performance zu verbessern. Die vorliegende Dissertation ist eine Zusammenstellung von drei eigenständige Aufsätzen und trägt zur Literatur bei, indem sie wichtige Probleme berücksichtigt, wie z.B. die Risikopräferenzen des Investors, die Nichtnegativitätsbeschränkung und auch das Vorhandensein von Strukturbrüchen.

Das erste Kapitel analysiert das Schätzrisiko einer effizienten Portfolioauswahl. Wir benutzen das Konzept des Sicherheitsäquivalents als Grundlage für eine wohldefinierte statistische Verlustfunktion und ein monetäres Maß zur Beurteilung des Schätzrisikos. Für eine gegebene Risikopräferenz stellen wir analytische Ergebnisse für verschiedene Quellen des Schätzrisikos, wie z.B. Stichprobengröße, die Dimension des Portfolio-Auswahlproblems und Korrelationsstruktur des Renditeprozesses. Unsere Ergebnisse zeigen auf, dass theoretisch suboptimale Portfolioauswahlstrategien sich als besser erweisen, wenn man die Risikopräferenzen berücksichtigt. Da das Schätzrisiko wesentlich von den Risikopräferenzen abhängt, wird die Entscheidung des Beurteilenden für ein gegebenes Portfolio endogen abhängig von der Stichprobengröße, der Anzahl der Assets und den Eigenschaften des Renditeprozesses. Wir zeigen auf, dass ein Shrinkage-Ansatz, der das Schätzrisiko berücksichtigt, einfache Strategien normalerweise übertrifft. Ferner kann das Fokussieren auf nur eine Quelle des Schätzrisikos, z.B. Riskioreduktion in Kovarianzschätzung, zu suboptimalen Portfolios führen.

Einführung der Portfoliobeschränkungen ist eine der effektivsten Methoden, "Plug-in"-Schätzungen eines Mean-Variance Portfolios zu verbessern. Jagannathan & Ma (2003) zeigen, dass Nichtnegativitätsbeschränkungen in der Konstruktion des globalen Minimum-Varianz-Portfolios eine Shrinkage Interpretation implizieren könnten, und die Performance selbst dann verbessern wenn die Beschränkungen bezüglich der Population falsch sind. Das zweite Kapitel verallgemeinert die theoretischen Ergebnisse von Jagannathan & Ma (2003) zum effizienten Portfolio, wo die Risikopräferenz eine entscheidende Rolle bei der Bildung des Portfolios einnimmt. Wir zeigen dass die Einführung von Nichtnegativität Beschränkungen auf effiziente Portfolios äquivalent ist zur Nutzung einer modifizierten Kovarianz-Matrix, welche von erwarteten Renditen der Assets und der Risikopräferenz des Investors abhängt. Wir führen eine Simulationsstudie mit realistischen Einsätzen durch um den "Zielkonflikt" zwischen dem theoretischen und empirischen Verlusten bei eingeschränkten Portfolios bezüglich der Risikopräferenz des Investors zu zeigen. Zusätzlich werden sowohl eingeschränkte als auch nicht-eingeschränkte Portfoliostrategien sowohl in Simulationen als auch in empirischen Studien verglichen. Wir sehen dass konservative aber nicht-eingeschränkte Portfoliostrategien, die in den jüngsten Studien vorgeschlagen werden, die eingeschränkte Portfolios selbst im Fall von kleinen Stichproben übertreffen können, wo der Mittelwert und die Kovarianzmatrix mit großen Schätzfehlern angenommen werden.

Das dritte Kapitel der Dissertation berücksichtigt mögliche Strukturbrüchen und analysiert das Schätzrisiko verschiedener Mean-Variance-Portfolios mit und ohne dem Hinzufügen von Beschränkungen. Eine Idee von Pesaran & Timmermann (2007) folgend, stellen wir einen analytischen Vergleich empirischer Portfolios, die mit "Vor-Bruch" Daten geschätzt werden mit reine-"Nach-Bruch" Strategien auf. Es zeigt sich, dass Portfolios die "Vor-Bruch"-Informationen berücksichtigen bezüglich ihres Sicherheitsäquivalents dominierend sein können; dieses Ergebnis gilt für verschiedene Risikoaversionslevels. Obwohl das theoretische Ergebnis unter der Annahme erzielt wurde, dass es nur einen einzelnen Strukturbruch gibt, dessen Datum unbekannt ist, so kann unser Ansatz, der durch Daten vor- und nach dem Strukturbruch erstellt wurde, sehr einfach generalisiert werden zu einem Fall mit multiplen Brüchen mit unbekannten Bruch-Zeitpunkten. Darüber hinaus stellen wir unter der Normalitätsannahme eine unverzerrte Methode um den Unterschied von Sicherheitsäquivalenten zwischen kombinerten Portfolios und reinen Post-break-Portfolios abzuschätzen, was uns in die Lage versetzt den Vorteil der pre-break-Informationen im Portfolioaufbau zu erkennen.

Literaturverzeichnis

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

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

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Chapter 1

Risk Preferences and Estimation Risk in Portfolio Choice

1.1 Introduction

Empirical estimates of mean-variance efficient portfolio weights often turn out to be unrealistic and reveal considerable standard errors. The problem is well-known in empirical finance and has been documented in several previous studies (e.g. Black & Litterman (1992), Best & Grauer (1991) and Britten-Jones (1999)). The low precision of estimated portfolio weights coincides with many findings from horse races between different portfolio selection strategies showing that theoretically sub-optimal portfolio choices do better in empirical applications than theoretically more efficient strategies. For instance, DeMiguel, Garlappi & Uppal (2009) show that the equally weighted portfolio outperforms the efficient portfolio and several other portfolio strategies in terms of out of sample prediction performance for several performance measures.

This paper sheds more light on classical portfolio selection rules in the tradition of Markowitz (1952) when parameters of the underlying return distribution have to be estimated and estimation risk is taken into account. Our paper is the first to study analytically the performance of empirical efficient portfolio in relation to its benchmark, i.e. the theoretical efficient portfolio. In particular, we focus on the loss in performance due to estimation depending on (i) sample size, (ii) dimension of the portfolio choice problem, (iii) correlation structure of returns and (iv) the investor's risk preferences. By concentrating on the efficient portfolio, our results generalize previous findings for portfolio strategies also satisfying the budget constraint and incorporate those as special cases.

Our analysis of the role of risk preferences for estimation risk provides novel insights into the functioning of shrinkage strategies in portfolio analysis. In particular, we can show that shrinkage of parameter estimates or shrinkage of the estimated portfolio weights have a common representation in terms of the risk preference parameter. In terms of the certainty equivalent (CE) loss, reduction of estimation risk via shrinkage turns out to be equivalent to a redefinition of the theoretical portfolio choice problem for an investor with a higher level of risk aversion. Taking risk preferences as given, the choice of the portfolio strategy and the estimation approach become endogenous, so that theoretically sub-optimal strategies can outperform other portfolio strategies once estimation risk is taken into account.

Despite its relevance for practice, so far only few attempts have been made to theoretically understand the mechanism determining the poor empirical performance of portfolio strategies in order to derive appropriate strategies reducing estimation risks. Notable exceptions are Jagannathan & Ma (2003), who analyze the potentially beneficial impact of imposing false restrictions in portfolio optimization. Kan & Zhou (2007) and Frahm & Memmel (2010) explicitly study the estimation risk of various portfolio estimation strategies. However, their analytical results are either restricted to the tangency portfolio without an adding-up restriction or focus on the global minimum variance portfolio where the investor is assumed to be extremely risk averse.

In the following, we use the loss in certainty equivalent compared to the theoretical efficient portfolio as a well-defined statistical loss function. In the literature, several evaluation criteria have been proposed to evaluate the performance of estimated portfolios. DeMiguel, Garlappi & Uppal (2009) use the Sharpe ratio, CE and the turnover rate to compare the performance of the MV portfolio and the equally weighted portfolio. However, among these evaluation rules, only the CE loss is a proper scoring rule which identifies the true optimal portfolio in the sense that an estimated portfolio can never dominate its theoretical counterpart. Therefore, a comparison of theoretical or empirical portfolio strategies with the theoretical CE of the efficient portfolio provides a clear ranking. The CE has the theoretical appeal of being a statistical loss function which assesses the additional loss an investor faces if he relies on estimated rather than on the true parameter values of the return process. However, contrary to conventional statistical loss functions, the CE based loss expresses estimation loss in terms of monetary units.

We analyze the loss in CE due to estimation analytically for the case of i.i.d.

multivariate-normally distributed asset returns and provide quantitative evidence for the extent of estimation risk of different portfolio strategies. In the presence of estimation risk, the global minimum variance portfolio, although theoretically inferior, can be shown to be the superior portfolio strategy even for an investor with a low level of risk aversion when estimation risk is high, e.g. in the presence of a high dimensional portfolio choice problem, for small or moderate sample sizes and/or in the presence of strong correlation dependencies in the theoretical return process.
Unlike previous studies which ignore the role of risk preferences for the magnitude of the financial loss caused by estimation uncertainty, we show that risk preferences, besides determining the usual trade-off between risk and return, are decisive in determining the extent to which estimation risk with respect to mean and variance contributes to the overall estimation risk. Our findings have a rather intuitive explanation: A risk neutral investor only cares about expected returns and not about risk. Therefore, his estimation risk with respect to the variance-covariance matrix of the return vector does not matter at all. On the contrary, a highly risk averse investor cares a lot about how precisely the variance-covariance matrix of the return vector can be estimated. His monetary loss due to estimation risk depends strongly on the quality of the estimation of the variance-covariance matrix. Therefore, the question regarding the superiority of various portfolio choice strategies taking financial and estimation risk into account can only be answered for given risk preferences.

The outline of the paper is as follows. In Section 1.2, we introduce the CE loss as a statistical loss function and monetary measure for suboptimal portfolio selection. We relate the CE loss and the expected CE loss for the case of parameter estimation of the efficient portfolio to their counterparts for the global minimum variance portfolio and the tangency portfolio. In Section 1.3, we give specific analytical results for the estimation risk based on the assumption of an iid normal return vector. Section 1.4, we propose an optimal shrinkage method tailored to the efficient portfolio with a budget constraint. In Section 1.5 we present some calibration results for a few selected data to provide evidence for the empirical relevance of analytical findings. Section 1.6 concludes and gives an outlook on future research.

1.2 Loss of certainty equivalent

1.2.1 Theoretical MV Efficient Portfolios

Suppose there are N risky assets and the investor can only allocate wealth to these assets. Let r_t denote the $N \times 1$ vector of returns of risky assets with mean $\mathrm{E}[r_t] = \mu$ and covariance matrix $\mathrm{V}[r_t] = \Sigma$. According to the standard mean-variance framework, the efficient frontier can be equivalently presented by the solution of the following optimization problem for the certainty equivalent with respect to the vector of portfolio weights $w = (w_1, w_2, \ldots, w_N)'$:

$$\max_{w,\,\iota' w=1} CE(w) = \max_{w,\,\iota' w=1} \{\mu' w - \frac{\gamma}{2} w'\Sigma w\}, \tag{1.2.1}$$

where the parameter $\gamma \in (0, \infty]$ reflects the investor's level of risk aversion and ι is a $N \times 1$ vector of ones. The objective function of the optimization problem (1.2.1) is the certainty equivalent (CE) of the investor. The closed form solution of (1.2.1) is given by w^*_{ep}, the weight vector of the efficient portfolio:

$$w^*_{ep} = w_{ep}(\mu, \Sigma) = w_{gmv} + \frac{1}{\gamma} \cdot A \cdot \mu, \tag{1.2.2}$$

where

$$A = \Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota} \tag{1.2.3}$$

is a semi-positive definite matrix and the weight vector $w_{gmv} = \Sigma^{-1}\iota/(\iota'\Sigma^{-1}\iota)$ refers to the global minimum variance portfolio (GMVP) as the solution of

$$\min_{w} w'\Sigma w \quad s.t \quad \iota' w = 1.$$

As we will compare in the following w^*_{ep} with the plug-in estimate of the efficient weight vector, $w_{ep}(\hat{\mu}, \hat{\Sigma})$, we use the superscript $*$ to indicate that $w_{ep}(\cdot, \cdot)$ is evaluated at the true parameters of the return process. Because $A \cdot \iota = 0$, the weight $w_z = \frac{1}{\gamma} \cdot A \cdot \mu$ is the weight vector of a zero-investment portfolio with weights summing up to zero, i.e. $\iota' w_z = 0$. Obviously, for the limiting case of an extremely risk averse investor ($\gamma \to \infty$), the weights of the efficient portfolio approach the weights of the GMVP, which solely depend on the variance of the return vector and are ,thus, only exposed to estimation risk of Σ. Therefore, as risk aversion increases, exposure to estimation risk with respect to mean returns decreases. Since the efficient portfolio weight is the sum of the GMVP weight and the zero-investment portfolio weight, $w^*_{ep} = w_{gmv} + w_z$, and since w_{gmv} represents the optimal choice if the investor is not willing to trade any risks against returns, w_z contains all relevant information concerning the extent to which the investor is willing to trade risk against return given his preferences and the nature of the return process.

In a similar fashion, it is also helpful to formulate the efficient portfolio weights (1.2.2) in terms of a linear combination of the weight vector of the GMVP and the weight vector of the tangency portfolio:

$$\begin{aligned} w^*_{ep} &= \frac{1}{\gamma}\Sigma^{-1}\mu + (1 - \frac{1}{\gamma}\iota'\Sigma^{-1}\mu)\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} \\ &= w_{tp} + (1 - \iota' w_{tp}) \cdot w_{gmv}, \end{aligned}$$

where $w_{tp} = \frac{1}{\gamma}\Sigma^{-1}\mu$, is the weight of the tangency portfolio, which is the solution of the optimization problem (1.2.1) when the adding-up restriction $\iota' w = 1$ is ignored. Therefore, if the mean and the covariance matrix are both known, the optimal investment strategy for an investor with risk aversion level γ is to allocate $\iota' w_{tp}$ of wealth to the optimal tangency portfolio and $1 - \iota' w_{tp}$ to the GMVP. As shown in the Appendix, the CE of the GMVP takes on the form

$$CE(w_{gmv}) = \mu_{gmv} - \frac{\gamma}{2}\sigma^2_{gmv} = \frac{\iota'\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\iota} - \frac{\gamma}{2} \cdot \frac{1}{\iota'\Sigma^{-1}\iota}, \tag{1.2.4}$$

where $\mu_{gmv} = \mu' w_{gmv} = \frac{\iota'\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\iota}$ is the mean return and $\sigma^2_{gmv} = w'_{gmv}\Sigma w_{gmv} = 1/\iota'\Sigma^{-1}\iota$ the variance of the GMVP. Substituting the weight of the efficient portfolio w^*_{ep} into the objective function, we obtain the theoretically highest CE an investment decision can achieve.

Proposition 1.2.1 (Decomposition of the Efficient CE)**.**
The CE of the efficient portfolio based on the weight vector w^*_{ep} given in (1.2.2) is:

$$CE(w^*_{ep}) = \mu' w^*_{ep} - \frac{\gamma}{2} w^{*\prime}_{ep} \cdot \Sigma \cdot w^*_{ep} = \frac{1}{2\gamma}\Delta_{SSR} + CE(w_{gmv}), \tag{1.2.5}$$

where Δ_{SSR} is the difference between the squared Sharpe ratios of the tangency portfolio and the GMVP,

$$\begin{aligned} \Delta_{SSR} = \mu' \cdot A \cdot \mu &= \frac{(\mu'\Sigma^{-1}\mu)(\iota'\Sigma^{-1}\iota) - (\iota'\Sigma^{-1}\mu)^2}{\iota'\Sigma^{-1}\iota} \\ &= \left(\frac{\mu' w_{tp}}{\sqrt{w'_{tp}\Sigma w_{tp}}}\right)^2 - \left(\frac{\mu' w_{gmv}}{\sqrt{w'_{gmv}\Sigma w_{gmv}}}\right)^2 \\ &= (\mu - \mu_{gmv} \cdot \iota)'\Sigma^{-1}(\mu - \mu_{gmv} \cdot \iota) > 0, \end{aligned} \tag{1.2.6}$$

$CE(w_{gmv})$ is the CE of the GMVP defined in Equation (1.2.4).

Proof 1.2.1. *See Appendix.*

The decomposition of the efficient CE given in Proposition 1.2.1 helps to understand the performance of the efficient theoretical portfolio relative to the theoretical

GMVP. For a portfolio choice based on the GMVP, the investor is assumed to be extremely risk averse and, thus, only cares about the risk of the investment. Therefore, by construction for the case of known parameters the GMVP always yields a lower CE than the optimal efficient portfolio based on w^*_{ep}. The extent of the theoretical dominance of the efficient portfolio over the GMVP in terms of the CE depends on Δ_{SSR}, which captures the additional return that the investor receives compared to the maximum risk averse investor.

The following two examples illustrate how population mean and covariance matrix affect the level of CE.

Example 1.2.1. *Consider the case where $\Sigma = \sigma^2 I$ where I denotes the identity matrix. We have:*

$$CE(w^*_{ep}) = \frac{1}{N}\frac{\sum_{i<j}(\mu_i - \mu_j)^2}{2\gamma\sigma^2} + \frac{1}{N}\sum_{i=1}^{N}\mu_i - \frac{\gamma}{2N}\sigma^2.$$

Hence, the theoretical CE is high if i) the average over the mean returns of the assets is high, ii) the variance is low, iii) the differences between means are high, and iv) the risk aversion parameter is low. It can be also seen that the additional CE of efficient portfolio compared to the GMVP, which is proportional to Δ_{SSR}, depends on the dissimilarity of mean returns (standardized by the return variances).

Consider now the impact of correlations on the CE. To make the result more intuitive, we consider the following bivariate case with only one correlation parameter.

Example 1.2.2. *Let $N = 2$ and $\sigma_1 = \sigma_2 = \sigma$, then:*

$$CE(w^*_{ep}) = \frac{(\mu_1 - \mu_2)^2}{4\gamma\sigma^2} \cdot \frac{1}{1-\rho} + \frac{\mu_1 + \mu_2}{2} - \frac{\gamma}{4}(1+\rho)\sigma^2$$

Therefore, in addition to the impact of the mean and variances, the CE is also high if returns are highly (positively or negatively) correlated. As the correlation approaches -1, the variance of the GMVP converges to zero, and, thus, the CE of the GMVP is equal to its expected return, which is the cross sectional average of the means in this concrete example. It is important to note that, as the correlation level approaches $+1$, Δ_{SSR} approaches infinity, in the case of $\mu_1 \neq \mu_2$, and, hence, the CE of the theoretically efficient portfolio is unbounded.

Similar to Proposition 1.2.1, the CE of the efficient portfolio can also be expressed in relation to the CE of the tangency portfolio.

$$
\begin{aligned}
CE(w_{ep}^{*}) =& \frac{1}{2\gamma}\mu'\Sigma^{-1}\mu - \frac{1}{2\gamma}\frac{1}{\iota'\Sigma^{-1}\iota}\left(\iota'\Sigma^{-1}\mu - \gamma\right)^2 \\
=& CE(w_{tp}) - \frac{1}{2\gamma}\frac{1}{\iota'\Sigma^{-1}\iota}\left(\iota'\Sigma^{-1}\mu - \gamma\right)^2.
\end{aligned}
$$

By definition, the theoretical CE of the efficient portfolio is never larger than the CE of the tangency portfolio. Finally, consider the weights of the Maximum Sharpe Ratio portfolio (MaxSR):

$$
w_{SR} = \frac{w_{tp}}{\iota' w_{tp}} = \frac{\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\mu}.
$$

For a given γ the CE of this portfolio is:

$$
CE(w_{SR}) = \frac{\mu'\Sigma^{-1}\mu}{(\iota'\Sigma^{-1}\mu)^2}(\iota'\Sigma^{-1}\mu - \frac{\gamma}{2}).
$$

The CE of the MaxSR portfolio and the CE of the GMVP are both positive if and only if $\iota'\Sigma^{-1}\mu > \frac{\gamma}{2}$. In this case the CE of the MaxSR portfolio is always larger than that of the GMVP.

Theoretically the GMVP yields a lower CE than the efficient portfolio, the tangency portfolio, and the MaxSR portfolio do. However, if the estimation risk is taken into account, the estimated GMVP can be much more reliable than the three other portfolio strategies.

1.2.2 Expected CE Loss

Consider the CE of the efficient portfolio $CE(w_{ep}^{*})$. By definition, the efficient portfolio dominates any other portfolio satisfying the adding up restriction in terms of the CE for given risk preferences, i.e. $CE(w_{ep}^{*}) \geq CE(w)$, where w is a weight vector satisfying the budget constraint and obtained by some other arbitrary portfolio selection rule based on true or estimated parameters. In particular, the efficient portfolio always dominates any estimated efficient portfolio $CE(w_{ep}^{*}) \geq CE(\hat{w}_{ep})$. Thus,

$$
\mathcal{L}(\hat{w}, w_{ep}^{*}) \equiv CE(w_{ep}^{*}) - CE(\hat{w}) \geq 0 \qquad (1.2.7)
$$

is a well defined statistical loss function with $\hat{w} = w(\hat{\mu}, \hat{\Sigma})$. In practice, when the mean and the covariance matrix are unknown and the optimized portfolio is based on estimated inputs, it is, by definition, inferior to its theoretical counterpart. Since $CE(\hat{w})$ is a random variable, the estimation loss is also random. The expectation over the loss function (1.2.7) defines the risk function

$$\mathcal{R}(\hat{w}|w_{ep}^*) \equiv \mathrm{E}\left[\mathcal{L}(\hat{w}, w_{ep}^*)\right] = CE(w_{ep}^*) - \mathrm{E}\left[CE(\hat{w})\right] > 0, \qquad (1.2.8)$$

where $\mathcal{R}(\hat{w}|w_{ep}^*)$ gives the expected loss in CE if an estimate of the portfolio weight is taken instead of the true efficient portfolio weight. The risk function can be interpreted as the average amount of money an investor is willing to pay that makes him indifferent to a portfolio based on estimated parameters versus a portfolio evaluated on the true parameters. Therefore, if two empirical portfolio strategies based on the estimated portfolio weights $\hat{w}$ and $\tilde{w}$ are compared, the comparison should be based on their expected CE measures. Hence, the portfolio based on $\hat{w}$ strictly dominates the portfolio based on $\tilde{w}$ if

$$\mathrm{E}\left[CE(\hat{w})\right] - \mathrm{E}\left[CE(\tilde{w})\right] = \mathcal{R}(\tilde{w}|w_{ep}^*) - \mathcal{R}(\hat{w}|w_{ep}^*) > 0. \qquad (1.2.9)$$

Thus proving the dominance of an estimated portfolio over any other estimated portfolio in terms of the expected CE is equivalent to the comparison of their corresponding risk functions.

Cho (2010) uses the CE to define the economic loss and shows that the loss of a suboptimal portfolio can be approximated by:

$$\mathcal{R}(\hat{w}|w) = CE(w) - \mathrm{E}\left[CE(\hat{w})\right] \cong \frac{\gamma}{2} tr\left(\mathrm{Cov}[\hat{w}] \cdot \Sigma\right).$$

He argues that this approximation can be applied to all MV portfolio problems of any given constraint, although it only holds when the estimated portfolio weights are assumed to be unbiased. The plug-in estimators of the portfolio weights are, however, generally nonlinear functions of the estimated mean and the estimated covariance matrix. Even if these estimates are unbiased, the weights as nonlinear functions are generally biased. Kan & Zhou (2007) provide a formal proof for the tangency portfolio under iid normality of the return vector and derive the size of the finite sample bias depending on the sample size T and the dimension of the portfolio choice problem N. In a similar fashion, Okhrin & Schmid (2006) show that the plug-in estimated efficient portfolio weights are also biased but have a smaller mean squared error than their unbiased counterparts. In any case, the unbiasedness assumption turns out to be very restrictive and we will show below that the bias in the estimated weights can be large and can even dominate the variance-covariance matrix of the vector of weights. In order to identify the exact CE loss of a suboptimal portfolio relative to the true efficient portfolio, we provide the following proposition:

Proposition 1.2.2 (CE Loss and Expected CE Loss)**.** *Let w_{ep}^* denote the solution of the MV-maximization problem (1.2.1) and let $\hat{w}$ be any portfolio weight vector satisfying $\iota'\hat{w} = 1$, then:*

$$\mathcal{L}(\hat{w}, w_{ep}^*) = CE(w_{ep}^*) - CE(\hat{w}) = \frac{\gamma}{2}(w_{ep}^* - \hat{w})'\Sigma(w_{ep}^* - \hat{w}).$$

The expected loss of CE is thus given by:

$$\mathcal{R}(\hat{w}|w_{ep}^*) = CE(w_{ep}^*) - \mathrm{E}\left[CE(\hat{w})\right] = \frac{\gamma}{2} tr\left(\Sigma \cdot \left[\mathrm{Cov}[\hat{w}] + Bias(\hat{w})^2\right]\right),$$

where $Bias(\hat{w})^{2=} \left(\mathrm{E}\left[\hat{w}\right] - w_{ep}^*\right)\left(\mathrm{E}\left[\hat{w}\right] - w_{ep}^*\right)'$.

Proof 1.2.2. *See Appendix.*

For the GMVP, the CE loss due to estimation error is given by:[1]

$$\mathcal{L}(\hat{w}, w_{gmv}) = CE(w_{gmv}) - CE(\hat{w}) = (w_{gmv} - \hat{w})'\Sigma(w_{gmv} - \hat{w}),$$

while a similar result can be obtained for the tangency portfolio[2]:

$$\mathcal{L}(\hat{w}, w_{tp}) = CE(w_{tp}) - CE(\hat{w}) = \frac{\gamma}{2}(w_{tp} - \hat{w})'\Sigma(w_{tp} - \hat{w}).$$

Although Proposition 1.2.2 reveals some similarities to the corresponding loss functions for the GMVP and the tangency portfolio, considerable differences occur if the mean and the covariance matrix are replaced by their estimates. Proposition 1.2.3 relates the CE loss directly to the estimated mean.

Proposition 1.2.3 (Expected CE for known Variance)**.** *If the true covariance matrix is known, then the CE loss of the efficient portfolio is:*

$$\mathcal{L}(w_{ep}(\hat{\mu}, \Sigma), w_{ep}^*) = \frac{1}{2\gamma}\left(\mu - \hat{\mu}\right)' \cdot A \cdot \left(\mu - \hat{\mu}\right)$$

with the expected CE loss given by:

$$\mathcal{R}(w_{ep}(\hat{\mu}, \Sigma)|w_{ep}^*) = \frac{1}{2\gamma} tr\left(A \cdot \left[\mathrm{Cov}[\hat{\mu}] + Bias(\hat{\mu})^2\right]\right),$$

where $\hat{\mu}$ is an arbitrary estimator for the mean returns.

[1] See Kempf & Memmel (2006) for a proof.
[2] See Frahm (2010)

Proof 1.2.3. *See Appendix.*

Since A is a positive semi-definite matrix, the CE loss is non-negative. Frahm (2010) derives the economic loss resulting from estimation for a given known covariance matrix for the case of the tangency portfolio:

$$\mathcal{L}(w_{tp}(\hat{\mu}, \Sigma), w_{tp}) = \frac{1}{2\gamma}(\mu - \hat{\mu}) \cdot \Sigma^{-1} \cdot (\mu - \hat{\mu}) \, .$$

Note that the estimation loss for the efficient portfolio is smaller than the estimation loss for the tangency portfolio because the former accounts for the budget constraint $\iota' w = 1$. In addition, the expected loss of the estimated efficient portfolio only depends on the differences between $(\mu_i - \hat{\mu}_i)$ and $(\mu_j - \hat{\mu}_j)$ for $i \neq j$. Proportional overestimation or underestimation of the mean implies no economic loss in the case of efficient portfolio but can significantly reduce the value of the estimated tangency portfolio. Thus, the estimated efficient portfolio can outperform the estimated tangency portfolio in practice.

1.2.2.1 Within and Out-of-sample Measures

Aside from the theoretical CE defined in (1.2.1), the out-of-sample CE concept is often considered:

$$CE_{os}(w) = \mathrm{E}\left[w'r\right] - \frac{\gamma}{2}\mathrm{V}\left[w'r\right].$$

This concept is often used in practice in comparative empirical studies when μ and Σ are unknown. Obviously, if all input elements are known, the out-of-sample CE, $CE_{os}(w)$, is identical to the CE definition defined in (1.2.1). However, if estimated portfolio weights are used instead of the true ones, both weights and returns are random and, thus, the two CE concepts differ.

Assumption 1.2.1.
Asset returns r_t are stochastically independent of trading strategy $\hat{w}$ selected by the investor.

Given Assumption 1.2.1, the out-of-sample CE can be computed as follows:

$$\begin{aligned} CE_{os} &= \mathrm{E}\left[\hat{w}'r_t\right] - \frac{\gamma}{2}\mathrm{V}\left[\hat{w}'r_t\right] \\ &= \mathrm{E}\left[\hat{w}'\,\mathrm{E}\left[r_t|\,\hat{w}\right]\right] - \frac{\gamma}{2}\left(\mathrm{E}\left[\hat{w}'\,\mathrm{V}\left[r_t|\,\hat{w}\right]\hat{w}\right] + \mathrm{V}\left[\mathrm{E}\left[r_t|\,\hat{w}\right]' \cdot \hat{w}\right]\right) \\ &= \mathrm{E}\left[CE(\hat{w})\right] - \frac{\gamma}{2}\mu'\,\mathrm{Cov}[\hat{w}]\mu \end{aligned} \qquad (1.2.10)$$

Therefore, if portfolio weights are estimated, CE_{os} is smaller than the theoretical CE which is based on the true mean and covariance matrix. In the following analysis, we use $\mathcal{R}_{os}(.|.)$ to denote the out-of-sample risk function (the out-of-sample expected loss) of estimated portfolio compared to the true optimal efficient portfolio. Based on Proposition 1.2.2, the out-of-sample risk function can be easily obtained by:

$$\mathcal{R}_{os}(\hat{w}|w_{ep}^*) = \frac{\gamma}{2} tr\left((\Sigma + \mu\mu') \cdot \text{Cov}[\hat{w}] + \Sigma \cdot Bias(\hat{w})^2\right). \tag{1.2.11}$$

1.2.3 Implied Mean of a Portfolio

The exact risk function given in Subsection 1.2.2 was derived under the assumption that the true covariance matrix is known and that the estimation risk is solely due to estimation of mean returns. Using Proposition 1.2.4 given below, we can represent any theoretical or empirical portfolio weight in terms of an efficient portfolio weight with a known covariance matrix and an implied mean vector. Therefore, a comparison of any portfolio weight in terms of the CE loss can be reduced to a comparison of the equivalent representation of this portfolio weight with the efficient portfolio weight evaluated at the true mean and covariance. The CE differences are simply reflected by the differences between the true and the implied mean vector.

Proposition 1.2.4 (Equivalent Representation). *Let S be the subspace of $\mathbb{R}^N$ which is orthogonal to ι and let the $N \times (N-1)$ matrix V be the basis matrix of S, i.e. the column vectors of V construct a basis of S. Let $\hat{w}$ denote the weight vector of a given portfolio. A is the matrix defined in Equation (1.2.3). Then, there is an implied mean vector $\hat{\mu}_{im} = c\iota + \hat{\mu}_{im}^0$ such that:*

$$\begin{aligned}\hat{w} &= w_{ep}(\hat{\mu}_{im}, \Sigma)\\ &= \frac{1}{\gamma}\Sigma^{-1}\hat{\mu}_{im} + (1 - \frac{1}{\gamma}\iota'\Sigma^{-1}\hat{\mu}_{im})\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota},\end{aligned}$$

where c is any arbitrary constant and

$$\hat{\mu}_{im}^0 = \gamma \cdot V(V'BV)^{-1}V' \cdot \Sigma \cdot \hat{w} \qquad \textit{and} \qquad \iota'\hat{\mu}_{im}^0 = 0.$$

with $B = \Sigma \cdot A$.

Proof 1.2.4. *See Appendix.*

The second term of the implied mean, $\hat{\mu}_{im}^0$, sums up to zero. For example, consider the case of the true efficient portfolio. Here, $\hat{\mu}_{im}^0$ measures the deviation from the average of the means for the single returns, i.e. $\mu - \bar{\mu}\iota$ with $\bar{\mu} = \iota'\mu/N$. Moreover, note there exists an infinite number of implied means $\hat{\mu}_{im}$, which generate the same portfolio and, hence, the same CE. Therefore, analyzing the $\hat{\mu}_{im}$ is fully equivalent to analyzing any other implied mean vector.

With the help of the equivalent representation given in Proposition 1.2.4 all errors in the suboptimal portfolio are contained in $\hat{\mu}_{im}$, and the difference between the true optimal portfolio and the estimated portfolio can be identified if the difference between the true population mean and the estimated (implied) mean is known. Thus the "best" implied mean can be defined as $\hat{\mu}^*_{im} = \hat{\mu}^0_{im} + \bar{\mu}\iota$, which has the lowest distance (measured by the Euclidean metric) to the true mean.

For given estimates of the mean and the covariance matrix, the implied mean $\hat{\mu}^0_{im}$ can be explicitly written as:

$$\hat{\mu}^0_{im} = V(V'BV)^{-1}V' \cdot \Sigma \cdot \hat{\Sigma}^{-1}(\gamma\iota + \hat{B}\hat{\mu}).$$

Even if the mean μ is estimated without error, the implied mean will differ from the true population mean if the covariance matrix is estimated with errors. Moreover, the difference will be large if the difference between the estimated inverse covariance matrix $\hat{\Sigma}^{-1}$ and the true Σ^{-1} is large. It is clear that, if the risk aversion parameter γ is large, the investor cares more about the variance of portfolio, and, therefore, the estimation risk in the covariance has a larger impact on the implied mean. In addition, the impact of the errors in the mean vector and the errors in the covariance matrix on the implied mean is not additive and the interaction might be large. This issue will be discussed below in more detail for cases where sample counterparts of the mean and the covariance matrix are used. With the help of Proposition 1.2.4, we can easily reformulate the CE loss and the expected CE loss of the efficient portfolio by replacing the estimated mean by the implied mean.

Example 1.2.3. *The theoretical GMVP has an implied mean equal to ι. Thus, the expected CE loss of the GMVP is given by*

$$\mathcal{R}(w_{gmv}|w^*_{ep}) = \mathcal{R}(w_{ep}(\iota, \Sigma)|w^*_{ep}) = \frac{1}{2\gamma}\mu \cdot A \cdot \mu = \frac{1}{2\gamma}\Delta_{SSR},$$

which is consistent with the result in Section 1.2.1.

Based on the equivalent representation by the implied mean Proposition 1.2.5 gives an upper bound for the CE loss.

Proposition 1.2.5 (Upper Bound of the CE loss)**.** *The loss of CE is bounded by:*

$$\begin{aligned}\mathcal{L}(\hat{w}, w_{ep}^*) \leq & \frac{1}{2\gamma}(\sum_{i=1}^{N} \lambda_i^{-1}) \cdot (\sum_{i=1}^{N} \sigma_i^{-2}) \parallel (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}) \parallel_2^2 \cdot \parallel \mu - \hat{\mu}_{im} \parallel_2^2 \\ = & \frac{N-1}{2\gamma}(\sum_{i=1}^{N} \lambda_i^{-1}) \cdot (\sum_{i=1}^{N} \sigma_i^{-2}) \left(\sum_{i=1}^{N} (\mu_i - \hat{\mu}_{im,i})^2 \right),\end{aligned}$$

where λ_i, $i = 1, \ldots, N$ is the eigenvalue of the correlation matrix.

Proof 1.2.5. *See Appendix.*

The upper bound of CE loss depends on the risk aversion level γ, the number of assets, the level of variances, the potential collinearity between assets and the estimation error (measured by $\parallel \mu - \hat{\mu}_{im} \parallel_2^2$). The upper bound gives the highest possible loss when a suboptimal portfolio strategy is used. It provides information regarding the outcome range of possible losses compared to the efficient portfolio strategy based on the true population parameters of the return process. As will be shown below, the risk function for cases of an unknown mean estimated by a sample mean and a given variance-covariance matrix turns out to be independent of the covariance matrix. This is because the squared estimation error of the first moment, $\mathrm{E}\left[(\hat{\mu} - \mu)'(\hat{\mu} - \mu)\right]$, is proportional to the true second moment Σ. Thus, the squared error in sample mean is to some extent compensated by Σ^{-1} in the loss function.

1.3 Expected CE Loss under Normality

Since the exact functional form of the expected CE loss depends on the underlying distributional properties of the return process, we follow Okhrin & Schmid (2006) and assume i.i.d. multivariate normality for the return process.

Assumption 1.3.1.

a) $r_t \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \Sigma)$ *for* $t = 1, \ldots, T$.

b) $T \geq N + 4$ *and* $N \geq 3$.

Population mean and population covariance matrix are estimated by their sample counterparts:

$$\hat{\mu} = \bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t \qquad \text{and} \qquad \hat{\Sigma} = S = \frac{1}{T-1}\sum_{t=1}^{T} (r_t - \bar{r})(r_t - \bar{r})'.$$

Under the normality assumption from above, the two estimators are distributed as:

$$\bar{r} \sim \mathrm{N}(\mu, \frac{1}{\mathrm{T}}\Sigma) \qquad \text{and} \qquad \mathrm{S} \sim \mathrm{W_N}(\mathrm{T}-1, \Sigma)/\mathrm{T}-1,$$

where $W_N(T-1, \Sigma)$ denotes the Wishart distribution with $T-1$ degrees of freedom and covariance matrix Σ.

1.3.1 Expected CE Loss of the Efficient Portfolio

Case I: Sample Mean - True Covariance Matrix

Consider first the case where the covariance matrix is known but the mean vector is estimated with errors. The expected CE loss using the sample mean of the return vector and the true population covariance matrix is:

$$\mathcal{R}(w_{ep}(\bar{r}, \Sigma)|w^*_{ep}) = \frac{1}{2\gamma} tr\left(A \cdot \mathrm{Cov}[\bar{r}]\right) = \frac{1}{2\gamma}\frac{N-1}{T}. \qquad (1.3.1)$$

Obviously, for this case estimation risk is negligible for the extremely risk averse investor as, for her, only the estimation risk concerning the covariance matrix matters. Moreover, a large sample size and a small number of assets in the portfolio reduces estimation risk as well. It should not be too surprising that the estimation risk of the tangency portfolio $\mathcal{R}(w_{tp}(\bar{r}, \Sigma)|w_{tp}(\mu, \Sigma)) = \frac{1}{2\gamma} \cdot \frac{N}{T}$ is larger than the one for the efficient portfolio given in (1.3.1), since the budget constraint is taken into account for the latter, which reduces estimation uncertainty.[3]

Consider now the out-of-sample CE loss caused by using the sample mean. Using (1.2.10) the difference between the out-of-sample CE and the within-sample CE is given by

$$\mathrm{E}\left[CE\left(w_{ep}(\bar{r}, \Sigma)\right) - CE_{os}\left(w_{ep}(\bar{r}, \Sigma)\right)\right] = \frac{1}{T} \cdot \frac{1}{2\gamma}\Delta_{SSR}.$$

Thus, the out-of-sample loss of CE due to estimation error in sample means is given by

$$\mathcal{R}_{os}(w_{ep}(\bar{r}, \Sigma)|w^*_{ep}) = \frac{1}{2\gamma} \cdot \left(\frac{N-1+\Delta_{SSR}}{T}\right). \qquad (1.3.2)$$

[3]See Kan & Zhou (2007) for a proof of the risk function of the tangency portfolio given Σ is known.

Case II: True Mean - Sample Covariance Matrix

Conditional on any given estimate of mean returns, the covariance of estimated portfolio weights using sample covariance matrix S is (see Okhrin & Schmid (2006)):

$$\mathrm{Cov}[w_{ep}(\hat{\mu}, S)|\hat{\mu}] = \frac{1}{T-N-1}\frac{A}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma^2}\left(c_1 A \cdot \hat{\mu}\hat{\mu}' \cdot A + c_2 \hat{\mu}' \cdot A \cdot \hat{\mu} \cdot A\right),$$

where

$$c_1 = \frac{(T-1)^2\,(T-N+1)}{(T-N)\,(T-N-1)^2\,(T-N-3)} \quad \text{and} \quad c_2 = \frac{(T-1)^2}{(T-N)\,(T-N-1)\,(T-N-3)}.$$

In addition, the conditional expectation of estimated weights using sample covariance matrix is:

$$\mathrm{E}\left[w_{ep}(\hat{\mu}, S)|\,\hat{\mu}\right] = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{T-1}{T-N-1}\cdot\frac{1}{\gamma}A\cdot\hat{\mu}.$$

Therefore, if the true mean μ is known, the expected CE loss due to estimation error in sample covariance matrix can be calculated as:

$$\begin{aligned} & \mathcal{R}(w_{ep}(\mu, S)|w_{ep}^{*}) \\ = \; & \frac{\gamma}{2} tr\left(\Sigma \cdot \left[\mathrm{Cov}[w_{ep}(\mu, S)] + Bias\left(w_{ep}(\mu, S)\right)^2\right]\right) \\ = \; & \frac{\gamma}{2}\frac{N-1}{T-N-1}\sigma^2_{gmv} + \frac{\Delta_{SSR}}{2\gamma}\left(c_1 + c_2\,(N-1) + \left(\frac{N}{T-N-1}\right)^2\right). \qquad (1.3.3) \end{aligned}$$

The first term in (1.3.3), $\frac{\gamma}{2}\frac{N-1}{T-N-1}\sigma^2_{gmv}$, can be interpreted as the baseline risk component, as it occurs in both, the risk function for any estimated efficient portfolio and the risk function for the estimated GMVP derived below. Contrary to Case I, where μ has to be estimated, risk aversion has an ambivalent effect on the expected CE loss when Σ has to be estimated. A higher degree of risk aversion increases the impact of baseline risk of the return process represented by σ^2_{gmv}. However, higher risk aversion reduces the effect of the overall earnings potential of the return process. It is easy to show that the expected CE loss, as a function of the degree of risk aversion, has a unique minimum, i.e. investors with different degrees of risk aversion may face the same expected CE loss. The less risk averse investor faces less estimation risk compared to a more risky investor. However, he loses money in terms of CE by pursuing a less profitable strategy in terms of the theoretical CE.

Similar to (1.3.3), we can also compute the out-of-sample CE loss when the sample covariance matrix is used. Using (1.2.11), if the true mean is known, the out-of-sample loss of CE due to estimation error of sample covariances is:

$$
\begin{aligned}
&\mathcal{R}_{os}(w_{ep}(\mu, S)|w_{ep}^*) \qquad (1.3.4)\\
=&\frac{\gamma}{2}tr\left((\Sigma + \mu\mu') \cdot \text{Cov}[w_{ep}(\mu, S)] + \Sigma \cdot Bias(w_{ep}(\mu, S))^2\right)\\
=&\frac{\gamma}{2}\frac{N-1+\Delta_{SSR}}{T-N-1}\sigma_{gmv}^2 + \frac{\Delta_{SSR}}{2\gamma}\left(c_1\left(1+\Delta_{SSR}\right) + c_2\left(N-1+\Delta_{SSR}\right) + \left(\frac{N}{T-N-1}\right)^2\right).
\end{aligned}
$$

Case III: Sample Mean and Sample Covariance Matrix

The covariance of estimated weights using sample mean and sample covariances is given by:[4]

$$
\begin{aligned}
\text{Cov}[w_{ep}(\bar{r}, S)] \quad = \quad & \frac{1}{T-N-1}\frac{A}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma^2}\left(c_1 A\mu\mu' A + c_2 \mu' A\mu A\right)\\
&+\frac{1}{T}\cdot\frac{A}{\gamma^2}\left(c_1 + c_2\left(N-1\right) + \frac{\left(T-1\right)^2}{\left(T-N-1\right)^2}\right).
\end{aligned}
$$

Using this result, we are able to derive the conditional loss of CE due to the estimation error when the sample mean and the sample covariance matrix are used:

$$
\begin{aligned}
&\mathcal{R}\left(w_{ep}(\bar{r}, S)|w_{ep}^*\right)\\
=&\frac{\gamma}{2}tr\left(\Sigma \cdot \left[\text{Cov}[w_{ep}(\bar{r}, S)] + Bias(w_{ep}(\bar{r}, S))^2\right]\right)\\
=&\frac{\gamma}{2}\frac{N-1}{T-N-1}\frac{1}{\iota'\Sigma^{-1}\iota} + \frac{\mu' \cdot A \cdot \mu}{2\gamma}\left(c_1 + c_2\left(N-1\right)\right)\\
&+\frac{N-1}{T}\cdot\frac{1}{2\gamma}\left(c_1 + c_2\left(N-1\right) + \frac{\left(T-1\right)^2}{\left(T-N-1\right)^2}\right) + \frac{1}{2\gamma}\cdot\left(\frac{N}{T-N-1}\right)^2 \cdot \mu' \cdot A \cdot \mu\\
=&\mathcal{R}(w_{ep}(\mu, S)|w_{ep}^*) + c_3\mathcal{R}(w_{ep}(\bar{r}, \Sigma)|w_{ep}^*). \qquad (1.3.5)
\end{aligned}
$$

with

$$
c_3 = \frac{(T-1)^2(T-2)}{(T-N-1)(T-N)(T-N-3)} > 1.
$$

Note that the overall expected loss due to estimation is larger than the sum of the risks of estimating μ and Σ, given that the other parameters are known. The coefficient c_3 can be interpreted as an interaction effect which gives more weight to the estimation risk with respect to Σ if mean returns also have to be estimated. c_3 increases with the dimension of the portfolio choice problem and decreases with sample size. Table 1.1 below gives some values of c_3.

[4] see Okhrin & Schmid (2006) for a proof.

Table 1.1: Scale Effect due to Estimation of μ

$T \setminus N$	5	10	15	20	25	30
60	1.31	1.75	2.43	3.50	5.30	8.60
120	1.14	1.30	1.50	1.74	2.03	2.40
180	1.09	1.19	1.30	1.43	1.57	1.74
240	1.07	1.14	1.22	1.30	1.39	1.50
300	1.05	1.11	1.17	1.23	1.30	1.37

Value of c_3 for different T (number of observations) and N (number of assets).

Sample size matters particularly for large portfolios. The scale factor decreases by more than 84% if the sample size increases, e.g. from 5 years of monthly data (T = 60) to 25 years (T = 300), for a portfolio of 30 assets while the reduction due to an increase in sample size is only 20 % for portfolios of 5 assets.

For the out-of-sample case, we obtain very similar results. The difference between within-sample and out-of-sample CE is given by:

$$\begin{aligned} & \mathrm{E}\left[CE\left(w_{ep}(\bar{r},S)\right) - CE_{os}\left(w_{ep}(\bar{r},S)\right)\right] \\ = \; & \frac{\gamma}{2}\mu' \,\mathrm{Cov}[w_{ep}(\bar{r},S)]\mu \\ = \; & \frac{\gamma}{2}\frac{1}{T-N-1}\frac{\mu' A\mu}{\iota'\Sigma^{-1}\iota} + \frac{(\mu' A\mu)^2}{2\gamma}(c_1+c_2) + \frac{1}{T}\cdot\frac{\mu' A\mu}{2\gamma}\left(c_1 + c_2\,(N-1) + \frac{(T-1)^2}{(T-N-1)^2}\right) \\ = \; & \frac{\gamma}{2}\frac{1}{T-N-1}\sigma^2_{gmv}\Delta_{SSR} + \frac{1}{2\gamma}(c_1+c_2)\cdot\Delta^2_{SSR} + c_3\cdot\frac{1}{T}\cdot\frac{\Delta_{SSR}}{2\gamma} \end{aligned}$$

Thus, the out-of-sample loss of CE due to the estimation error in the sample means and the sample covariances is:

$$\mathcal{R}_{os}\left(w_{ep}(\bar{r},S)|w^*_{ep}\right) = \mathcal{R}_{os}\left(w_{ep}(\mu,S)|w^*_{ep}\right) + c_3\cdot\mathcal{R}_{os}\left(w_{ep}(\bar{r},\Sigma)|w^*_{ep}\right). \qquad (1.3.6)$$

The composition of the out-of-sample CE loss is the same as for the unconditional CE loss given by (1.3.5). However, because the term on the right hand side in (1.3.6) is larger than its counterpart for the expected CE loss of the unconditional case, the out-of-sample expected CE loss is clearly larger than the unconditional expected CE loss.

1.3.2 Expected CE loss of the GMVP

Frahm (2010) suggests using the mean of the GMVP as an estimator for the mean vector of returns:

$$\hat{\mu}_{gmv} = \bar{r}'\hat{w}_{gmv} \cdot \iota = \frac{\bar{r}'\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota} \cdot \iota \tag{1.3.7}$$

Because $A \cdot \iota = 0$, the efficient frontier reduces to the GMVP when all means are equal. The CE loss in this case is equal to the CE loss using the estimated GMVP. Therefore in terms of the loss function, the estimator for the means given in (1.3.7) is equivalent to any estimator which is proportional to ι. The estimated portfolio weights of the GMVP using the sample covariance matrix are essentially unbiased (Okhrin & Schmid (2006)). Therefore, for a given γ, the expected CE difference between the theoretical GMVP and the empirical GMVP is:

$$\mathrm{E}\left[CE(w_{gmv}) - CE\left(w_{gmv}(S)\right)\right] = \frac{\gamma}{2}\left[\mathrm{V}\left[w_{gmv}'r_t\right] - \mathrm{V}\left[w_{gmv}(S)'r_t\right]\right].$$

Thus the within-sample CE loss and the out-of-sample CE loss of the empirical GMVP for a given γ are

$$\mathcal{R}(w_{gmv}(S)|w_{ep}^*) = \mathcal{R}(w_{ep}(c \cdot \iota, S)|w_{ep}^*) = \frac{\gamma}{2} \cdot \frac{N-1}{T-N-1} \cdot \sigma_{gmv}^2 + \frac{1}{2\gamma}\Delta_{SSR} \tag{1.3.8}$$

and

$$\mathcal{R}_{os}\left(w_{gmv}(S)|w_{ep}^*\right) = \mathcal{R}_{os}(w_{ep}(c \cdot \iota, S)|w_{ep}^*) = \frac{\gamma}{2} \cdot \frac{N-1+\Delta_{SSR}}{T-N-1} \cdot \sigma_{gmv}^2 + \frac{1}{2\gamma}\Delta_{SSR}, \tag{1.3.9}$$

respectively. Also, for this special case, the expected CE losses are nonlinear functions of the risk preference parameter with a unique minimum. Therefore, individuals with different risk attitudes may face the same expected CE loss.

1.4 Shrinkage Estimation of the Efficient Portfolio

Based on a Bayesian analysis, Jorion (1986) proposes a shrinkage estimator for the mean of the form:

$$\hat{\mu}_{shrink} = \eta \cdot \hat{\mu} + (1-\eta) \cdot \hat{\mu}_{gmv} \cdot \iota,$$

where η denotes the shrinkage parameter and $\hat{\mu}_{gmv}$ is the shrinkage target which is equal to the estimated expected return of GMVP as defined in (1.2.4). The optimal shrinkage parameter of Jorion's Bayes-Stein is given by:

$$\eta_{BS} = 1 - \frac{N+2}{(N+2) + T(\hat{\mu} - \hat{\mu}_{gmv} \cdot \iota)'\hat{\Sigma}^{-1}(\hat{\mu} - \hat{\mu}_{gmv} \cdot \iota)} = \frac{\hat{\Delta}_{SSR}}{\hat{\Delta}_{SSR} + \frac{N+2}{T}}.$$

As shown above, the computation of the efficient portfolio using the shrinkage target $\hat{\mu}_{gmv}$ is equivalent to shrinking the mean to any target of form $c \cdot \iota$, where c is an arbitrary constant. For $c = 0$, this shrinkage approach is equivalent to biasing the estimated mean towards zero. Thus estimation risk in the GMVP weight has no impact on the final result. Furthermore, shrinking the mean is also equivalent to directly applying shrinkage estimation to the efficient portfolio weight with the GMVP as the shrinkage target, i.e.

$$w_{shrink}(\eta, \hat{\mu}, \hat{\Sigma}) = \eta \cdot w_{ep}(\hat{\mu}, \hat{\Sigma}) + (1-\eta) \cdot w_{gmv}(\hat{\Sigma}) = w_{gmv}(\hat{\Sigma}) + \eta \hat{w}_z. \tag{1.4.1}$$

Equation (1.4.1) also reveals that shrinking mean returns to the mean of the GMVP is nothing but reducing the investors arbitrage opportunities by lowering the contribution of the (estimated) zero-investment portfolio.

Kan & Zhou (2007) argue that the shrinkage portfolio suggested by Jorion (1986) "can be suboptimal, because it is not constructed for holding optimal position", and propose the optimal shrinkage estimator of the mean for a tangency portfolio. In the following, we derive the optimal shrinkage estimator tailored for the efficient portfolio given in (1.2.2). Using $\bar{r}$ and S as estimators for the shrinkage weight (1.4.1) in the risk function given in Proposition 1.2.2 leads to:

$$\begin{aligned} & \mathcal{R}(w_{shrink}(\eta, \bar{r}, S)|w^*_{ep}) = \mathcal{R}(w_{ep}(\eta \cdot \bar{r}, S)|w^*_{ep}) & (1.4.2)\\ = \; & \frac{\gamma}{2} trace\left(\Sigma \cdot \mathrm{Cov}[w_{shrink}(\eta, \bar{r}, S)] + \Sigma \cdot Bias\left(w_{shrink}(\eta, \bar{r}, S)\right)^2\right) & \\ = \; & \frac{\gamma}{2}\frac{N-1}{T-N-1}\sigma^2_{gmv} + \frac{\eta^2}{2\gamma}\left[(c_1 + c_2(N-1))\Delta_{SSR} + c_3\frac{N-1}{T}\right] & \\ & + \frac{\Delta_{SSR}}{2\gamma}\left(1 - \frac{(T-1)\eta}{T-N-1}\right)^2. & (1.4.3) \end{aligned}$$

The optimal shrinkage factor η^* can be obtained by minimizing (1.4.2) and solving the first order condition:

$$\eta^* = \frac{\Delta_{SSR}}{c_3\left(\Delta_{SSR} + \frac{N-1}{T}\right)} \cdot \frac{T-1}{T-N-1} = \frac{(T-N)(T-N-3)}{(T-1)(T-2)} \cdot \frac{\Delta_{SSR}}{\left(\Delta_{SSR} + \frac{N-1}{T}\right)} < 1.$$

With this optimal shrinkage parameter η^*, the expected CE loss is:

$$\mathcal{R}(w_{shrink}(\eta^*, \bar{r}, S)|w_{ep}^*) = \frac{\gamma}{2}\left(\frac{N-1}{T-N-1}\right)\sigma_{gmv}^2 + \frac{1}{2\gamma}\Delta_{SSR}\left(1 - \frac{T-1}{T-N-1}\eta^*\right). \tag{1.4.4}$$

Comparing the expected CE loss of the optimal shrinkage portfolio to the expected loss of the empirical GMVP (see Equation (1.3.8)), we see that using the shrinkage approach leaves the baseline risk component unchanged but reduces the theoretical loss of the GMVP. Therefore, the optimal shrinkage portfolio outperforms the GMVP only marginally when the sample size is small or the theoretical loss of the GMVP is small.

The comparison of the risks of the optimal shrinkage portfolio with the risk of the sample efficient portfolio given by (1.3.5) shows that the expected loss of the optimal shrinkage portfolio in (1.4.4) is similar to the partial loss of the sample efficient portfolio, which is caused by the estimation error of the covariance matrix (see Equation (1.3.3)). However, the optimal shrinkage portfolio does not contain the partial loss, which is due to the estimation error in the sample mean and the interaction term, $c_3 \cdot \frac{N-1}{T}$. Thus, in finite samples, the optimal shrinkage portfolio can yield a much higher performance than the sample efficient portfolio. Interestingly, although the CE loss depends on the risk aversion parameter, the true variance of the GMVP and the difference in the squared Sharpe ratios, Δ_{SSR}, the optimal shrinkage parameter is only a function of Δ_{SSR} for a given sample size and number of assets. Therefore optimal shrinkage estimation is valid for any type of investor.

Obviously, the optimal shrinkage parameter η^* is infeasible because it depends on the unknown Δ_{SSR}. This term has to be estimated, which introduces an additional source of estimation error. Contrary to the case of optimal shrinkage tangency portfolio proposed by Kan & Zhou (2007), where both Δ_{SSR} and the mean of the GMVP have to be estimated, η^* only depends on one unknown parameter. This is because, as argued above, shrinking the portfolio weights to the weights of the GMVP is equivalent to shrinking the mean to zero. Therefore, estimates based on the optimal shrinkage parameter η^* are likely to be more reliable than the estimates of shrinkage tangency portfolio along the lines of Kan & Zhou (2007).

Based on the normality assumption and on $\bar{r}$ and S as plug-in estimates, the distribution for the $\hat{\Delta}_{SSR}$ is given by:[5]

$$\frac{T-N+1}{N-1}\hat{\Delta}_{SSR} \sim F_{(N-1,T-N+1)}(T\Delta_{SSR}),$$

where $F_{(N-1,T-N+1)}(T\Delta_{SSR})$ is non-central F distribution with $N-1$ and $T-N+1$ degrees of freedom and noncentrality parameter $T\Delta_{SSR}$. Therefore, the unbiased estimator of Δ_{SSR} is given by:

$$\hat{\Delta}^{u}_{SSR} = \frac{T-N-1}{T}\hat{\Delta}_{SSR} - \frac{N-1}{T}.$$

When N is large, the unbiased estimator may be negative which would be equivalent to assuming that all assets have negative mean returns. This will cause large short positions in the constructed portfolio. To ensure the positiveness of $\hat{\Delta}_{SSR}$, we can also bound the unbiased estimator by zero:

$$\hat{\Delta}^{mod}_{SSR} = max\left\{\frac{T-N-1}{T}\hat{\Delta}_{SSR} - \frac{N-1}{T}, 0\right\},$$

where $\hat{\Delta}_{SSR}$ is the plug-in estimator of Δ_{SSR} based on the sample mean and the sample covariance matrix.

1.5 Calibration to Real Data

The results presented so far are purely analytical. In the following, we therefore provide quantitative evidence on the expected CE loss for realistic magnitudes on the population parameters of the return process for different levels of risk aversion, sample size and portfolio size. More specifically, we consider three different samples containing 1) 5 industry portfolios; 2) 10 industry portfolios; 3) 30 industry portfolios based on monthly returns for the sample period 07/1926 - 09/2009 published on Kenneth French's Web site.[6] For our simulations, we choose the sample estimates from these three data sets for the true mean and true covariance matrix .

1.5.1 Properties of the Theoretical CE

Table 1.2 below gives the annualized theoretical CE of the efficient portfolio for different levels of γ. The table also reports the values for the difference in the squared Sharpe ratio, Δ_{SSR}, and the variance of the GMVP, σ^2_{gmv}, based on monthly data

[5] See Kan & Zhou (2007) for a proof.
[6] *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

for the three different portfolio sizes. While Δ_{SSR}, which reflects the potential financial gains over the minimum variance portfolio, has a positive effect on the CE, the second measure reflects the baseline risk inherent in the portfolio choice problem.

Table 1.2: Theoretical CE for Different Portfolios and Degrees of Risk Aversion

	Annualized CE (%)								Δ_{SSR}	σ^2_{gmv}
γ	0.04	0.5	1	2	4	6	8	10	(%)	(%)
5-PF	43.51	14.06	12,07	9.97	6.72	3.67	0.68	-2.30	0.2085	0.2452
10-PF	106.01	18.02	13.79	11.05	8.41	6.40	4.56	2.78	0.6348	0.1405
30-PF	427.21	43.44	26.43	17.40	11.85	9.08	7.00	5.20	2.7786	0.1152

Annualized theoretical CE of efficient portfolio, Δ_{SSR} and variance of the GMVP, σ^2_{gmv}, for three different samples. All values are scaled by 10^2.

A comparison of the values for Δ_{SSR} and σ^2_{gmv} reveals that, as the number of asset increases, Δ_{SSR} increases but σ^2_{gmv} decreases only moderately. The two quantities are similar in size in 5 industry portfolio case, but Δ_{SSR} is almost 25 times higher than σ^2_{gmv} in the 30 portfolio case. This explains why the theoretical CE substantially increases when more assets are considered. Since all three portfolio data sets are constructed from the same stocks, the heterogeneity between assets is large in the large dimension cases. However, grouping the assets to larger portfolios lowers the dissimilarity of the asset candidates and consequently restricts the space of portfolio optimization reflected by the values for Δ_{SSR} and σ^2_{gmv}.

Moreover, for all three data sets, the corresponding values of Δ_{SSR} are relatively small with a maximum of 2.78% for the 30 asset case. As shown previously, the difference between the expected CE loss and the expected CE loss for the out-of-sample case mainly depends on the magnitude of Δ_{SSR}. Since Δ_{SSR} is changing only moderately for our data constellations, we refrain from reporting the simulation results for the out of sample case.

The risk aversion level of $\gamma = 0.04$ (first column) is equivalent to a risk tolerance of 50 considered by Chopra & Ziemba (1993). In this case, the investor is nearly risk-neutral and can achieve with the MV portfolio a very high annualized CE. As the γ increases, however, the investor cares more about the risk and the theoretical CE decreases. The CE can even become negative if the investor is too risk averse and, in this case, the risky investment becomes unattractive.

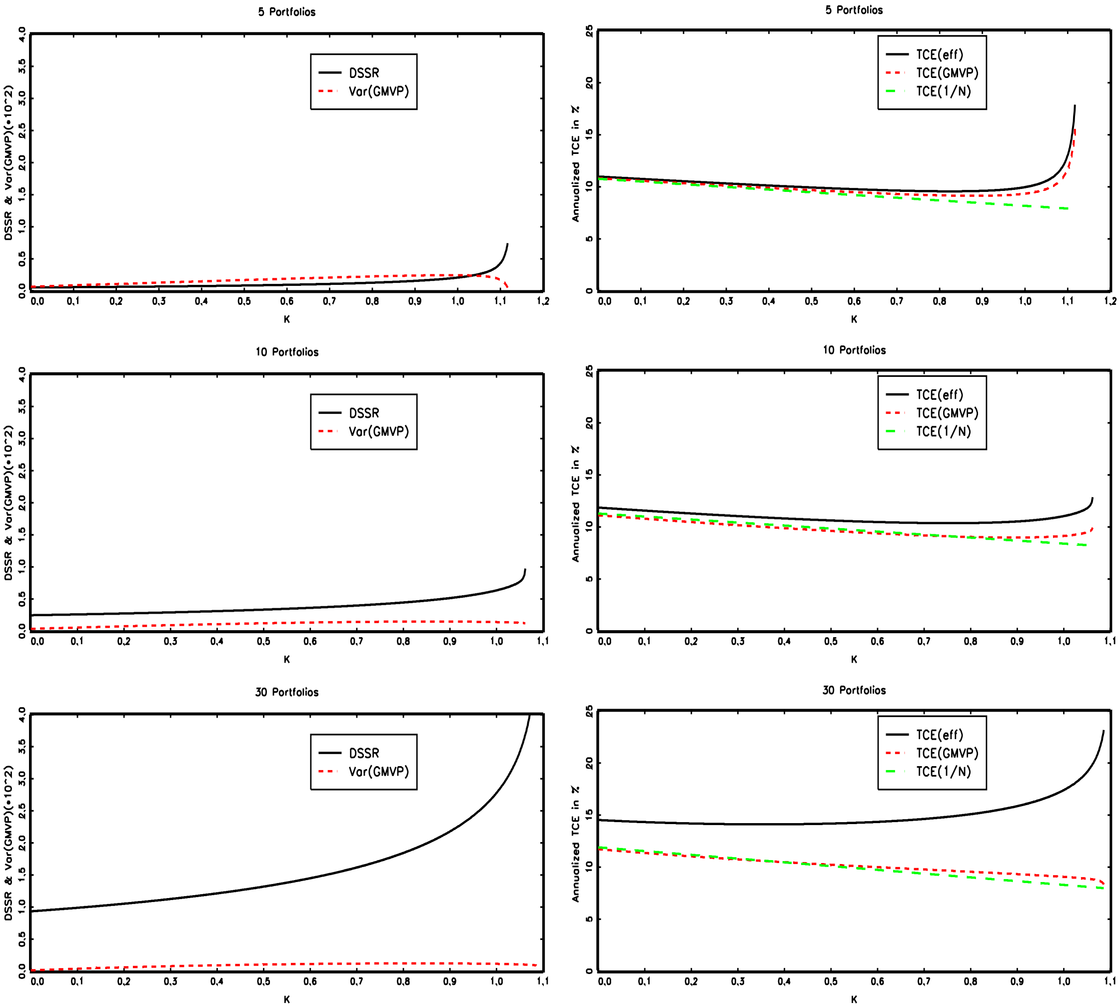


Figure 1.1: Left Panel: Impact of correlation on Δ_{SSR} (solid line) and σ^2_{gmv} (dashed line) for the three portfolios. Right Panel: Impact of correlation on annualized theoretical CE (in %) on 1) efficient portfolio (TCE(eff), solid); 2), GMVP (TCE(GMVP), small dashes); 3) equally weighted portfolio (TCE(1/N), long dashes). $\gamma = 2$

As given by (1.3.3) and (1.3.8), Δ_{SSR} and σ^2_{gmv} have different effects on the risk of the estimated efficient portfolio and the GMVP for different levels of γ. For instance, if we use the estimated GMVP, the impact of estimation error of the covariance matrix cannot be neglected in the case of 5 industry portfolios. It is, however, less relevant in the case of 30 industry portfolios where the CE loss of the estimated GMVP is mainly caused by the theoretical difference between the GMVP and the efficient portfolio. In addition, we can also show that as γ increases, the impact of Δ_{SSR} decreases, whereas the impact of σ^2_{gmv} increases.

Theoretical CE and the Impact of Near-Multicollinearity

Theoretical CE and Δ_{SSR} can be heavily affected by the population properties of the asset pool under consideration. Example 1.2.2 demonstrates the dependency of the CE on the correlation for the two asset case analytically. In the following, we show how the correlation level affects Δ_{SSR}, σ^2_{gmv} and the CE using the correlation matrix from the three portfolios as the benchmark. More specifically, we keep the means and the variances of the portfolios unchanged but replace the correlation matrix by

$$C_k = k \cdot C_0 + (1 - k) \cdot I,$$

where C_0 is the original correlation matrix estimated from the data and k is the parameter determining the strength of the correlation structure. Here, we assume that all correlations change proportionally. It is known that all eigenvectors of C_k are the same as the eigenvector of C_0, but the eigenvalues of C_k are $\lambda_{k,j} = k\lambda_j + (1 - k)$, $j = 1, \dots, N$, where λ_j is the j-th eigenvalue of the original correlation matrix C_0. To ensure the positive definiteness of C_k, all $\lambda_{k,j}$ $j = 1, \dots, N$ must be positive. Thus, we select the value of k from the interval $[0, (1 - \lambda_{0,min})^{-1})$, where λ_{min} is the smallest eigenvalue of C_0.

From Figure 1.1 we see that, with increasing correlation level, Δ_{SSR} increases substantially and, therefore, the theoretical CE of the efficient portfolio increases. Based on Equation (1.3.8) the theoretical loss of the GMVP, $\frac{1}{2\gamma}\Delta_{SSR}$, also increases as the correlation level increases. However, since the σ^2_{gmv} is almost unchanged, we can conclude that the expected CE loss of the GMVP caused by estimation risk, which is proportional to σ^2_{gmv}, does not change too much as the correlation level increases (see Equation (1.3.8) for comparison). Hence, the aggregate loss of the GMVP can be dominated by its theoretical part in the case where asset returns are highly correlated. In addition, in the case of large dimensional portfolio choice problem, Δ_{SSR} is large for high correlations. Thus, the GMVP also performs poorly in this case and the difference between the GMVP and the equally weighted portfolio is small.

Impact of Risk Aversion

Table 1.3 gives the annualized CE of the efficient theoretical portfolio, the theoretical GMVP and the equally weighted portfolio for our three different data sets. By definition, $CE(w^*_{ep})$ dominates the two other strategies and all CEs are monotonically decreasing functions of the risk aversion level. Their relative performance, however, depends on the specific parameter constellations. Note that, even in this theoretical scenario where estimation risk is still ignored, the performance of the two suboptimal

strategies comes close to the efficient portfolio strategy. For instance, in 5 industry portfolio case, the performance of the GMVP is quite close to the performance of the efficient portfolio for γ larger than 2. For the 30 industry portfolio case, however, the theoretical efficient portfolio outperforms the GMVP significantly, even for high levels of risk aversion due to the strong increase of Δ_{SSR}.
In addition, because the GMVP has a relatively small theoretical loss in the 5 portfolio case, it is more attractive than the equally weighted portfolio even for a less risk averse investor. But in the large dimension case, the ranking of the GMVP and the equally weighted portfolio can change when different risk aversion levels are considered. For the less risk averse investor, the equally weighted portfolio could be more attractive.

Table 1.3: Annualized CE (in %) of efficient portfolio, GMVP and equally weighted portfolio.

γ	0.04	0.5	1	2	4	6	8	10
5-PF								
w^*_{ep}	43.51	14.06	12.07	9.97	6.72	3.67	0.68	-2.30
w_{gmv}	12.23	11.56	10.82	9.35	6.41	3.46	0.52	-2.42
$1/N$	11.52	10.73	9.88	8.18	4.77	1.36	-2.05	-5.46
10-PF								
w^*_{ep}	106.01	18.02	13.79	11.05	8.41	6.40	4.56	2.78
w_{gmv}	10.79	10.40	9.98	9.14	7.46	5.77	4.08	2.40
$1/N$	11.68	10.91	10.08	8.40	5.06	1.71	-1.64	-4.99
30-PF								
w^*_{ep}	427.21	43.44	26.43	17.40	11.85	9.08	7.00	5.20
w_{gmv}	10.42	10.10	9.76	9.06	7.68	6.30	4.92	3.54
$1/N$	12.03	11.15	10.20	8.29	4.48	0.67	-3.14	-6.95

Our findings highlight the importance of the level of risk aversion for the evaluation of portfolio strategies. In many horse races of various portfolio strategies presented in the literature, the risk aversion parameter is usually arbitrarily given, e.g. a popular value for the risk aversion parameter is 2 (e.g. Best & Grauer (1991) and DeMiguel, Garlappi & Uppal (2009)), which is assumed to be the market average risk aversion level. However, when this specific risk aversion level is used to compare the GMVP with the equally weighted portfolio, it is likely that no large differences in the performance of the two strategies will be found and one may be inclined to conclude that complete ignorance of any portfolio optimization strategy by using the equally weighted portfolio is quite meaningful.

1.5.2 Properties of the Expected CE Loss

As shown before, the theoretical CE loss of the GMVP decreases as the risk aversion level increases. This relationship, however, no longer holds for the expected CE loss based on the plug-in estimated portfolio weights. To illustrate this, in the following we use the 10 and 30 industry portfolios based on sample size $T = 60$ because, in this case, the estimation risk of the GMVP also contributes a considerable fraction to the overall risk of the estimated GMVP. The results are depicted in Figure 1.2.

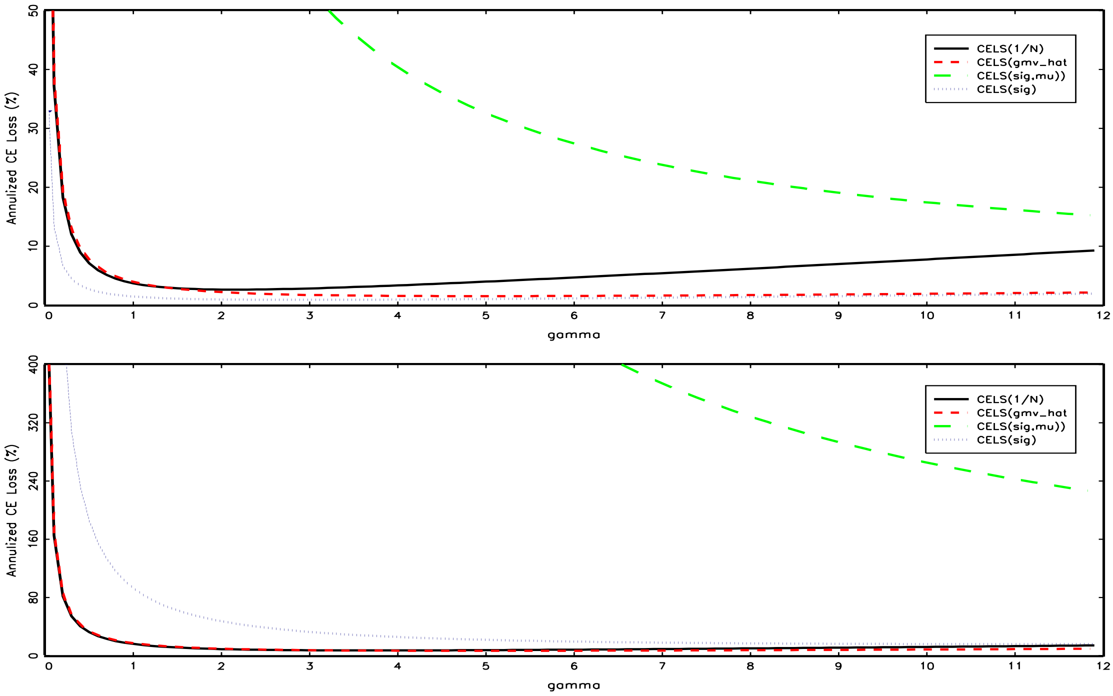


Figure 1.2: Relation between γ and expected CE loss for different portfolio strategies:
1) CELS(1/N) - solid line: Equally weighted portfolio, $\mathcal{R}((1/N)\iota|w^*_{ep})$;
2) CELS(gmv_hat) - dark dashed line: Empirical GMVP, $\mathcal{R}(w_{gmv}(S)|w^*_{ep})$;
3) CELS(sig,mu) - light dashed line: Empirical efficient PF, $\mathcal{R}(w_{ep}(\bar{r}, S)|w^*_{ep})$;
4) CELS(sig) - dotted line: Empirical efficient PF with μ known, $\mathcal{R}(w_{ep}(\mu, S)|w^*_{ep})$.
Upper panel: 10 portfolio case, lower panel: 30 portfolio case. Sample size: $T = 60$.

Figure 1.2 shows the (annualized) expected CE losses of different portfolio strategies in the 10-portfolio (upper panel) and the 30-portfolio (lower panel) cases. In the 5 portfolio case, the result is similar to that in the case of 10 portfolios and is not reported here. Accounting for estimation risk changes the ranking of the portfolio strategies completely. The empirical counterpart of the efficient portfolio (plug-in estimator) is inferior to any other portfolio strategy considered over the entire range of γ. The equally weighted portfolio is only a strong competitor for investors with low levels of risk aversion. If one assumes, that professional investors are less risk averse than private investors our findings imply that in particular professional investors should be cautious in applying the the empirical counterparts of efficient theoretical portfolio strategies. The comparison of $\mathcal{R}(w_{ep}(\bar{r}, S)|w^*_{ep})$ with $\mathcal{R}(w_{ep}(\mu, S)|w^*_{ep})$ (case 3 and 4 in Figure 1.2) reveals that the gains of knowing the true mean return vector are considerable. However, in the large dimension (30 portfolio) case, simply estimating Σ and adopting the empirical GMVP turn out to be better than the efficient portfolio strategy based on the true mean. Even in the 10 portfolio case, the empirical GMVP performs as well as this (infeasible) efficient portfolio strategy for $\gamma \geq 2$. Note that, with the exception of the equally weighted portfolio, the expected CE loss is declining in γ, if the value of gamma is within a reasonable range. Because the theoretical CE is always declining in γ, we can conclude that the decrease in the expected CE for the empirical portfolio weights is less pronounced, i.e. the more risk averse investor loses less due to estimation risk than the less risk averse investor.

Relevance of Estimation Risk: Mean vs. Covariance

As shown for the cases 3 and 4 in Figure 1.2 knowledge of subsets of the parameters of the return distribution can lead to a major reduction in estimation risk. In the following, we provide additional numerical evidence on the relative importance of estimation risk with respect to the mean and the covariance matrix. For this, we compute the percentage of $\mathcal{R}(w_{ep}(\bar{r}, \Sigma)|w^*_{ep})$ and $\mathcal{R}(w_{ep}(\mu, S)|w^*_{ep})$ in the total expected CE loss of the empirical efficient portfolio.

Table 1.4 (upper panel) provides results on the expected CE loss for the efficient portfolio if both μ and Σ are estimated by their sample counterparts. in most of the cases, using plug-in portfolio weights leads to substantial expected losses. Only the risk averse investor focusing on small portfolios and using large (in practice rather unrealistic) sample sizes can expect minor losses due to estimation. The expected estimation loss literally explodes for the less risk averse investor with large portfolios (N = 30). In this case, even large sample size does not really mitigate the problem.

Table 1.4: Relative expected CE Loss due to Estimation Error in Mean and Covariance (in %)

Annualized Expected Loss of Sample Efficient Portfolios

	5 PF			10 PF			30 PF		
$T \backslash \gamma$	1	2	8	1	2	8	1	2	8
60	52.55	26.44	7.43	159.27	79.87	21.13	2585.39	1293.73	328.62
120	22.88	11.52	3.27	59.13	29.67	7.94	359.98	180.33	46.77
180	14.59	7.35	2.09	35.97	18.05	4.85	173.50	86.95	22.75

Impact of Interaction Effect

	5 PF			10 PF			30 PF		
$T \backslash \gamma$	1	2	8	1	2	8	1	2	8
$s_{(\bar{r})}$					(%)				
60	76.11	75.64	67.32	56.51	56.34	53.25	11.22	11.21	11.03
120	87.43	86.84	76.55	76.10	75.83	70.85	40.28	40.20	38.75
180	91.37	90.74	79.73	83.39	83.08	77.32	55.72	55.59	53.12
$s_{(S)}$					(%)				
60	0.50	1.11	12.00	0.92	1.21	6.64	3.59	3.67	5.19
120	0.50	1.17	12.89	0.88	1.23	7.72	3.41	3.60	7.08
180	0.50	1.19	13.18	0.87	1.24	8.09	3.28	3.50	7.78
$s_{(\bar{r},S)}$					(%)				
60	23.39	23.24	20.69	42.57	42.45	40.11	85.19	85.12	83.78
120	12.07	11.99	10.57	23.02	22.94	21.43	56.31	56.20	54.17
180	8.13	8.07	7.09	15.74	15.68	14.59	41.00	40.91	39.09

i.) $s_{(\bar{r})} = \dfrac{\mathcal{R}(w_{ep}(\bar{r},\Sigma)|w^*_{ep})}{\mathcal{R}(w_{ep}(\bar{r},S)|w^*_{ep})}$ = share of the expected CE loss due to estimation of μ;

ii.) $s_{(S)} = \dfrac{\mathcal{R}(w_{ep}(\mu,S)|w^*_{ep})}{\mathcal{R}(w_{ep}(\bar{r},S)|w^*_{ep})}$ = share of expected CE loss due to estimation of Σ;

iii.) $s_{(\bar{r},S)} = \dfrac{(c_3-1)\mathcal{R}(w_{ep}(\bar{r},\Sigma)|w^*_{ep})}{\mathcal{R}(w_{ep}(\bar{r},S)|w^*_{ep})}$ = interaction effect.

The lower panel of Table 1.4 reports on the partial risks and the overall risk of the sample efficient portfolio based on the three data sets. With the help of Equation (1.3.5), we can decompose the overall estimation risk into its three components: i.) the expected CE loss resulting from estimating the mean, ii.) the expected CE loss resulting from estimating the variance and iii) the interaction effect for the latter if, in addition, the mean has to be estimated. Interestingly, the expected CE loss due to estimation of Σ is rather small in percentage terms for all three portfolio sizes and sample sizes. However, its contribution to the overall risk becomes relevant if the mean has to be estimated as well. For T = 60 and N = 30, the scale effect explains almost 84% of the expected CE loss. Our calibration exercise makes clear that horse races for the "best" covariance estimator, which take mean returns as given, assume away a large fraction of estimation uncertainty.

1.5.3 Shrinkage Portfolio

Although under the normality assumption the true parameters of the return process are estimated efficiently, the evidence provided in this study reveals that the empirical weights of the efficient portfolio imply a high estimation risk, while theoretically less efficient strategies perform comparatively well. In the following, we compare the expected CE loss of different portfolio strategies discussed in the previous sections. Since the modified estimator for Δ_{SSR} introduces an additional source of estimation risk for the shrinkage portfolio and the analytical expected CE loss of this estimated shrinkage portfolio is difficult to obtain, we compute the expected CE loss by means of Monte-Carlo methods using 10000 replications. The expected losses for the different portfolio strategies are depicted in Figure 1.3.

The performance of the five approaches varies considerably depending on the sample size. Their relative performance is, however, qualitatively the same for the two risk aversion levels $\gamma = 2$ and $\gamma = 8$. Figure 1.3 confirms the strength of the equally weighted portfolio performance over optimized portfolio strategies for small and moderate sample sizes ($T \leq 240$) and $\gamma = 2$. Only for very large, i.e. in practice unrealistic, sample sizes, the equally weighted portfolio is clearly beaten by the optimized empirical portfolios. In any case, using the empirical GMVP generally dominates the equally weighted portfolio.

If the optimal shrinkage parameter is known, the shrinkage method optimally combines the sample efficient portfolio and the empirical GMVP and significantly reduces the expected CE loss caused by estimation error. This benefit is particularly large if the theoretical risk of the GMVP is large, e.g. in the 30 portfolio case. However, in practice, the estimation error in the shrinkage parameter reduces the benefits of using the shrinkage portfolio. As was shown in the previous section, the shrinkage approach cannot reduce the baseline risk component. Therefore, as the dimension of the portfolio choice problem increases, the baseline risk component dominates, so that shrinkage portfolio also performs poorly and can be even worse than the equally weighted portfolio. Furthermore, if the risk aversion level is high, the optimal shrinkage portfolio, the estimated shrinkage portfolio and the empirical GMVP perform similarly because the true GMVP is theoretically close to the true efficient portfolio and has relative low theoretical CE loss.

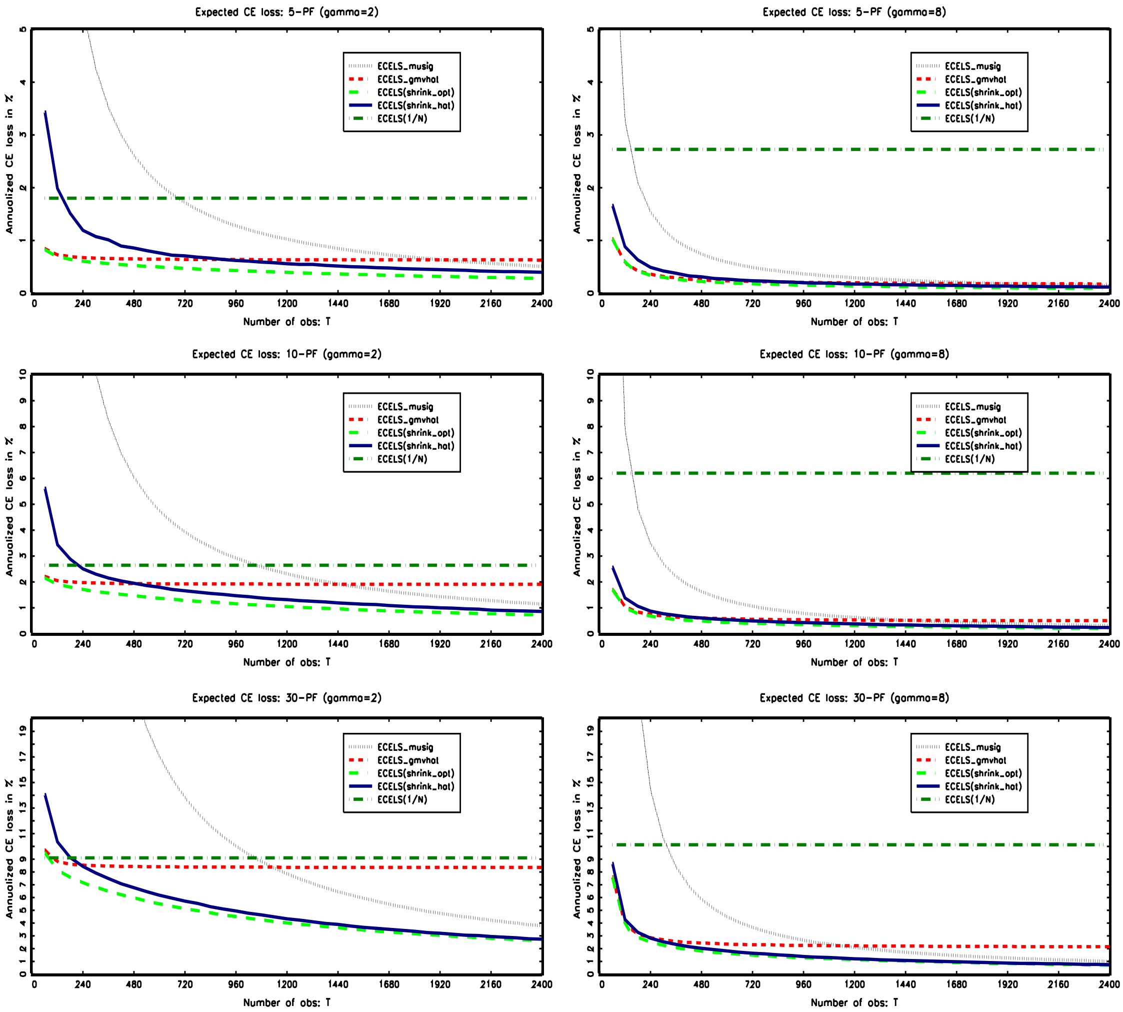


Figure 1.3: Expected CE loss of estimated efficient portfolio, estimated GMVP, equally weighted portfolio as well as theoretical and estimated shrinkage portfolios for two different degrees of risk aversion:
1) ECELS_musig - dotted line: $\mathcal{R}(w_{ep}(\bar{r}, S)|w^*_{ep})$, empirical efficient PF
2) ECELS_gmvhat - short dashed line: $\mathcal{R}(w_{gmv}(S)|w^*_{ep})$, empirical GMVP;
3) ECELS(shrink_opt) - long dashed line: $\mathcal{R}(w_{ep}(\eta^*, \bar{r}, S)|w^*_{ep})$, efficient shrinkage PF;
4) ECELS(shrink_hat) - solid line: $\mathcal{R}(w_{ep}(\hat{\eta}, \bar{r}, S)|w^*_{ep})$, shrinkage PF;
5) ECELS(1/N) - dotted-slashed line: $\mathcal{R}((1/N)\iota|w^*_{ep})$, equally weighted portfolio.

1.5.3.1 Deviation from Normality

The assumption of i.i.d. normality for the return vector can be easily relaxed if we only consider the estimation risk of the mean given Σ is known. In this case, we only need the assumption of a serially uncorrelated return series to derive the expected CE loss analytically. In the following, we study the robustness of our results for deviations from normality, if the covariance matrix has to be estimated. For this, we simulate i.i.d. student-t distributed returns with the same mean vectors and covariance matrices as we used previously, i.e.

$$r_t = \sqrt{\frac{\nu - 2}{W_t}} \cdot Y_t + \mu,$$

where $Y \overset{iid}{\sim} N(0, \Sigma)$, $W \overset{iid}{\sim} \chi^2_v$, ν is the degree of freedom of the multivariate t distribution, μ and Σ are the specified mean and covariance of r_t.

In order to study the effect of large kurtosis, we assume 5 degrees of freedom and compute the expected CE losses for the sample efficient portfolio, the empirical GMVP and the estimated shrinkage portfolio. Our estimates of the expected CE losses are based on 10000 simulations. The result is reported in Table 1.5. If returns are student-t distributed, the performance of the estimated portfolio deteriorates compared to the results based on normal returns, but the difference turns out to be rather small.

Table 1.5: Expected CE Loss for t-distributed Returns

	CE of $\hat{w}_{ep}$ (%)				CE of $\hat{w}_{gmv}$ (%)				CE of $\hat{w}_{shrink}$ (%)			
	$\gamma = 2$		$\gamma = 8$		$\gamma = 2$		$\gamma = 8$		$\gamma = 2$		$\gamma = 8$	
T	$\mathcal{N}$	t_5	$\mathcal{N}$	t_5	$\mathcal{N}$	t_5	$\mathcal{N}$	t_5	$\mathcal{N}$	t_5	$\mathcal{N}$	t_5
	5-PF											
60	-16,40	-17.73	-6.73	-7.25	9.13	9.03	-0.35	-0.74	7.84	7.75	-0.65	-1.03
180	2.65	2.46	-1.41	-1.65	9.28	9.24	0.25	0.08	8.95	8.91	0.16	-0.02
300	5.73	5.62	-0.54	-0.67	9.31	9.28	0.36	0.25	9.15	9.13	0.32	0.21
420	6.98	6.96	-0.18	-0.27	9.32	9.30	0.41	0.32	9.26	9.24	0.39	0.30
540	7.67	7.65	0.02	-0.06	9.33	9.31	0.43	0.36	9.32	9.30	0.43	0.36
660	8.10	8.07	0.14	0.07	9.33	9.32	0.45	0.39	9.35	9.34	0.46	0.40
	10-PF											
60	-68,23	-73,09	-16,42	-17,59	8,83	8,78	2,85	2,62	6,52	6,43	2,24	2,04
180	-6,88	-7,56	-0,26	-0,47	9,05	9,03	3,73	3,64	8,53	8,44	3,58	3,49
300	1,02	0,80	1,86	1,75	9,09	9,08	3,87	3,82	8,93	8,90	3,83	3,77
420	4,09	3,97	2,68	2,60	9,10	9,10	3,94	3,89	9,15	9,14	3,94	3,90
540	5,73	5,57	3,12	3,06	9,11	9,10	3,97	3,93	9,28	9,26	4,01	3,97
660	6,74	6,70	3,40	3,34	9,12	9,11	3,99	3,96	9,39	9,40	4,06	4,03
	30-PF											
60	-1232,35	-1349,00	-310,62	-345,43	7,68	7,63	-0,61	-0,86	4,33	3,98	-1,47	-1,82
180	-67,05	-71,67	-15,12	-16,44	8,80	8,78	3,84	3,74	8,42	8,42	3,74	3,63
300	-22,68	-24,61	-3,58	-4,12	8,92	8,90	4,32	4,27	9,60	9,53	4,49	4,43
420	-8,64	-9,78	0,11	-0,22	8,96	8,95	4,51	4,45	10,37	10,36	4,85	4,80
540	-1,84	-2,59	1,90	1,68	8,99	8,98	4,60	4,56	10,96	10,96	5,10	5,06
660	2,16	1,71	2,95	2,77	9,00	9,00	4,66	4,63	11,48	11,48	5,28	5,25

1.6 Conclusion

This paper takes a closer look at the quality of efficient portfolios compared to other portfolio strategies accounting for the budget constraint in a world where the parameters of the return process are unknown and have to be estimated by the investor. The relative performance of the different empirical portfolio strategies depends on the magical quadrangle between i.) the theoretical properties of the return process (e.g. the eigenvalues of the covariance matrix), ii.) the estimation properties of their sample counterparts, notably sample size, iii.) the number of assets and iv.) the investor's risk preferences. While in theory there is a well-defined dominant portfolio strategy, the ranking of portfolio strategies becomes unclear when the portfolio weights are based on estimated parameters. In this case, the ranking depends on the particular parametric constellation within the magical quadrangle. Therefore, in the light of estimation risk, a theoretically suboptimal portfolio can turn into a reasonable choice once sample size, size of the portfolio and risk preferences are taken into account.

Using the concept of implied means, we are able to represent the weights of any portfolio with the theoretically efficient portfolio weights. This allows us to decompose the overall estimation risk into their single components. Unlike previous empirical studies comparing the estimation risks in means and covariances separately, our study clearly shows why and when a precise estimate of covariances becomes necessary. Although the estimation of mean returns is crucial for almost all financial decisions, the impact of estimation error in covariances can never be neglected in the presence of estimation risk in the mean. Thus, to evaluate different covariance estimators, comparing their pure statistical metrics to the true covariance is questionable. An estimate of covariance matrix which has lower statistical risk is not necessarily superior if a financial decision rule is considered.

Based on the property that the weights of the efficient portfolio can be represented as the sum of the weights of the global minimum variance portfolio and the zero investment portfolio, we can show that Bayes-Stein shrinkage estimation of mean returns, shrinkage of the portfolio weights towards the weights of the global minimum variance portfolio as well as plug-in estimation of the efficient portfolio assuming a higher degree of risk aversion are equivalent strategies to reduce estimation risk. In a calibration study, we show that the expected CE loss of the efficient portfolio due to estimation is non-negligible for realistic empirical scenarios. Moreover, we show that shrinkage leads to superior choices of the portfolio weights compared to the empirical efficient portfolio but also compared to the simple $1/N$ strategy.

Admittedly, the results presented in this paper are based on the most simple portfolio set-up allowing us to derive finite sample properties for the estimated portfolio weights and the CE based on estimated portfolio weights. Extending our findings to the case of a dynamic price process, where the investor forms expectations on the return process given past filtration, would be desirable.
Moreover, as sample size is a major determinant of estimation risk particularly for large portfolios, optimal empirical portfolio strategies in the presence of structural breaks should be derived combing optimally pre- and post-break information.

Bibliography

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

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

Britten-Jones, M. (1999), 'The sampling error in estimates of mean-variance efficient portfolio weights', *The Journal of Finance* **54**(2), 655–671.

Cho, D. D. (2010), Estimation risk in covariance. Journal of Asset Management, (forthcoming).

Chopra, V. K. & Ziemba, W. T. (1993), 'The effect of errors in means, variances, and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**(2), 6–11.

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

Frahm, G. (2010), An analytical investigation of estimators for expected asset returns from the perspective of optimal asset allocation. University of Cologne, Seminar of Economic and Social Statistics, Discussion Paper 1/10.

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

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

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

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

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

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

1.7 Appendix

Lemma 1.7.1. *Let I denote the identity matrix. Define matrix $B = \Sigma \cdot A = I - \frac{\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}$. We have:*

1. *A is semi-positive definite and $A \cdot \Sigma \cdot A = A \cdot B = A$,*
2. *$A \cdot x = 0$ if and only if $x = \iota$,*
3. *$tr(B) = N - 1$.*

Proof 1.7.1. *See e.g .Okhrin & Schmid (2006).*

Proof 1.7.2 (*Proposition 1.2.1*)**.** *Substituting the solution (1.2.2) in the CE leads to:*

$$
\begin{aligned}
&CE(w_{ep})\\
=&\mu'\left(w_{gmv}+\frac{1}{\gamma}\cdot A\cdot\mu\right)-\frac{\gamma}{2}\left(w_{gmv}+\frac{1}{\gamma}\cdot A\cdot\mu\right)'\Sigma\left(w_{gmv}+\frac{1}{\gamma}\cdot A\cdot\mu\right)\\
=&\mu' w_{gmv}-\frac{\gamma}{2}w_{gmv}'\Sigma w_{gmv}+\frac{1}{\gamma}\mu'\cdot A\cdot\mu\\
&-\frac{\gamma}{2}\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)'\Sigma\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)-\gamma w_{gmv}'\Sigma\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)
\end{aligned}
\tag{1.7.1}
$$

Since $A \cdot \Sigma \cdot A = A$, and $\iota' \cdot A = 0$, we have

$$
\frac{\gamma}{2}\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)'\Sigma\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)=\frac{1}{2\gamma}\mu'\cdot A\cdot\mu
$$

and

$$
w_{gmv}'\Sigma\left(\frac{1}{\gamma}\cdot A\cdot\mu\right)=\frac{1}{\gamma}\cdot\frac{\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\Sigma\cdot A\cdot\mu=0
$$

Substituting the GMVP weight $w_{gmv} = \Sigma^{-1}\iota/\left(\iota'\Sigma^{-1}\iota\right)$ in equation (1.7.1) leads to:

$$
\begin{aligned}
CE(w_{ep}) =&\mu' w_{gmv}-\frac{\gamma}{2}w_{gmv}'\Sigma w_{gmv}+\frac{1}{\gamma}\mu'\cdot A\cdot\mu-\frac{1}{2\gamma}\mu'\cdot A\cdot\mu\\
=&CE(w_{gmv})+\frac{1}{2\gamma}\mu'\cdot A\cdot\mu\\
=&\frac{1}{2\gamma}\mu'\cdot A\cdot\mu+\frac{1}{\iota'\Sigma^{-1}\iota}(\iota'\Sigma^{-1}\mu-\frac{\gamma}{2})
\end{aligned}
$$

Proof 1.7.3 (*Proposition 1.2.2*).

$$\begin{aligned}
&CE(w_{ep}) - CE(\hat{w}) \\
=&[\mu' w_{ep} - \frac{\gamma}{2} w_{ep}' \Sigma w_{ep}] - [\mu'\hat{w} - \frac{\gamma}{2}\hat{w}'\Sigma\hat{w}] \\
=&[\mu' w_{ep} - \frac{\gamma}{2} w_{ep}' \Sigma w_{ep}] - [\mu'(w_{ep} + \hat{w} - w_{ep}) - \frac{\gamma}{2}(w_{ep} + \hat{w} - w_{ep})'\Sigma(w_{ep} + \hat{w} - w_{ep})] \\
=&[\mu' w_{ep} - \frac{\gamma}{2} w_{ep}' \Sigma w_{ep}] - \mu'(\hat{w} - w_{ep}) - \mu' w_{ep} \\
&+ \frac{\gamma}{2}(\hat{w} - w_{ep})'\Sigma(\hat{w} - w_{ep}) + \gamma(\hat{w} - w_{ep})\Sigma w_{ep} + \frac{\gamma}{2} w_{ep}' \Sigma w_{ep} \\
=&-\mu'(\hat{w} - w_{ep}) + \frac{\gamma}{2}(\hat{w} - w_{ep})'\Sigma(\hat{w} - w_{ep}) + \underbrace{\gamma(\hat{w} - w_{ep})'\Sigma w_{ep}}_{=(*)}
\end{aligned}$$

Since $w_{ep} = \frac{1}{\gamma}\Sigma^{-1}\mu + (1 - \frac{1}{\gamma}\iota'\Sigma^{-1}\mu)\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$ *and* $\sum_i w_i = \iota' w = 1$, *we have :*

$$\begin{aligned}
(*) &= (\hat{w} - w_{ep})'\mu + \underbrace{(\hat{w} - w_{ep})'\iota}_{=0} \frac{\gamma}{\iota'\Sigma^{-1}\iota} - \underbrace{(\hat{w} - w_{ep})'\iota}_{=0} \frac{\iota'\Sigma^{-1}\mu}{\iota'\Sigma^{-1}\iota} \\
&= (\hat{w} - w_{ep})'\mu
\end{aligned}$$

Therefore:

$$CE(w_{ep}) - CE(\hat{w}) = \frac{\gamma}{2}(w_{ep} - \hat{w})'\Sigma(w_{ep} - \hat{w}).$$

The risk function can be easily obtained by taking the expectation of the CE loss.

Proof 1.7.4 (*Proposition 1.2.3*). *If there is no error in covariance matrix we have:*

$$w_{ep} - w_{ep}(\hat{\mu}, \Sigma) = \frac{1}{\gamma} \cdot A \cdot (\mu - \hat{\mu})$$

Therefore the loss of CE is:

$$CE(w_{ep}) - CE\left(w_{ep}(\hat{\mu}, \Sigma)\right) = \frac{1}{2\gamma}(\mu - \hat{\mu})' \cdot A' \cdot \Sigma \cdot A(\mu - \hat{\mu}) = \frac{1}{2\gamma}(\mu - \hat{\mu})' \cdot A(\mu - \hat{\mu}).$$

The risk function can be easily obtained by taking the expectation of the CE loss.

Proof 1.7.5 (*Proposition 1.2.4*). *In this proposition we show that for any given portfolio* $\hat{w}$, *there is a implied mean vector* $\hat{\mu}_{im}$ *such that:*

$$\hat{w} \stackrel{!}{=} w_{ep}(\hat{\mu}_{im}, \Sigma) = w_{gmv} + \frac{1}{\gamma} \cdot A \cdot \hat{\mu}_{im} = w_{gmv} + \frac{1}{\gamma}\Sigma^{-1} \cdot B \cdot \hat{\mu}_{im}$$

Since ι*and the column vectors of* V *compose a basis of* $\mathbb{R}^N$, *the vector* $\hat{\mu}_{im}$ *can be written as linear combination of this basis, i.e.* $\hat{\mu}_{im} = c \cdot \iota + V \cdot c_0$ *with* $c_0 \in \mathbb{R}^{(N-1)}$. *Then:*

$$\hat{w} \stackrel{!}{=} w_{ep}(\hat{\mu}_{im}, \Sigma) = w_{gmv} + \frac{1}{\gamma}\Sigma^{-1} \cdot B \cdot \hat{\mu}_{im} = w_{gmv} + \frac{1}{\gamma}\Sigma^{-1} B \cdot V \cdot c_0 \tag{1.7.2}$$

because $A \cdot \iota = 0$. The solution of equation (1.7.2) is:

$$\begin{aligned} c_0 &= \gamma \cdot \left(V'BV\right)^{-1} \cdot V' \cdot \Sigma \cdot (\hat{w} - w_{gmv}) \\ &= \gamma \cdot \left(V'BV\right)^{-1} \cdot V' \cdot \Sigma \cdot \hat{w} - \gamma \cdot \left(V'BV\right)^{-1} \cdot V' \cdot \Sigma \cdot \frac{\Sigma^{-1}\iota}{\iota'\Sigma\iota} \\ &= \gamma \cdot \left(V'BV\right)^{-1} \cdot V' \cdot \Sigma \cdot \hat{w} \end{aligned}$$

because $V'\iota = 0$. Thus, $\hat{\mu}_{im} = c \cdot \iota + V \cdot c_0 = c \cdot \iota + \gamma \cdot V \cdot (V'BV)^{-1} \cdot V' \cdot \Sigma \cdot \hat{w}$

Proof 1.7.6 (*Proposition 1.2.5*)**.** *Based on proposition 1.2.3 and proposition 1.2.4, the CE loss of an suboptimal portfolio $\hat{w}$ can be reformulated as:*

$$\begin{aligned} & CE(w_{ep}) - CE(\hat{w}) \\ = \;& \frac{1}{2\gamma}(\mu - \hat{\mu}^*_{im})' \cdot A \cdot (\mu - \hat{\mu}^*_{im}) \\ = \;& \frac{1}{2\gamma}(\mu - \hat{\mu}^*_{im})' \cdot A \cdot \Sigma \cdot A \cdot (\mu - \hat{\mu}^*_{im}) \\ = \;& \frac{1}{2\gamma}(\mu - \hat{\mu}^*_{im})' \cdot \Sigma^{-\frac{1}{2}} \cdot (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota})' \cdot (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}) \cdot \Sigma^{-\frac{1}{2}} \cdot (\mu - \hat{\mu}^*_{im}) \end{aligned}$$

where $\hat{\mu}^*_{im}$ is the mean implied by the portfolio weight $\hat{w}$.
Let C denote the correlation matrix and D denote the diagonal matrix containing the standard deviations. Then $\Sigma = D \cdot C \cdot D$ and we have:

$$\begin{aligned} & CE(w_{ep}) - CE(\hat{w}) \\ \leq & \parallel (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}) \parallel_2^2 \parallel C^{-\frac{1}{2}} \parallel_2^2 \parallel D^{-1} \parallel_2^2 \parallel (\mu - \hat{\mu}^*_{im}) \parallel_2^2 \\ = & \parallel (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}) \parallel_2^2 \cdot trace\left(C^{-1}\right) \cdot trace\left(D^{-2}\right) \cdot \parallel (\mu - \hat{\mu}^*_{im}) \parallel_2^2 \\ = & \frac{1}{2\gamma}(\sum_{i=1}^{N} \lambda_i^{-1}) \cdot (\sum_{i=1}^{N} \sigma_i^{-2}) \parallel (I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}) \parallel_2^2 \parallel (\mu - \hat{\mu}^*_{im}) \parallel_2^2 \\ = & \frac{1}{2\gamma}(\sum_{i=1}^{N} \lambda_i^{-1}) \cdot (\sum_{i=1}^{N} \sigma_i^{-2})(N-1) \parallel (\mu - \hat{\mu}^*_{im}) \parallel_2^2 \end{aligned}$$

The last equation holds because:

$$
\begin{aligned}
\parallel I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota} \parallel_2^2 &= trace\left(\left[I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}\right]' \left[I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}\right]\right) \\
&= trace\left(\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\left[I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}\right]' \left[I - \frac{\Sigma^{-\frac{1}{2}}\iota\iota'\Sigma^{-\frac{1}{2}}}{\iota'\Sigma^{-1}\iota}\right]\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\right) \\
&= trace\left(\Sigma^{\frac{1}{2}} \cdot A \cdot \Sigma \cdot A \cdot \Sigma^{\frac{1}{2}}\right) \\
&= trace\left(\Sigma^{\frac{1}{2}} \cdot A \cdot \Sigma^{\frac{1}{2}}\right) \\
&= trace\left(\Sigma \cdot A\right) \\
&= N - 1
\end{aligned}
$$

Chapter 2

Portfolio with Non-negativity Constraints: Better or Worse?

2.1 Introduction

In practice, many portfolios are required by law or by policy to satisfy particular constraints. The most common and most important constraints are the adding-up (budget) and the non-negativity (no-short-sale) constraints which are both binding in the "standard portfolio selection" model defined by Markowitz (1952) and Markowitz (1987). Imposing the non-negativity constraint additional to the adding-up constraint in the portfolio optimization problem prevents extreme positions in optimized portfolios and makes them easy to implement in practice.

From a theoretical perspective, imposing constraints definitively hurts portfolio performance. As indicated by Green & Hollifield (1992), if the covariance structure is supposed to be dominated by a single factor, negative positions will be needed to diversify systematic risks[1]. However, from practical perspective, many studies have pointed out that portfolio constraints can significantly reduce the estimation risk, which is usually very large in the small sample case and dominates the theoretical loss caused by constraints. Jagannathan & Ma (2003) show that, even in the special scenario considered by Green & Hollifield (1992) where the true minimum variance portfolio contains extreme negative positions, restricting portfolio weights to be positive can indeed improve the performance of the plug-in estimator of portfolio weights based on the sample covariance matrix. Their theoretical results show that the non-negativity constraint on the portfolio weights can be interpreted as a form of shrinkage estimator of the covariance matrix. Since the shrinkage estimation

[1] In addition, if assets' exposures to this dominating factor are similar, diversifying the systematic risk could lead to extremely long and short positions in the resulted portfolio.

approach has proven to be a successful way to reduce the estimation risk of plug-in estimates in finite sample, it is not surprising that constraints also help even if they are not theoretically justified. DeMiguel, Garlappi, Nogales & Uppal (2009) generalize the non-negativity constraint to the L_1 norm constraint and relate it to the LASSO regression model. Fan, Zhang & Yu (2012) provide a novel statistical insight into the L_1-norm constraint in vast portfolio selection problems and show that the constraint can prevent the optimized portfolios from aggregation of estimation errors in the covariance matrix. However, to avoid the effect of estimation errors in the mean on portfolio weights, all these studies give up effort on mean estimation and limit themselves on the global minimum variance portfolio (GMVP).

In contrast, some recent studies suggest to incorporate the information of the estimated mean in portfolio selection models, though it is hard to estimate precisely. Kan & Zhou (2007) argue that combining two stochastically independent portfolios can significantly reduce the variance of estimated portfolio weights and thus improves the out-of-sample performance. They therefore propose a three-fund portfolio rule to combine the cash, the estimated tangency portfolio and the estimated GMVP, and show that the three-fund portfolio rule can outperform the GMVP in terms of investor's utility even in the finite sample case. Liu & Pohlmeier (2013) focus on efficient portfolios satisfying the adding-up constraint and show that the estimation risk of estimated portfolios crucially depends on investors' risk preferences. They propose an optimal shrinkage approach to shrink the efficient portfolio with a given risk preference parameter to the GMVP where the shrinkage intensity can be equivalently expressed as a scaling parameter on the expected returns of assets[2].

It is well known that there is no closed-form solution to the portfolio optimization problem with the non-negativity constraint. Although several empirical studies have been made in the literature, many crucial properties of constrained portfolios still remain unclear: for instance, how the performance of constrained portfolios is affected by the investor's risk preference and what potential risks there are when imposing constraints on portfolio choice problem. In this paper, we focus on the general mean-variance portfolio selection problem for investors with different risk preferences rather than the GMVP where the investor is assumed to be extremely risk averse. We also theoretically derive a shrinkage interpretation of the non-negativity con-

[2] In this case, shrinking the efficient portfolio to the GMVP is equivalent to shrinking the estimated mean vector to the portfolio mean of GMVP. This is similar to the shrinkage estimator of the mean proposed by Jorion (1986). Antoine (2012) proposes a similar shrinkage approach for the tangency portfolio and interprets the shrinkage intensity as *corrected risk aversion parameter* which is larger than the original investor's risk aversion parameter: an investor selects portfolios under parameter uncertainty should be more conservative than in the situation where everything is known.

straint on efficient portfolios, in which not only the covariance matrix but also the mean vector are shrunk to particular targets and the shrinkage intensities depend on the investor's risk preference. In addition, since the shrinkage target of the mean is proportional to the original mean, the shrinkage effect on the mean can be easily represented as a scaling effect on the variance elements in the covariance matrix.

To demonstrate the effect of the non-negativity constraint intuitively, we conduct a simulation study with realistic inputs from several financial data sets. The empirical and theoretical losses of estimated portfolios are defined in terms of their certainty equivalents (CE). We find that the trade-off between the theoretical and empirical losses of constrained portfolios is highly affected by the investor's risk aversion level. It is known that, when the non-negativity constraint is binding, the efficient frontier is constructed by different segments and in each segment only a subset of assets is active, i.e. their corresponding portfolio weights are strictly positive. Therefore, we further focus on the active assets in the constrained portfolio and compute the probability of the active assets being correctly identified. Our simulation results show that, in the present of estimation risk, it seems very difficult to identify the active asset set for a large range of risk aversion levels. From practical perspective, this implies large re-balancing needs and consequently high transaction costs.

Finally, different portfolio strategies with and without the non-negativity constraint are compared on simulated and empirical data. It is shown that, for a wide range of risk preference parameters, constrained portfolio strategies, especially the constrained GMVP, perform usually worse than some conservative but unconstrained portfolios such as the unconstrained GMVP, the dominating estimator of GMVP proposed by Frahm & Memmel (2010), and the shrinkage portfolio proposed by Liu & Pohlmeier (2013).

This article is organized as follows. Section 2.2 reviews the theoretical and empirical findings of Jagannathan & Ma (2003) and proposes a shrinkage interpretation of the non-negativity constraint for the efficient portfolio case. Section 2.3 illustrates the impacts of the non-negativity constraint on portfolio performance by simulation. Section 2.4 compares different competitive portfolio strategies using empirical data. Section 2.5 concludes.

2.2 Shrinkage Interpretation

Suppose there are N risky assets and the investor can only invest in those assets. Let the $N \times 1$ random vector $\mathbf{r}_t = (r_{1t}, \ldots, r_{Nt})$ denote the returns of risky assets at time t with mean $\mathrm{E}[r_t] = \mu$ and covariance matrix $\mathrm{V}[r_t] = \Sigma$. Let w denote the weight vector of a portfolio and ι be the vector of ones. In addition, let $\hat{\mu}$ and $\hat{\Sigma}$ denote the estimates of μ and Σ, respectively.

2.2.1 Global Minimum Variance Portfolio

The global portfolio variance minimization problem with non-negativity constraint is given by:

$$\begin{aligned} &\min_w \; w'\Sigma w \\ \text{s.t.} \quad &\iota' w = 1 \\ &w_i \geq 0 \quad \forall\, i = 1, \ldots, N \end{aligned} \tag{2.2.1}$$

This problem is solved by minimizing the following Lagrange function:

$$L = \frac{1}{2} w'\Sigma w - \xi' w - \lambda_1(\iota' w - 1)$$

where $\xi = (\xi_1, \ldots, \xi_N)$ is the vector of Lagrange multipliers for the non-negativity constraint and λ_1 is the Lagrange multiplier for the adding-up constraint. Because of the non-negativity constraint, the standard Lagrange multiplier method is no longer sufficient for solving the constrained optimization problem (2.2.1), and the well known Kuhn-Tucker conditions are needed:

$$\begin{aligned} \Sigma w - \lambda_1 \iota &= \xi, \\ \xi' w &= 0, \\ \xi_i \geq 0 \text{ and } w_i \geq 0 \quad &\forall \quad i = 1 \ldots N. \end{aligned}$$

Denote the solution to the constrained portfolio variance minimization problem (2.2.1) as $w^+_{gmv}(\Sigma)$. Jagannathan & Ma (2003) show that the constrained problem (2.2.1) is equivalent to the unconstrained variance minimization problem:

$$\min_w \; w'\tilde{\Sigma} w \qquad \text{s.t.} \quad \iota' w = 1 \tag{2.2.2}$$

with the modified covariance matrix

$$\tilde{\Sigma} = \Sigma - (\iota'\xi + \xi'\iota) \tag{2.2.3}$$

which is positive semi-definite. Therefore, the non-negativity constraint has a shrinkage-like effect on the covariance matrix used in forming the GMVP.

Note that, the shrinkage interpretation of the non-negativity constraint can be expressed by either using the population covariance matrix Σ or its estimate $\hat{\Sigma}$. While shrinking the estimated covariance matrix generally reduces the estimation error, shrinking the population covariance matrix introduces the specification error and consequently causes the theoretical loss of the constrained GMVP relative to the true unconstrained one. This trade-off between the theoretical and empirical losses plays a crucial role for the performance of constrained portfolios.

Jagannathan & Ma (2003) examine the trade-off between the theoretical and empirical losses using both simulation and empirical studies. They consider three different cases in the simulation where the non-negativity constraint is correct, moderately wrong or severely wrong in population[3]. They find that, if the sample size is small, restricting portfolio weights to be positive can significantly reduce the estimation error which dominates the theoretical loss caused by the constraint, and therefore can improve the performance of empirical GMVP even in the case where the constraint is wrong in population. This improvement is especially large if the portfolio dimension is large. On the other hand, if the constraint is correct in population, constrained GMVP based on the sample covariance matrix performs better than the empirical unconstrained GMVP but worse than the equally weighted portfolio, because in this case, the equally weighted portfolio is not far away from the theoretical GMVP based on the population covariance matrix.

In their empirical study, Jagannathan & Ma (2003) compare constrained and unconstrained portfolio strategies based on both monthly and daily data. They also consider various estimators of the covariance matrix suggested in the literature, such as the sample covariance matrix, the shrinkage estimator as well as covariance matrices estimated from different factor models. Their empirical results based on monthly data coincide with the theoretical and simulation results. Interestingly, they find that, if the daily data is used instead of monthly data, the unconstrained GMVP based on the sample covariance matrix performs the best in terms of the

[3] It means whether the GMVP based on the population covariance matrix contains negative positions or not.

out-of-sample annualized standard deviation of the portfolio. If the non-negativity constraint is in place, the GMVPs based on different covariance estimators perform all similar. Therefore it seems difficult to improve the portfolio performance by adopting precise estimates if the non-negativity constraint is binding.

2.2.2 Efficient Portfolio

Jagannathan & Ma (2003) analyze intensively the impact of the non-negativity constraint on the GMVP and highlight the benefit of imposing the constraint. However, their empirical findings also show some potential problems: while imposing the constraint shrinks the sample covariance matrix and reduces the estimation risk in the empirical portfolio, an improved portfolio performance can only be achieved if the loss due to the estimation risk is large and dominates the theoretical loss caused by the constraint. The trade-off between the theoretical and empirical losses depend on the population properties of return and is much more complicated in the efficient portfolio case where the investors' risk preferences are taken into account.

The mean-variance optimization problem with both adding-up and non-negativity constraints can be formulated as:

$$\begin{aligned} & \max_w CE(w) = \max_w \{\mu' w - \frac{\gamma}{2} w' \Sigma w\}, \\ \text{s.t.} \quad & \iota' w = 1 \\ & w_i \geq 0 \quad \forall\, i = 1, \ldots, N \end{aligned} \tag{2.2.4}$$

where the parameter $\gamma \in (0, \infty]$ reflects the investor's risk aversion level. This maximization problem can be solved by minimizing the following Lagrange function:

$$L = \frac{1}{2} w' \Sigma w - \frac{1}{\gamma} w' \mu - \xi' w - \lambda_1 (\iota' w - 1)$$

where $\xi = (\xi_1, \ldots, \xi_N)$ are the Lagrange multipliers for the non-negativity constraint and λ_1 is the Lagrange multiplier for the adding-up constraint. Then the Kuhn-Tucker conditions for the constrained maximization problem (2.2.4) are given by:

$$\Sigma w - \frac{1}{\gamma} \mu - \lambda_1 \iota = \xi, \tag{2.2.5}$$

$$\xi' w = 0, \tag{2.2.6}$$

$$\xi_i \geq 0 \text{ and } w_i \geq 0 \quad \forall \quad i = 1 \ldots N. \tag{2.2.7}$$

Let $w^{+}_{ep}(\mu, \Sigma)$ denote the solution to the constrained optimization problem (2.2.4) based on μ and Σ. Let $\mu_p^+ = \mu' w_{ep}^+(\mu, \Sigma)$ and $\sigma_p^{2+} = w_{ep}^+(\mu, \Sigma)'\Sigma w_{ep}^+(\mu, \Sigma)$ denote the portfolio mean and the portfolio variance evaluated at $w_{ep}^+(\mu, \Sigma)$, respectively. Similar to the constrained GMVP case, we have the following result for efficient portfolios with the non-negativity constraint:

Lemma 2.2.1. *Assume that the portfolio variance and mean are strictly positive: $\sigma_p^{2+} > 0$ $\mu_p^+ > 0$. Let $\xi = (\xi_1, \ldots, \xi_N)$ denote the Lagrange multipliers for the non-negativity constraint and λ_1 denote the Lagrange multiplier for the adding-up constraint. Define:*

$$\begin{aligned} \tilde{\Sigma} &= \Sigma - (\iota\xi' + \xi\iota') + k_1 \cdot \mu\mu' + k_2 \cdot \iota\iota' \\ \tilde{\mu} &= (1 + k_1 \cdot \gamma \cdot \mu_p^+)\mu \end{aligned}$$

where

$$k_1 = \frac{1}{\lambda_1\gamma^2 + \mu_p^+\gamma}, \qquad k_2 = 2\frac{\mu_p^+}{\gamma}.$$

Then

1. *$\tilde{\Sigma}$ is symmetric and positive semi-definite,*
2. *the constrained optimization problem* (2.2.4) *is equivalent to unconstrained mean-variance optimization problem:*

$$\max_w CE(w) = \max_w \{\tilde{\mu}'w - \frac{\gamma}{2} w'\tilde{\Sigma}w\}, \qquad \text{s.t.} \quad \iota'w = 1 \tag{2.2.8}$$

Proof 2.2.1. *See Appendix A.*

Lemma 2.2.1 shows that, for a given risk preference parameter γ, the weight of the constrained efficient portfolio constructed from μ and Σ is exactly the same as the weight of the unconstrained efficient portfolio constructed from the modified mean $\tilde{\mu}$ and the modified covariance matrix $\tilde{\Sigma}$. Since the GMVP is only a special portfolio on the efficient frontier, this result can be considered as a generalization of the result from Jagannathan & Ma (2003): if the investor is extremely conservative with the risk preference parameter γ approaching infinity, we obtain the result of Jagannathan & Ma (2003) described in Section 2.2.1. However, different from the GMVP case, the shrinkage target of the covariance matrix is composed of three matrices: the shrinkage target in the constrained GMVP case, $\mu\mu'$ and $\iota\iota'$. Obviously, both of the latter two target matrices are singular. Therefore, if the investor's preference parameter is very small, the modified covariance matrix $\tilde{\Sigma}$ is nearly singular.

Note that, to make the shrinkage interpretation comparable with the result from Jagannathan & Ma (2003), the mean is also shrunk toward a certain target depending on the investor's risk preference. This is not the only way to express the shrinkage-like effect of the non-negativity constraint. Since the shrinkage target is proportional to the original mean vector, this shrinkage effect on the mean can be represented as a scaling effect on the covariance matrix, and we have the following result.

Proposition 2.2.1.
Assume that the portfolio variance and mean are strictly positive: $\sigma_p^{2+} > 0$ *and* $\mu_p^+ > 0$. *Then the constrained optimization problem* (2.2.4) *is equivalent to the unconstrained mean-variance optimization problem:*

$$\max_w CE(w) = \max_w \{\mu' w - \frac{\gamma}{2} w' \bar{\Sigma} w\}, \quad \text{s.t.} \quad \iota' w = 1 \tag{2.2.9}$$

with the positive semi-definite matrix

$$\bar{\Sigma} = c_1 \Sigma - c_1 (\iota' \xi + \xi' \iota) + c_2 \cdot \mu\mu' + c_3 \cdot \iota\iota' \tag{2.2.10}$$

where

$$c_1 = \frac{\lambda_1 \gamma + \mu_p^+}{\lambda_1 \gamma + 2\mu_p^+}, \quad c_2 = \frac{1}{\lambda_1 \gamma^2 + 2\gamma\mu_p^+}, \quad c_3 = \frac{2\mu_p^+(\lambda_1 \gamma + \mu_p^+)}{\lambda_1 \gamma^2 + 2\gamma\mu_p^+}$$

Proof 2.2.2. *See Appendix A.*

In Proposition 2.2.1, both the original covariance matrix Σ and the shrinkage target for GMVP, $(\iota' \xi + \xi' \iota)$, are scaled by a constant c_1 which is strictly smaller than one for all positive and finite values of γ. This is equivalent to scaling all variance elements in the covariance matrix toward zero but keeping correlations unchanged. As γ goes to zero, c_1 converges to $1/2$ and the scaling effect become less important, while the last two shrinkage targets, $\mu\mu'$ and $\iota\iota'$, become dominating and lead to increasing correlations in the modified covariance matrix $\bar{\Sigma}$.

In addition, as γ converges to zero, the shrinkage intensity c_2 approaches faster to infinity than c_3. Thus, the matrix $\mu\mu'$ has the dominating impact on the modified covariance matrix $\bar{\Sigma}$ when the investor's risk aversion is small. In this case, if the mean is estimated with large estimation errors, a precise estimate of the covariance matrix might be shrunk to the noisy target. This implies a potential risk of volatility evaluation approaches based on portfolio performance which will be shown later in the simulation study.

2.3 Simulated Data

In this section, we use simulations to demonstrate the effect of the non-negativity constraint on the portfolio selection problem. We consider four different data sets containing monthly returns of 1) 5 industry portfolios (5PF); 2) 25 portfolios formed on Size and Book-to-Market (25PF); 3) 30 industry portfolios (10PF); 4) 48 industry portfolios (48PF) for the sample period 07/1926 - 09/2009. These datasets are published on Kenneth French's Web site[4] and used as the standard test assets in recent empirical studies to compare different portfolio strategies[5].

We first provide some intuition about the data. It is known that, when the non-negativity constraint is in place, the efficient frontier is constructed by several hyperbolic pieces which is called efficient segments by Markowitz (1987). Two adjacent efficient segments are linked to each other and the points at which the efficient segments intersect is called corner portfolios. In each corner portfolio, and therefore in each segment, only a subset of assets are active, i.e. their corresponding portfolio weights are strictly positive. For a given risk preference parameter γ, we denote this subset containing active assets by IN_γ and let $\sharp IN_\gamma$ denote the number of active assets in the IN_γ set. Based on the sample means and the sample covariance matrices of the four data sets, we compute all corner portfolios and report their corresponding risk preference parameters, portfolio means and the numbers of active assets in Table 2.1. For a portfolio between the two corner portfolios, the number of active assets is the same as the corner portfolio with the higher risk aversion level, for instance, in the 5PF case, the portfolio with $\gamma = 10$ contains 4 active assets.

We can observe that, for all data sets and all risk preference levels, the non-negativity constraint is wrong in population, and it becomes severely wrong as the risk aversion parameter approaches to zero. The reason is clear: if the risk aversion level is low, an investor cares more about the expected return than the variance of the portfolio. In this case, the investor has smaller need of risk diversification and concentrates on the assets providing highest expected returns.

[4] *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

[5] See for instance Kan & Zhou (2007) DeMiguel, Garlappi & Uppal (2009), Brodie, Daubechies, Molc, Giannone & Loris (2009).

Table 2.1: Risk preference, portfolio mean and number of active assets of corner portfolios.

			5PF				
γ	$+\infty$	3.352	1.564	0.897			
μ_p^+	0.010	0.010	0.010	0.011			
$\sharp IN_\gamma$	4	3	2	1			
			25PF				
γ	$+\infty$	11.754	7.204	6.479			
μ_p^+	0.009	0.009	0.010	0.010			
$\sharp IN$	3	4	3	4			
- - -	- - -	- - -	- - -	- - -	- - -		
γ	5.077	3.421	2.392	1.496	1.111		
μ_p^+	0.010	0.012	0.014	0.015	0.017		
$\sharp IN$	5	4	3	2	1		
			30PF				
γ	$+\infty$	475.090	13.496	12.403	8.895	4.196	
μ_p^+	0.009	0.009	0.010	0.010	0.010	0.011	
$\sharp IN$	6	7	6	7	8	7	
- - -	- - -	- - -	- - -	- - -	- - -	- - -	
γ	3.385	2.503	1.824	0.210	0.160	0.059	
μ_p^+	0.012	0.012	0.012	0.012	0.012	0.013	
$\sharp IN$	6	5	4	3	2	1	
			48PF				
γ	$+\infty$	64.935	47.475	18.216	16.449	8.982	
μ_p^+	0.009	0.010	0.010	0.011	0.011	0.011	
$\sharp IN$	9	10	11	10	9	10	
- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
γ	6.150	4.813	4.772	4.361	3.275	2.964	
μ_p^+	0.012	0.013	0.013	0.013	0.013	0.014	
$\sharp IN$	11	10	9	8	7	6	
- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -
γ	2.579	2.437	2.177	1.980	1.629	1.463	0.372
μ_p^+	0.014	0.014	0.014	0.015	0.015	0.015	0.015
$\sharp IN$	7	6	5	4	3	2	1

Comparing to the 30PF case, constrained portfolios in the 25 case contains much less active assets although the portfolio dimensions in both cases are similar. Thus we can expect that constrained portfolios in the 25PF case have larger theoretical loss than those in the 30PF case.

In addition, we can observe that the means of corner portfolios are vary similar to each other. This is different from the case of unconstrained efficient portfolios. For a given γ, the mean and the variance of the unconstrained efficient portfolio are given by:

$$\begin{aligned} \mu_p &= \mu_{gmv} + \frac{1}{\gamma}\Delta_{SSR} \\ \sigma_p^2 &= \sigma_{gmv}^2 + \frac{1}{\gamma^2}\Delta_{SSR} \end{aligned}$$

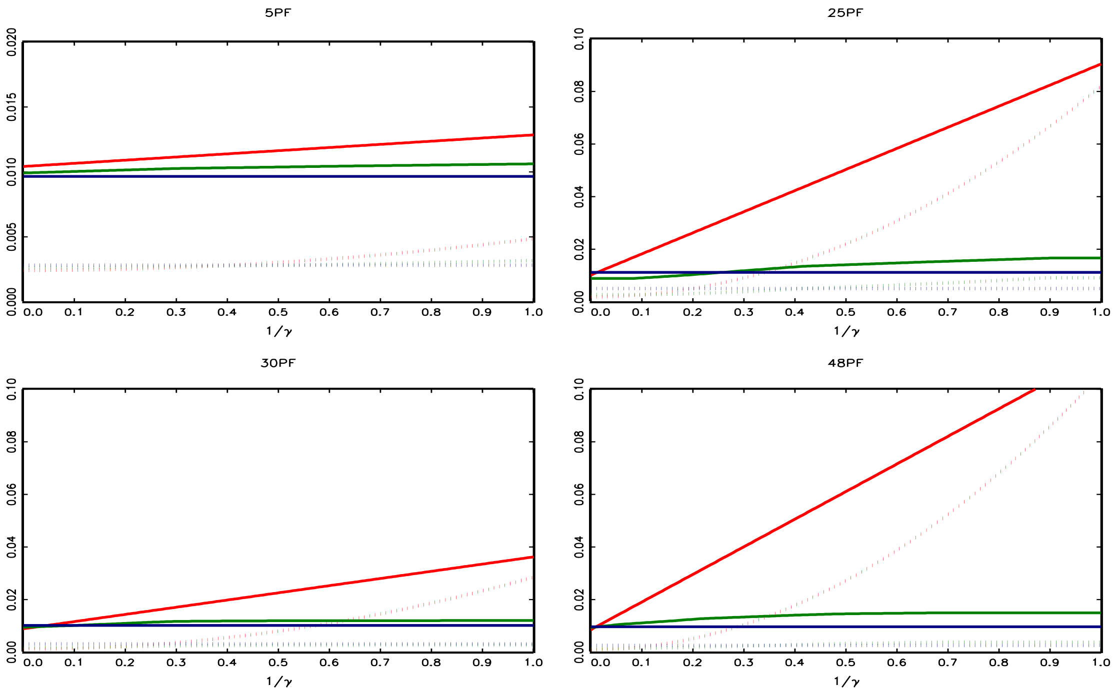


Figure 2.1: Portfolio mean and portfolio variance of:
1. theoretical unconstrained efficient portfolio: red solid line(portfolio mean) and red dotted line (portfolio variance);
2. theoretical non-negativity constrained portfolio: green solid line(portfolio mean) and green dotted line (portfolio variance);
3. equally weighted portfolio: blue solid line(portfolio mean) and blue dotted line (portfolio variance).

where $\mu_{gmv} = \frac{\mu'\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$ and $\sigma_p^2 = \frac{1}{\iota'\Sigma^{-1}\iota}$ are the mean and variance of GMVP, and

$$\Delta_{SSR} = \mu' A \mu \quad \text{with} \quad A = \Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}. \tag{2.3.1}$$

Therefore, in the unconstrained case, the portfolio mean is linear in $1/\gamma$ and the portfolio variance is quadratic in $1/\gamma$. Figure 2.1 plots the mean and variance of unconstrained efficient portfolios, non-negativity constrained efficient portfolios, and the equally weighted portfolio. For all four data sets considered, while the mean and the variance of the theoretical efficient portfolio are both increasing in $1/\gamma$, the mean and the variance of the constrained portfolio stay almost unchanged and are similar to those of the equally weighted portfolio. This suggests a large theoretical loss of constrained portfolios relative to the true theoretical efficient portfolios for investors with small risk aversion levels.

Moreover, since the risk aversion level of investors are difficult to identify, the mean-variance optimization problem sometimes is expressed as:

$$
\begin{aligned}
&\min_{w} w'\Sigma w, \\
\text{s.t.}\quad &\iota' w = 1, \quad \iota' w = \mu_0 \\
&w_i \geq 0 \quad \forall\, i = 1, \ldots, N
\end{aligned} \tag{2.3.2}
$$

where μ_0 is the investor's target expected return. Obviously, if the non-negativity constraint is in place, two investors with very different risk preferences could have similar target returns, i.e. a small change in the target return could lead to a totally different portfolio decision. Therefore, calculating the optimal portfolio based on the optimization problem (2.3.2) might be misleading, especially when the expected returns of assets are close to each other.

2.3.1 Theoretical v.s. Empirical Losses

In the following section, we compare the performance of estimated constrained and unconstrained portfolios in simulation. The portfolio performance is evaluated in terms of the Certainty Equivalent return (CE). Since the efficient portfolio calculated from the true mean and the true covariance matrix dominates all other portfolios subject to the adding-up constraint, the CE difference:

$$
\mathcal{L}(\hat{w}, w^*_{ep}) \equiv CE(w^*_{ep}) - CE(\hat{w}) \geq 0 \tag{2.3.3}
$$

is a well defined loss function and will be used to compare different portfolio strategies.

We choose the sample estimates from the four data sets as the true mean and true covariance matrix for our simulations. We consider four different sample sizes: $T = 60,\ 120,\ 180$, and 240. For each sample size, we simulate 10,000 times i.i.d multivariate normally distributed return vector, and estimate sample portfolio weights with and without the non-negativity constraint for different risk preference parameters. Figure 2.2 displays the means as well as the 90% confidence intervals of their CE Losses for $T = 60$ and 240[6]. To clearly illustrate the simulation results, we split the value of γ in two ranges: $\gamma > 1$ and $\gamma \leq 1$.

In general, constrained portfolios dominate the corresponding unconstrained portfolios for most risk aversion levels, but the dominance becomes insignificant if the investor is highly risk averse. A large sample size leads to significant improvement

[6]The results for $T = 120,\ 180$ can be found in Appendix B

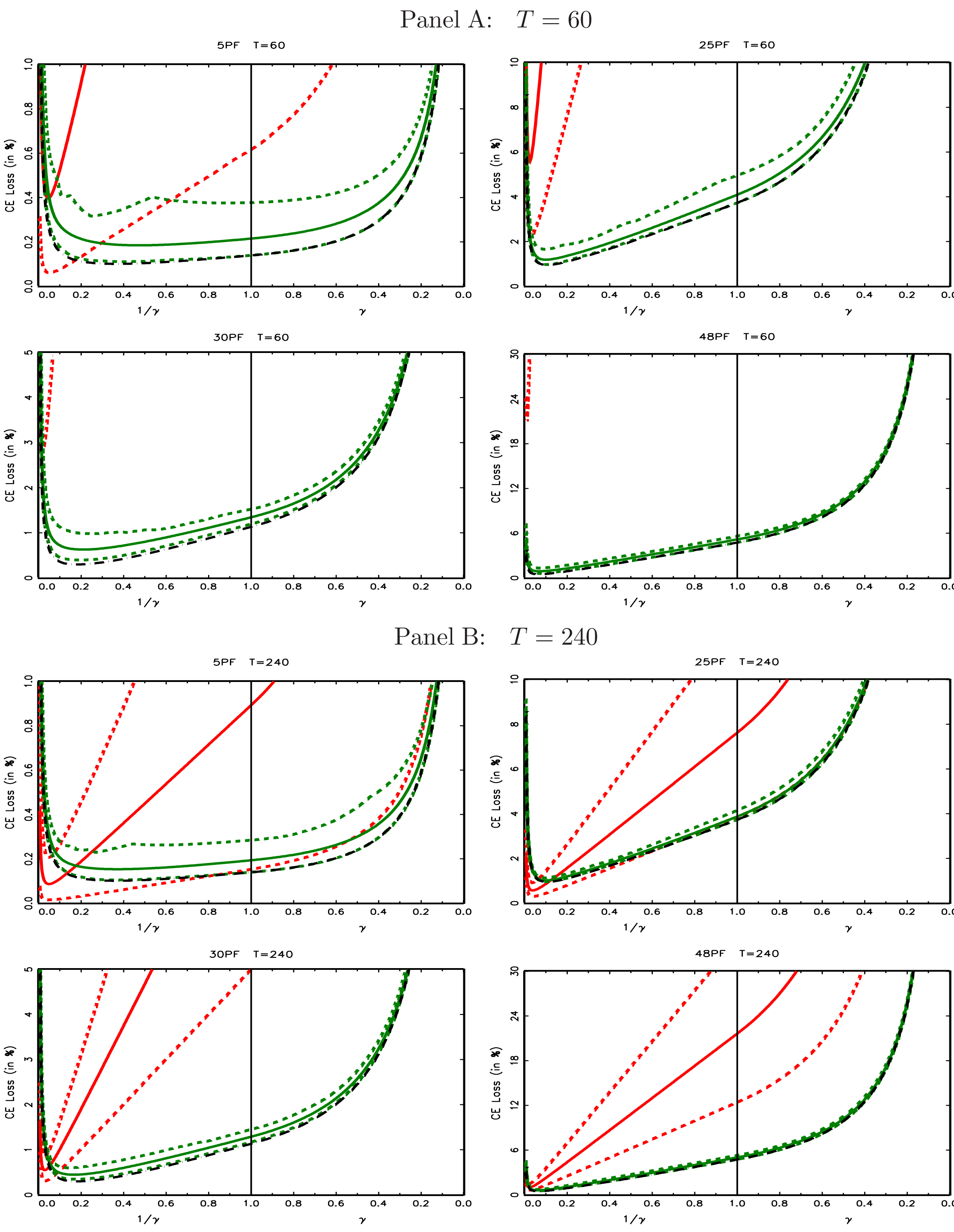


Figure 2.2: CE loss of the estimated unconstrained efficient portfolio and the estimated non-negativity constrained portfolio:
1. unconstrained portfolio: red solid line (mean), red dotted lines (90% confidence interval);
2. constrained portfolio: green solid line (mean), green dotted lines (90% confidence interval).

for the empirical unconstrained portfolio, while such improvement is limited in the constrained case.

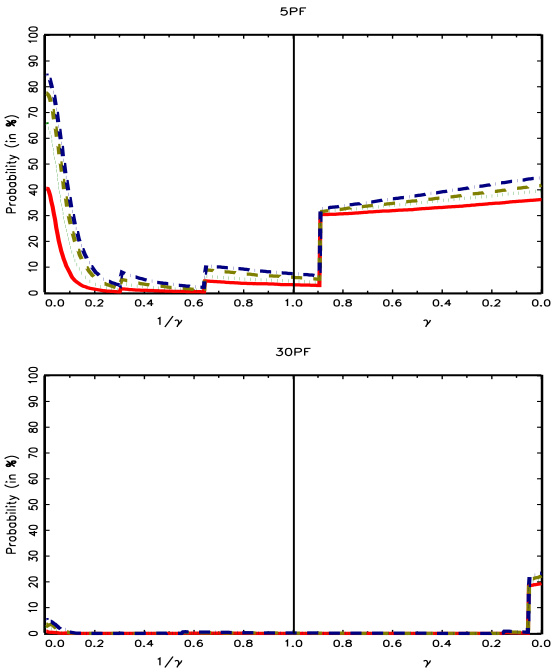


Figure 2.3: Probability of correctly identifying the IN_γ set for different sample size: $T = 60$ (red solid line); $T = 120$ (green closely spaced dots); $T = 180$ (brown dashed line); $T = 240$ (blue dots and dashes).

In addition, different from the case of estimated unconstrained portfolio where the CE loss is exclusively caused by estimation errors, the CE loss of estimated constrained portfolios can be decomposed into the empirical and theoretical losses. Figure 2.2 shows that, for investors with very high or very low risk aversion level, the theoretical loss dominates, but for investors with moderate risk aversion level, the empirical loss dominates. This fact can be explained by the following simulation study.

Since each non-negativity constrained portfolio only contains a subset of assets with non-zero weights, we use simulation to study how the constraint affects the appearance of asset candidates in this IN_γ set. Figure 2.3 shows the probability of correctly identifying the IN_γ set with respect to different sample sizes for our four data sets. We see that the probability of correct identification is in general very low, and almost equal to zero for most values of risk aversion in large dimension cases (the 30PF and 48PF cases). Even in the 5PF case, the probabilities of obtaining the correct active assets are very low for the risk aversion parameter γ in the range from 1 to 10, which is however commonly adopted in the portfolio literature. Therefore, although the true data generating process stays unchanged, estimating the non-negativity constrained portfolio from different samples might result in different subsets of active assets which leads to a large need of portfolio re-balancing and consequently large transaction costs.

2.3.2 Evaluating Covariance Estimators

Portfolio performance is a natural and widely accepted economic measure for evaluating the forecast of the covariance matrix. The better forecast of the covariance matrix should provide superior portfolio performance. However, our theoretical results show that, the non-negativity constraint shrinks the covariance matrix to the target depending on the mean. Therefore, if the mean is estimated with large errors, using portfolio performance to judge the accuracy of the forecast might be misleading.

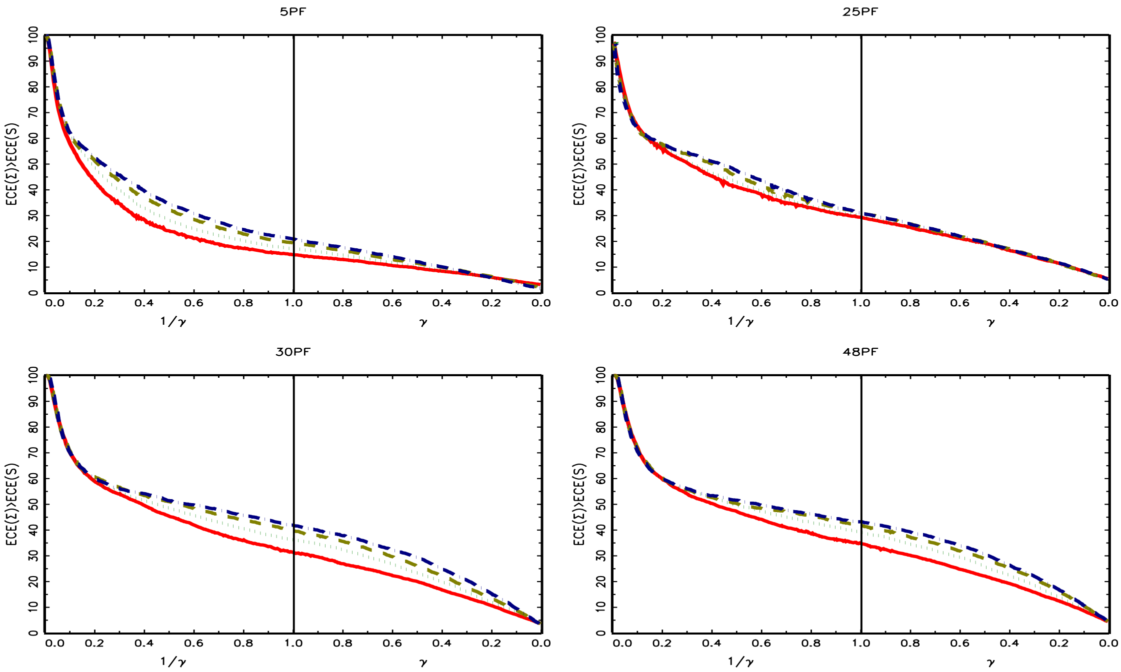


Figure 2.4: Probability of the portfolio constructed from the true covariance matrix being dominating, i.e. $\Pr\left(\mathrm{E}[CE(w_{ep}(\hat{\mu}, \Sigma))] > \mathrm{E}[CE(w_{ep}(\hat{\mu}, S))]\right)$ where Σ and S are the true and sample covariance matrices, respectively.
Sample size: $T = 60$ (red solid line); $T = 120$ (green closely spaced dots); $T = 180$ (brown dashed line); $T = 240$ (blue dots and dashes).

We calculate the probability that portfolios constructed from the true covariance matrix outperform those portfolios estimated by the sample covariance matrix and plot it in Figure 2.4. We discover that the probability decreases dramatically with decreasing risk aversion level. Increasing the sample size can only slightly improve the result. Therefore, in the case of finite risk aversion level, using the CE of estimated non-negativity constrained portfolio based on estimated expected return can potentially lead to wrong covariance matrix selection.

2.3.3 Comparison of Portfolios

In this section, we compare the performance of some competitive portfolio strategies in terms of their expected CE losses. Let $\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and $S = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})(r_t - \bar{r})'$ denote the sample mean and the sample covariance matrix respectively. We consider the following portfolio strategies:

1. equally weighted portfolios (EW):

$$w_{ew} = \frac{\iota}{N}; \tag{2.3.4}$$

2. non-negativity constrained efficient portfolio (NEP) based on sample mean and sample covariance matrix;

3. global minimum variance portfolio (GMVP) based on the sample covariance matrix:

$$\hat{w}_{gmv}(\Sigma) = \frac{S^{-1}\iota}{\iota' S^{-1}\iota}; \tag{2.3.5}$$

4. non-negativity constrained global minimum variance portfolio (NGMVP) based on the sample covariance matrix;

5. dominating estimator of GMVP of Frahm & Memmel (2010)(DOM) based on the sample covariance matrix:

$$w_{dom} = c\frac{\iota}{N} + (1-c)\hat{w}_{gmv}$$

with

$$c = \frac{T-3}{T-N+2}\frac{\hat{\sigma}^2_{gmv}}{\hat{\sigma}^2_{\iota} - \hat{\sigma}^2_{gmv}}$$

where $\hat{\sigma}^2_{gmv} = \frac{1}{\iota' S^{-1}\iota}$ and $\hat{\sigma}^2_{\iota} = \iota' S \iota$ are the estimated variance of GMVP and equally weighted portfolio based on the sample covariance matrix, respectively.

6. shrinkage portfolio of Liu & Pohlmeier (2013) (SHR) based on sample mean and sample covariance matrix:

$$w_{shrink}(c^*, \hat{\mu}, \hat{\Sigma}) = \frac{\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota} + \frac{c^*}{\gamma} \cdot \hat{A} \cdot \bar{r}$$

where

$$c^* = \frac{(T-N)(T-N-3)}{(T-1)(T-2)} \cdot \frac{\Delta_{SSR}}{\left(\Delta_{SSR} + \frac{N-1}{T}\right)}$$

with the matrix Δ_{SSR} given in equation (2.3.1) is the theoretical optimal shrinkage parameter and $\hat{A}$ is the plug-in estimator of the matrix A based on the sample covariance matrix. To avoid additional estimation risks from the nuisance parameter Δ_{SSR}, we first assume that it is known, i.e. calculated from the true mean and the true covariance matrix.

Portfolio strategies 1, 2 and 4 could be referred to non-negativity constrained portfolios and the other three are unconstrained portfolios. In addition, the last two portfolios, DOM and SHR, can be considered as special cases of the general portfolio combination rule:

$$w_{comb} = a_1 \frac{\iota}{N} + a_2 \hat{w}_{gmv} + a_3 \hat{w}_{ep} \tag{2.3.6}$$

with $a_1 + a_2 + a_3 = 1$. The DOM portfolio and the SHR portfolio represent the cases where $a_1 = 0$ and $a_3 = 0$ respectively, but they are not nested. Figure 2.5 shows the expected CE losses of these portfolio strategies. In the 5PF case, because the dimension is small relative to the sample size, the CE loss due to the estimation error is small. Thus, negative positions in portfolios can help. The unconstrained portfolios, SHR, DOM as well as GMVP have similar performance but outperform other three constrained portfolio strategies. In large dimensional cases where $N = 30$ and 48, the non-negativity constraint reduces the estimation error in the sample covariance matrix, and thus the constrained GMVP performs the best for investors with large risk aversion. However, in the 25PF case, since the non-negativity constraint leads to larger theoretical loss , the three unconstrained portfolios perform similarly to the constrained GMVP for high degrees of risk aversion, while the SHR performs much better than others for low degrees of risk aversion. Increasing the sample size significantly improves the performance of unconstrained portfolios, but the improvement of constrained portfolios is limited. In the case of $T = 240$, unconstrained portfolios perform better than constrained portfolios and the SHR portfolio performs the best in all four data sets.

Different from the DOM portfolio, the SHR portfolio depends on the nuisance parameter Δ_{SSR} which is unknown in practice. Under the normality assumption, Liu & Pohlmeier (2013) propose an unbiased estimator of Δ_{SSR} based on the sample mean and the sample covariance matrix. But it is shown that Δ_{SSR} depends heavily on the correlations: as the correlation level increases, Δ_{SSR} increases dramatically (see Liu & Pohlmeier (2013)). Since the correlations in sample covariance matrix seem to be overestimated in the small sample case, the estimated Δ_{SSR} based on the sample covariance matrix could contains large estimation errors. Therefore, instead of using the sample covariance, we adopt a diagonal matrix with diagonal elements equal to sample asset variances to calculate $\hat{\Delta}_{SSR}$. Figure 2.6 compares the SHR portfolio based on the estimated $\hat{\Delta}_{SSR}$ with other competitive portfolios. It can be observed that the estimation errors in $\hat{\Delta}_{SSR}$ reduce the performance of SHR port-

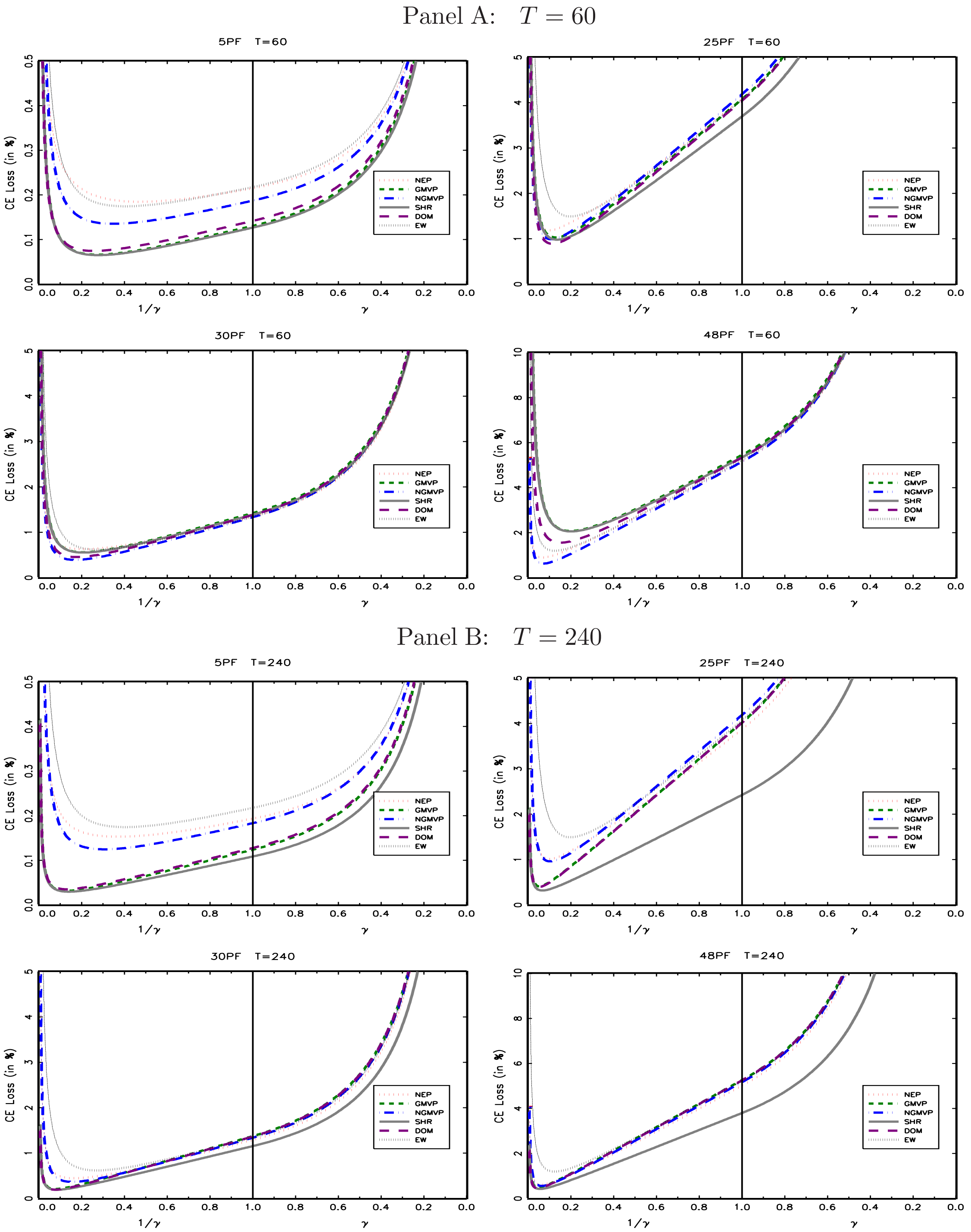


Figure 2.5: Expected CE loss of different portfolio strategies:
1. Red dotted line: non-negativity constrained efficient portfolio (NEP);
2. Green short dashed line: global minimum variance portfolio (GMVP);
3. Blue dots and dashes: non-negativity constrained global minimum variance portfolio (NGMVP);
4. Grey solid line: shrinkage portfolio of Liu & Pohlmeier (2013) based on true Δ_{SSR} (SHR);
5. Magenta dashed line: dominating estimator of GMVP (DOM);
6. Black closely spaced dots:equally weighted portfolio (EW).

folio. In the small sample case, it becomes even inferior for 25PF and 30PF data sets. But if $T = 240$, the estimated SHR performs similarly to the theoretical one and outperforms other considered portfolios.

2.4 Empirical Results

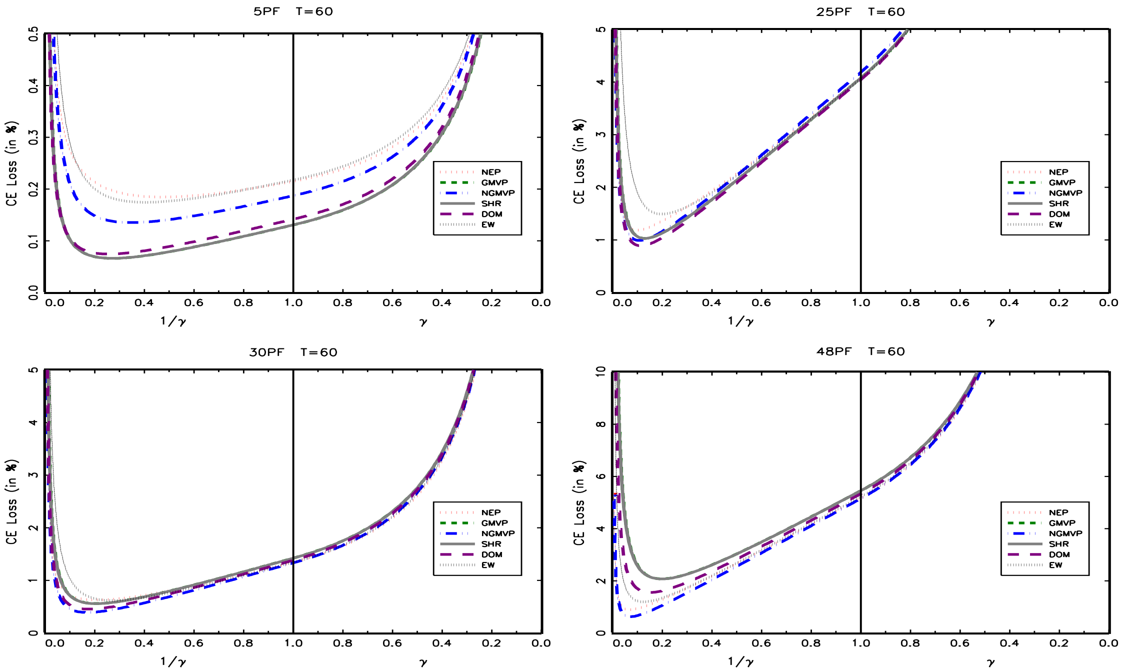
In this section we use the 25PF data set to compare the six competitive portfolios described in the previous section. There are two reasons for focusing on the 25PF rather than the other data sets. First, as indicated in the previous section, the non-negativity constraint causes a larger theoretical loss in the 25PF case than other cases. This allows us to better investigate the trade-off between the theoretical and empirical losses of constrained portfolios. Second, different from industry portfolios constructed by firm's SIC codes, the 25PF is formed from economic properties of firms and updated every six months. Therefore, comparing to industry portfolios, the 25PF data set is less sensitive to structural breaks.

To estimate portfolio weights, we use a "rolling window" approach with the length of estimation window equal to M. In each month t, starting from $t = M$, the mean and the covariance matrix are estimated by their sample counterparts from asset returns in the previous M months. We then use the estimated mean and covariance matrix to calculate the portfolio weights from t to $t+1$ and record the portfolio return in month $t+1$. This process is continued by adding the return for the next period and dropping the return for the earliest period. At the end, we obtain a series of $T-M$ monthly out-of-sample returns. Let $\bar{r}_p$ and $\hat{\sigma}_p^2$ denote the sample estimates of the portfolio mean and variance based on recorded out-of-sample returns, respectively. Then the out-of-sample portfolio CE can be obtained by:

$$CE_o = \bar{r}_p - \frac{\gamma}{2}\hat{\sigma}_p^2 \tag{2.4.1}$$

Figure 2.7 graphs the out-of-sample portfolio CEs. We see that the result is not exactly the same as but similar to the results of simulation study from the previous section(see Figure 2.6 Panel B). In general, the constrained GMVP, which is the most conservative portfolio from both theoretical and empirical perspective, performs the worst. Taking the mean into account, the constrained efficient portfolio performs similar to the constrained GMVP for investors with high risk aversions but much better for less risk averse investors. The three unconstrained portfolios perform almost the same for all risk aversion levels. They perform worse than the equally weighted and the unconstrained efficient portfolio in the case of $M = 60$ for investors with the same risk aversion level, but provide similar performance for highly risk

Panel A: $T = 60$



Panel B: $T = 240$

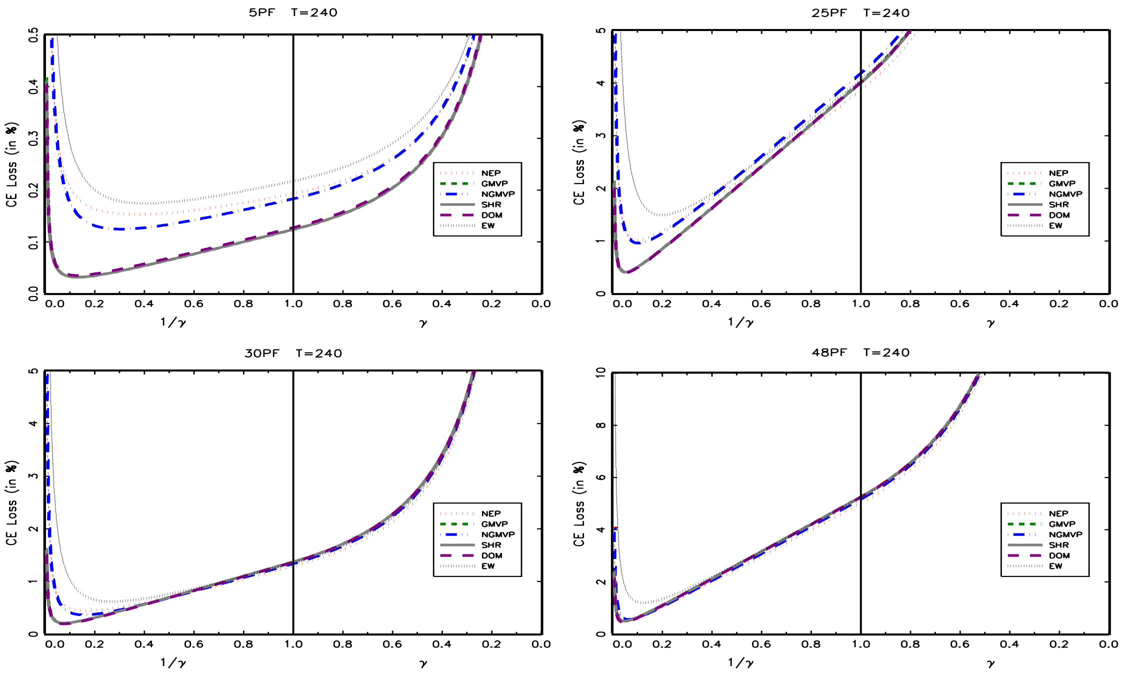


Figure 2.6: Expected CE loss of different portfolio strategies:
1. Red dotted line: non-negativity constrained efficient portfolio (NEP);
2. Green short dashed line: global minimum variance portfolio (GMVP);
3. Blue dots and dashes: non-negativity constrained global minimum variance portfolio (NGMVP);
4. Grey solid line: shrinkage portfolio of Liu & Pohlmeier (2013) based on estimated Δ_{SSR} (SHR);
5. Magenta dashed line: dominating estimator of GMVP (DOM);
6. Black closely spaced dots: equally weighted portfolio (EW).

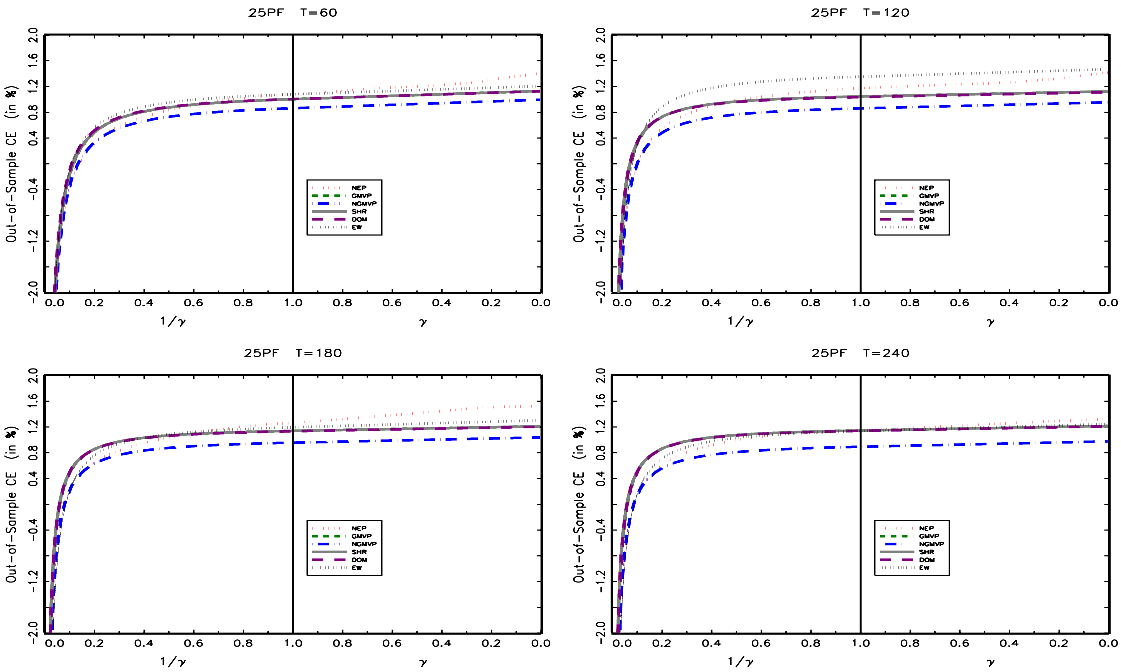


Figure 2.7: Out-of-sample CE of different portfolios:
1. Red dotted line: non-negativity constrained efficient portfolio (NEP);
2. Green short dashed line: global minimum variance portfolio (GMVP);
3. Blue dots and dashes: non-negativity constrained global minimum variance portfolio (NGMVP);
4. Grey solid line: shrinkage portfolio of Liu & Pohlmeier (2013) based on estimated Δ_{SSR} (SHR);
5. Magenta dashed line: dominating estimator of GMVP (DOM);
6. Black closely spaced dots: equally weighted portfolio (EW).
Size of estimation window: $T = 60$ (left upper panel), $T = 120$ (right upper panel), $T = 180$ (left lower panel),$T = 240$ (right lower panel).

averse investors. In cases of $M = 180$ and $M = 240$, they outperform the three unconstrained portfolios for investors with high risk aversions. Because we consider the 25PF case where the non-negativity constraint is severely wrong in population for all risk aversion levels, restricting portfolio weights to be positive could lead to considerable theoretical loss which might dominates the empirical loss caused by the estimation error in the covariance matrix. Therefore, for highly risk averse investors, since the estimation risk in mean is less important, we should relax the constraint and include negative position into the portfolio to diversify the risk. For less risk averse investors, the non-negativity constraint can effectively restrict the huge estimation risk in the mean and the covariance matrix and provide better performance.

2.5 Conclusion

Imposing a non-negativity constraint on the portfolio selection problem leads to theoretically inferior portfolios, but in the mean time it also limits the estimation risk and improves the practical performance of portfolios. This paper provides insights into the trade-off between the theoretical and empirical losses of the non-negativity constrained portfolio and intensively explains why and when the non-negativity constraint helps.

We generalize the result of Jagannathan & Ma (2003) to the efficient portfolio case where portfolio weights satisfy the adding-up constraint, and show theoretically that imposing the non-negativity constraint on the efficient portfolio is equivalent to using a modified covariance matrix depending on the mean and the risk preference of investors in portfolio optimization. As the investor's risk aversion decreases, the modified covariance matrix approaches to a singular matrix formed by the product of the mean vector and its transpose. Therefore, for low degrees of risk aversions, restricting portfolio weights lead to considerable theoretical losses which cannot be reduced by increasing the estimation precision.

We also conduct a simulation study to examine the trade-off between the theoretical and empirical losses of the non-negativity constrained portfolio. We find that, for investors with very high or very low risk aversions, the theoretical loss is dominating, while for investors with moderate risk aversions, the empirical loss becomes dominating. To explain this evidence, we calculate the probability that the active assets, which have non-zero portfolio weights, are correctly identified. We find that it is almost impossible to obtain the correct active asset set for a large range of moderate risk aversions which are frequently adopted in the literature.

Based on simulated and empirical data, we compare the performance of six competitive portfolio strategies including three constrained and three unconstrained portfolios in terms of their Certainty Equivalents. We select several financial data sets to obtain realistic inputs for the simulation. We find that, in the small dimensional case, unconstrained portfolios perform better than constrained portfolios, while in large dimensional case all portfolios perform similarly except the optimal shrinkage portfolio strategy proposed by Liu & Pohlmeier (2013). In general, the optimal shrinkage portfolio performs better than others, especially for investors with relatively low risk aversions.

In our empirical study, we select the French's 25 portfolio data set which is formed based on size and book-to-market ratio. The constrained portfolios constructed from this data set contains less active assets with non-zero weights which suggests that the non-negativity constraint is severely wrong in population relative to other data sets analyzed in the simulation study. We find that the constrained global minimum variance portfolio performs the worst in general. For investors with relative high risk aversions, relaxing the constraint leads to better portfolio performance, while for less risk averse investors, the non-negativity constraint reduces the estimation risk which could be huge in the small sample case, and largely improves the portfolio performance.

Bibliography

Antoine, B. (2012), 'Portfolio selection with estimation risk: A test-based approach', *Journal of Financial Econometrics* **10**(1), 164–197.

Brodie, J., Daubechies, I., Molc, C. D., Giannone, D. & Loris, I. (2009), 'Sparse and stable markowitz portfolios', *PNAS* **106**(30), 12267–12272.

DeMiguel, V., Garlappi, L., Nogales, F. J. & Uppal, R. (2009), 'A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms', *Management science* **55**(5), 798–812.

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

Fan, J., Zhang, J. & Yu, K. (2012), 'Vast portfolio selection with gross-exposure constraints', *Journal of the American Statistical Association* **107**(498), 592–606.

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Green, R. C. & Hollifield, B. (1992), 'When will mean-variance efficient portfolios be well diversified?', *Journal of Finance* **47**(5), 1785–1809.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

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

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

Liu, H. & Pohlmeier, W. (2013), Risk preferences and estimation risk in portfolio choice. Working Paper.

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

Markowitz, H. M. (1987), *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell.

2.6 Appendix

2.6.1 Appendix A

Proof 2.6.1 (*Lemma 2.2.1*). *Firstly, we show that the first order condition of the unconstrained optimization problem* (2.2.8) *is satisfied if the Kuhn-Tucker conditions* (2.2.6)-(2.2.7) *hold. Suppose that* (w, ξ, λ_1) *is a solution of the constrained optimization problem* (2.2.4) *and* $\mu_p = \mu' w$ *is the corresponding portfolio mean, we have*

$$\begin{aligned}
\tilde{\Sigma} w &= \left(\Sigma - (\iota\xi' + \xi\iota') + k_1 \cdot \mu\mu' + k_2 \cdot \iota\iota'\right) w \\
&= \Sigma w - (\iota\xi' + \xi\iota')w + \frac{1}{\lambda_1\gamma^2 + \mu_p\gamma} \cdot \mu\mu' w + 2\frac{\mu_p}{\gamma} \cdot \iota\iota' w \\
&= \Sigma w - \xi + \frac{\mu_p}{\lambda_1\gamma^2 + \mu_p\gamma} \cdot \mu + 2\frac{\mu_p}{\gamma} \cdot \iota \\
&= \lambda_1 \iota + \frac{1}{\gamma}\mu + \frac{\mu_p}{\lambda_1\gamma^2 + \mu_p\gamma} \cdot \mu + 2\frac{\mu_p}{\gamma} \cdot \iota \\
&= \left(\lambda_1 + 2\frac{\mu_p}{\gamma}\right)\iota + \frac{1}{\gamma}\left(1 + \frac{\mu_p}{\lambda_1\gamma + \mu_p}\right)\mu \\
&= \left(\lambda_1 + 2\frac{\mu_p}{\gamma}\right)\iota + \frac{1}{\gamma}\tilde{\mu}
\end{aligned}$$

The third equality holds because $\iota' w = 1$ *and* $\mu' w = \mu_p$*. The forth equality follows from the condition* (2.2.6)*. Therefore the fact that*

$$\tilde{\Sigma} w - \frac{1}{\gamma}\tilde{\mu} - \left(\lambda_1 + 2\frac{\mu_p}{\gamma}\right)\iota = 0$$

shows that w *is the solution of the unconstrained mean-variance optimization problem based* $\tilde{\Sigma}$ *and* $\tilde{\mu}$*, if the matrix* $\tilde{\Sigma}$ *is symmetric and positively semi-definite.*
the matrix $\tilde{\Sigma}$ *is obviously symmetric. Now we show that it is positively semi-definite. Let* x *be an arbitrary vector. Then:*

$$\begin{aligned}
x'\tilde{\Sigma} x &= x\left(\Sigma - (\iota\xi' + \xi\iota') + k_1 \cdot \mu\mu' + k_2 \cdot \iota\iota'\right) x \\
&= x'\Sigma x - x'(\iota\xi' + \xi\iota')x + k_1 \cdot x'\mu\mu' x + k_2 \cdot x'\iota\iota' x \\
&= x'\Sigma x - 2(x'\iota)(x'\xi) + k_1 \cdot (x'\mu)^2 + k_2 \cdot (x'\iota)^2
\end{aligned}$$

By the Kuhn-Tucker conditions (2.2.6)-(2.2.7)*, we have*

$$x'\xi = x'\Sigma w - \lambda_1(x'\iota) - \frac{1}{\gamma}(x'\mu) \tag{2.6.1}$$

$$w'\Sigma w = \lambda_1 + \frac{1}{\gamma}\mu_p \geq 0 \tag{2.6.2}$$

Therefore,

$$\begin{aligned}
x'\tilde{\Sigma}x &= x'\Sigma x - 2(x'\iota)\left(x'\Sigma w - \lambda_1(x'\iota) - \frac{1}{\gamma}(x'\mu)\right) + k_1\cdot(x'\mu)^2 + k_2\cdot(x'\iota)^2 \\
&= \left[x'\Sigma x - 2(x'\iota)x'\Sigma w + (x'\iota)^2 w'\Sigma w\right] - (x'\iota)^2 w'\Sigma w \\
&\quad +2\lambda_1(x'\iota)^2 + 2\frac{1}{\gamma}(x'\iota)(x'\mu) + k_1\cdot(x'\mu)^2 + k_2\cdot(x'\iota)^2 \\
&= \left[x'\Sigma x - 2(x'\iota)x'\Sigma w + (x'\iota)^2 w'\Sigma w\right] - (x'\iota)^2\left(\lambda_1 + \frac{1}{\gamma}\mu_p\right) \\
&\quad +2\lambda_1(x'\iota)^2 + 2\frac{1}{\gamma}(x'\iota)(x'\mu) + k_1\cdot(x'\mu)^2 + k_2\cdot(x'\iota)^2 \\
&= \left[x'\Sigma x - 2(x'\iota)x'\Sigma w + (x'\iota)^2 w'\Sigma w\right] + \left(\lambda_1 + \frac{\mu_p}{\gamma}\right)(x'\iota)^2 + \frac{2}{\gamma}(x'\iota)(x'\mu) + \frac{(x'\mu)^2}{\lambda_1\gamma^2 + \mu_p\gamma} \\
&= \left[x'\Sigma x - 2(x'\iota)x'\Sigma w + (x'\iota)^2 w'\Sigma w\right] + \left(\lambda_1 + \frac{\mu_p}{\gamma}\right)\left[(x'\iota) + \frac{1}{\lambda_1\gamma + \mu_p}(x'\mu)\right]^2
\end{aligned}$$

Because
small

$$|(x'\iota)x'\Sigma w| \le |(x'\iota)|\cdot(x'\Sigma x)^{\frac{1}{2}}\cdot(w'\Sigma w)^{\frac{1}{2}},$$

we have

$$\begin{aligned}
x'\tilde{\Sigma}x &= \left[x'\Sigma x - 2(x'\iota)x'\Sigma w + (x'\iota)^2 w'\Sigma w\right] + \left(\lambda_1 + \frac{\mu_p}{\gamma}\right)\left[(x'\iota) + \frac{1}{\lambda_1\gamma + \mu_p}(x'\mu)\right]^2 \\
&\ge \left[x'\Sigma x - 2\left|(x'\iota)x'\Sigma w\right| + (x'\iota)^2 w'\Sigma w\right] + \left(\lambda_1 + \frac{\mu_p}{\gamma}\right)\left[(x'\iota) + \frac{1}{\lambda_1\gamma + \mu_p}(x'\mu)\right]^2 \\
&= \left[(x'\Sigma x)^2 - \left|(x'\iota)\right|(w'\Sigma w)^{\frac{1}{2}}\right]^2 + \left(\lambda_1 + \frac{\mu_p}{\gamma}\right)\left[(x'\iota) + \frac{1}{\lambda_1\gamma + \mu_p}(x'\mu)\right]^2 \\
&\ge 0.
\end{aligned}$$

The last inequality follows from the fact that $\lambda_1 + \frac{1}{\gamma}\mu_p = w'\Sigma w \ge 0$. *Thus,* $\tilde{\Sigma}$ *is positive-semi definite.*

Proof 2.6.2 (*Proposition 2.2.1*)**.** *It is well known that the closed form solution of the unconstrained mean-variance optimization problem based on* $\tilde{\Sigma}$ *and* $\tilde{\mu}$ *is given by:*

$$w_{ep}(\tilde{\mu},\tilde{\Sigma}) = \frac{\tilde{\Sigma}^{-1}\iota}{(\iota'\tilde{\Sigma}^{-1}\iota)} + \frac{1}{\gamma}\cdot\left(\tilde{\Sigma}^{-1} - \frac{\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}}{\iota'\tilde{\Sigma}^{-1}\iota}\right)\cdot\tilde{\mu}, \tag{2.6.3}$$

It is easy to show that, for any constant $\tau \neq 0$,

$$w_{ep}(\tilde{\mu},\tilde{\Sigma}) = w_{ep}(\tau\tilde{\mu},\tau\tilde{\Sigma}). \tag{2.6.4}$$

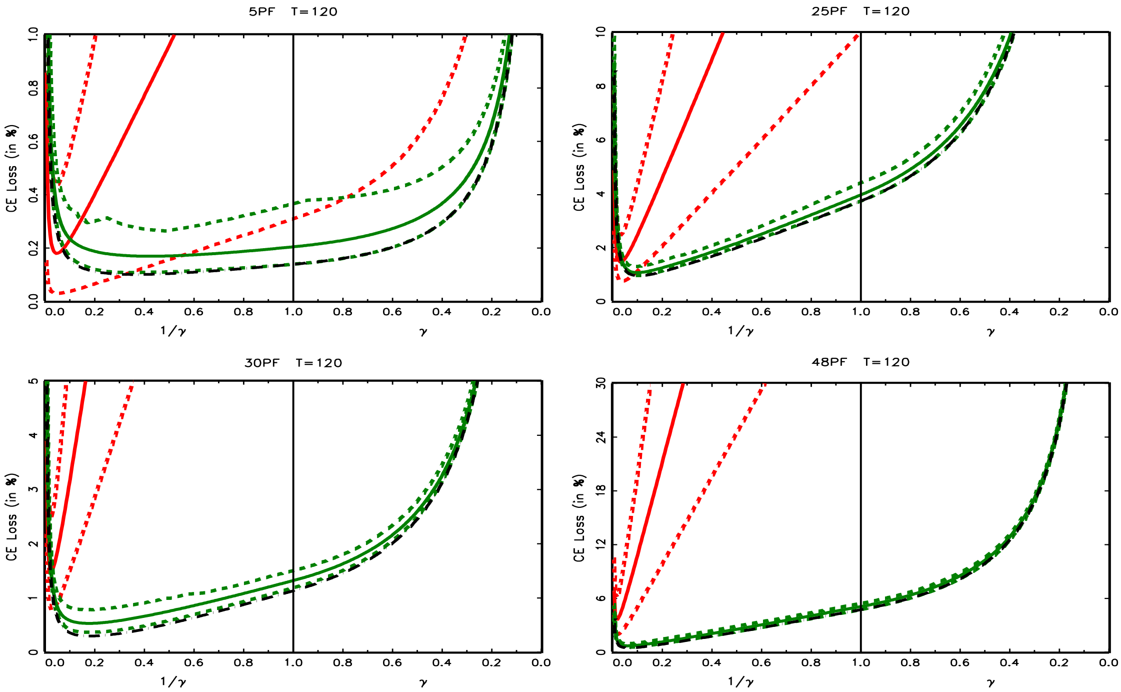
From (2.6.2)*, we have*

$$1 + k_1\cdot\gamma\cdot\mu_p = 1 + \frac{\mu_p}{\lambda_1\gamma + \mu_p} \ge 1 \tag{2.6.5}$$

Therefore, the result in proposition can be easily shown by some simple calculation.

2.6.2 Appendix B

Panel A : $T = 120$



Panel B : $T = 180$

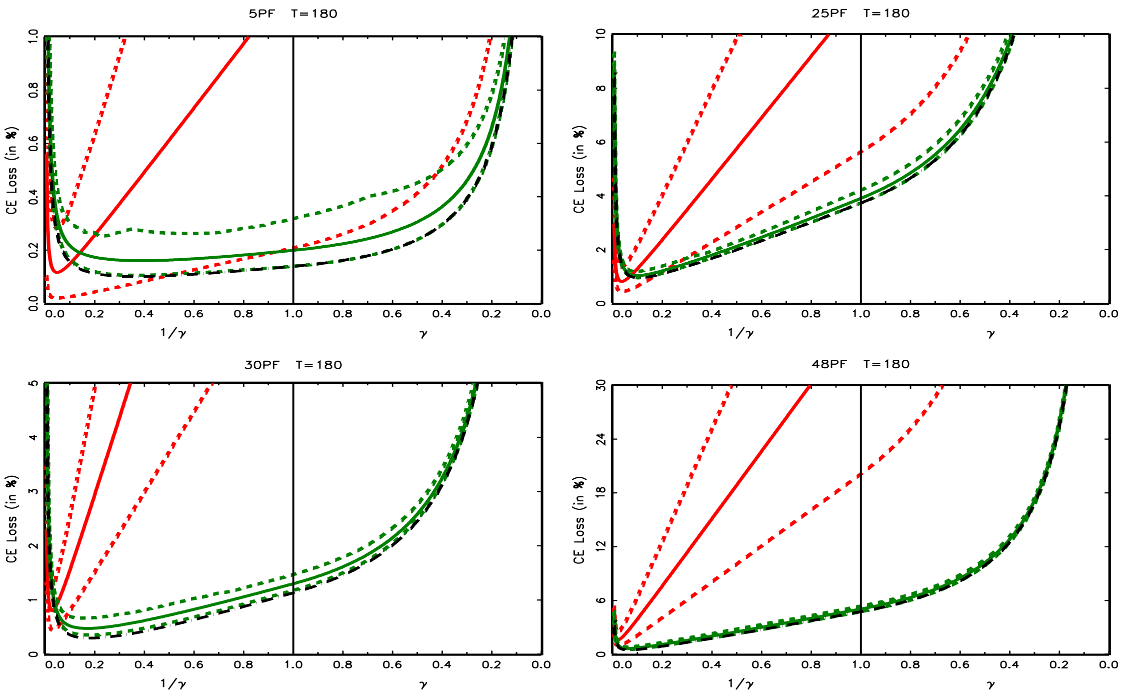
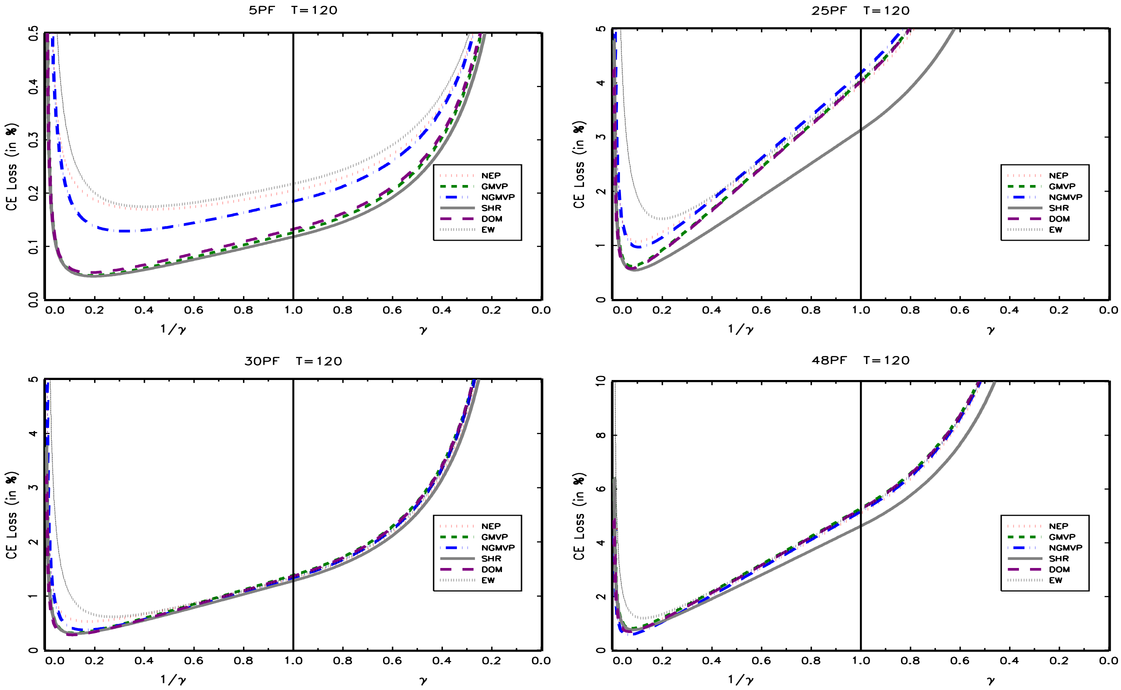


Figure 2.8: CE loss of the estimated unconstrained efficient portfolio and the estimated non-negativity constrained portfolio:
1. unconstrained portfolio: red solid line (mean), red dotted lines (90% confidence interval);
2. constrained portfolio: green solid line (mean), green dotted lines (90% confidence interval).

Panel A: $T = 120$



Panel B: $T = 180$

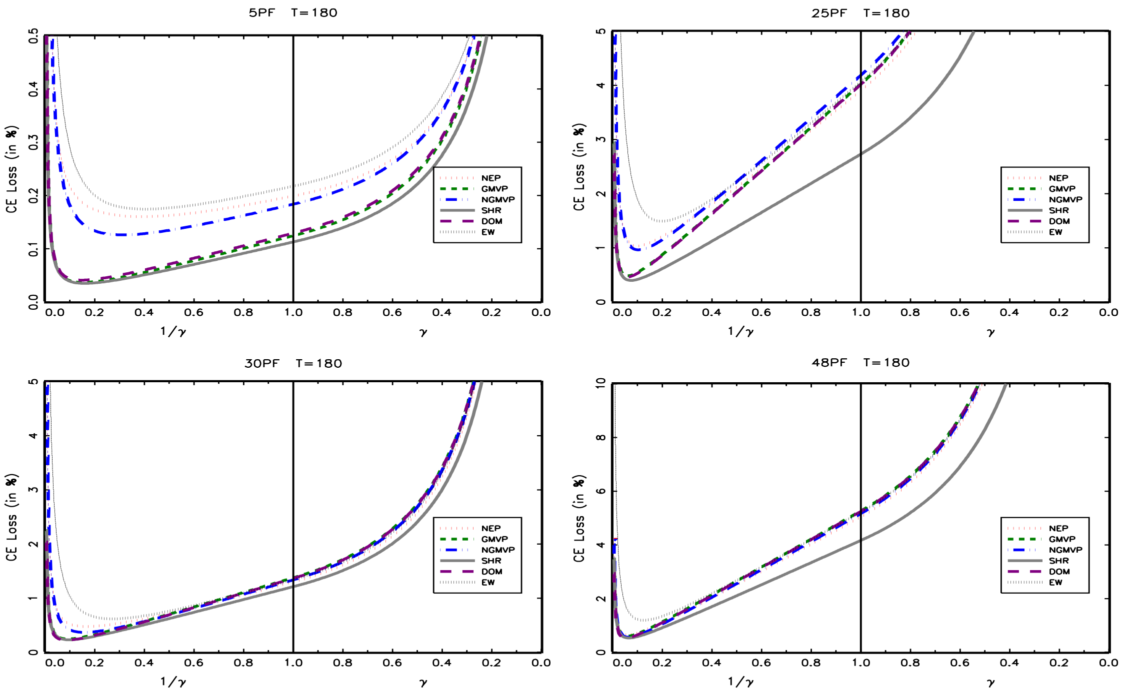


Figure 2.9: Expected CE loss of different portfolio strategies:
1. Red dotted line: non-negativity constrained efficient portfolio (NEP);
2. Green short dashed line: global minimum variance portfolio (GMVP);
3. Blue dots and dashes: non-negativity constrained global minimum variance portfolio (NGMVP);
4. Grey solid line: shrinkage portfolio of Liu & Pohlmeier (2013) based on true Δ_{SSR} (SHR);
5. Magenta dashed line: dominating estimator of GMVP (DOM);
6. Black closely spaced dots:equally weighted portfolio (EW).

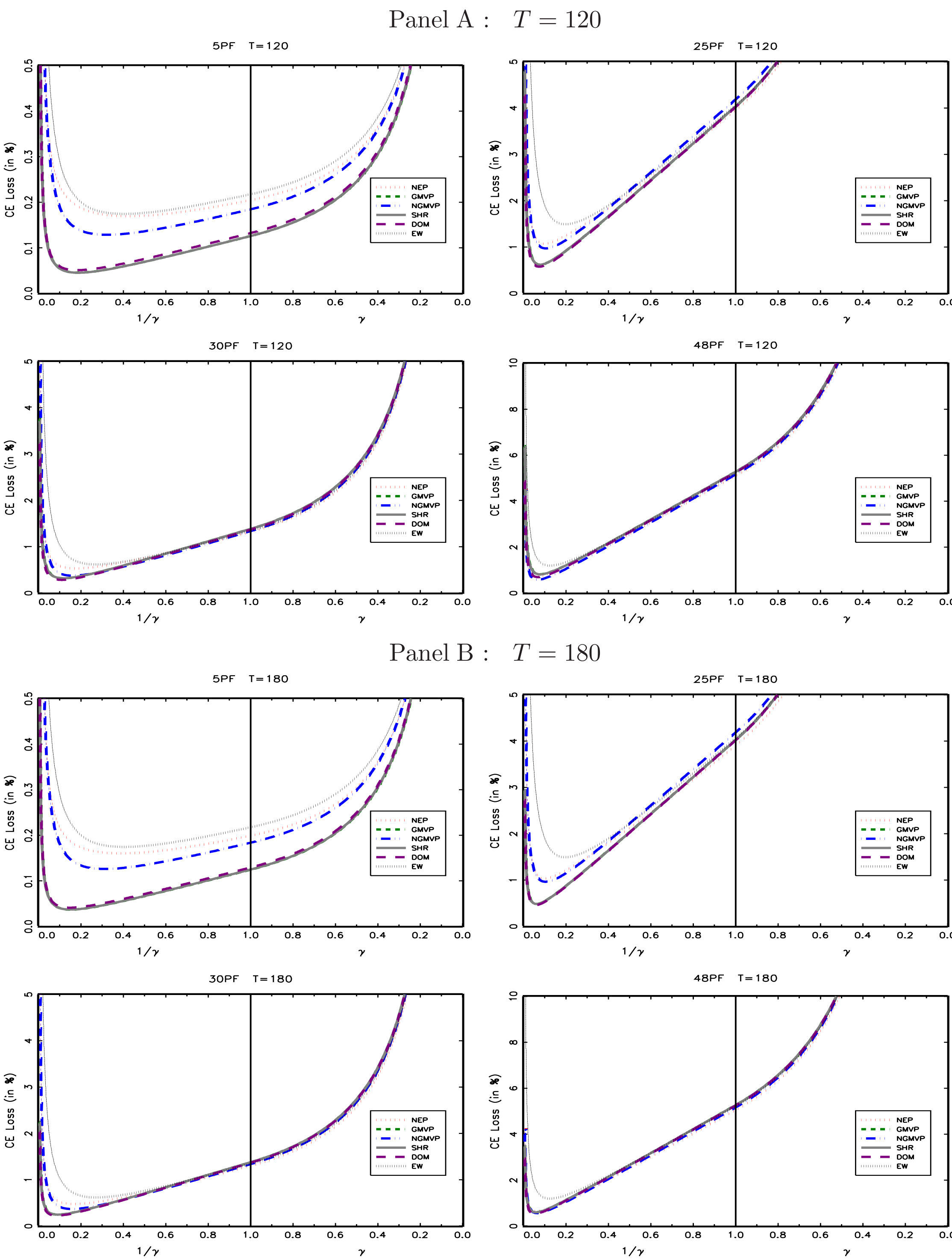


Figure 2.10: Expected CE loss of different portfolio strategies:
1. Red dotted line: non-negativity constrained efficient portfolio (NEP);
2. Green short dashed line: global minimum variance portfolio (GMVP);
3. Blue dots and dashes: non-negativity constrained global minimum variance portfolio (NGMVP);
4. Grey solid line: shrinkage portfolio of Liu & Pohlmeier (2013) based on estimated Δ_{SSR} (SHR);
5. Magenta dashed line: dominating estimator of GMVP (DOM);
6. Black closely spaced dots:equally weighted portfolio (EW).

Chapter 3

Portfolio Choice: Combining Pre- and Post-Break Information

3.1 Introduction

There is no doubt that portfolio selection under structural breaks is certainly the most realistic but simultaneously a much more difficult and complicated problem. Despite its crucial importance, technical difficulties in the structural break analysis heavily limit the ability to investigate the impact of structural breaks on the portfolio selection problem. In the empirical portfolio selection literature, most studies emphasize the small sample properties of the estimated portfolios and highlight that, if the sample is finite, the estimation error in input parameters could lead to extremely bad portfolio performance[1]. As indicated by Pesaran & Timmermann (2005), the presence of structural breaks is also the main reason why the small sample properties are of particular interest: even if the entire sample is very large, the occurrence of a structural break means that the post-break sample will often be quite small.

Instead of designing and estimating the post-break model, Pesaran & Timmermann (2007) focus on the selection of the sampling window and propose a new research avenue to deal with structural breaks. They consider the forecasting problem based linear regression model and conclude that the inclusion of some pre-break data in parameter estimation leads to improved forecasting performance. Since the weights of mean-variance optimal portfolios can be computed from the slope coefficients of an OLS-regression (see Kempf & Memmel (2006), Britten-Jones (1999) and Brodie

[1]See for instance Okhrin & Schmid (2006), Kan & Zhou (2007), DeMiguel, Garlappi & Uppal (2009), Frahm & Memmel (2010) and Liu & Pohlmeier (2013). Especially, Frahm & Wiechers (2011) define the ratio of the sample size over the dimension as the effective sample size and argue that the precision of the estimated portfolio depends heavily on this effective sample size.

et al. (2009)), it is natural to ask whether the pre-break sample of returns is informative for the post-break portfolio selection and can be used to improve the portfolio performance. It seems not trivial to answer this question from the finding of Pesaran & Timmermann (2007). Different from the regression model with strictly exogenous regressors considered by Pesaran & Timmermann (2007), in the regression representation of the mean-variance portfolio selection problem, the regressors are asset returns and therefore endogenous. This implies that, if a structural break in fact occurred, then one has to face changes not only in regression coefficients but also in the distribution of regressors which make the analysis much more complicated. In addition, portfolio selection problem has a totally different objective than the forecasting problem: the former aims at seeking the optimal trade-off between the return and risk of the investment to maximize the investor's economic gain, while the latter only focuses on reducing the forecasting error based on some given statistical evaluation measures.

This paper takes the structural break into account and analytically studies the performance of different mean-variance portfolio strategies with and without adding-up constraint. The certainty equivalent (CE) is used as a monetary measure to compare the performance of empirical portfolios estimated using exclusively post-break data with those incorporating pre-break information[2]. Since the bad performance of empirical portfolios is mainly caused by the large estimation error in means, we first consider the case where the pre-break information is used in mean estimation, while the covariance matrix is estimated solely by post-break data. It is shown that including pre-break data in estimation could lead to the portfolio weights that have lower variance at the cost of greater bias, and by trading off the bias and the variance of estimated portfolio weights, the portfolio performance measured by the CE can be substantially improved.

The situation is however different when focusing on the estimation of return covariance matrix. In the presence of structural breaks in the return distribution, no matter in the mean or in the covariance matrix, the distributional properties of the sample covariance matrix are intractable even under the normality assumption. To solve the problem, we propose a combination approach to reduce the estimation error in covariance matrix: given a particular mean estimate, the portfolios estimated

[2]Although both the certainty equivalent and the Sharpe ratio are widely used as a performance measure in literature, Engle & Colacito (2006) reveal the fact that "selecting the best covariance matrix estimator based on a Sharpe ratio criterion may be misleading". Frahm, Wickern & Wiechers (2012) argue that, if the investor's risk preferences are taken into account, it is appropriate to compare the portfolio performance based on the CE rather than on the Sharpe ratio.

from the post-break sample covariance matrix are combined with the one estimated from the pre-break sample covariance matrix. It is shown that, under some simplifying assumptions, the combined portfolio dominates the uncombined one in terms of its CE.

The structure of this paper is as follows. Section 3.2 describes necessary assumptions used in further analysis and some elementary results in literature are also reviewed. Section 3.3 analytically compares the portfolio incorporating pre-break information with the one estimated by post-break data. Since our analytical results for portfolio comparison depend on some unknown parameters associated with the population return distribution, Section 3.4 derives the unbiased estimator for these parameters under the normality assumption. Section 3.5 presents some calibration results for a few selected data sets to provide evidence for the empirical relevance of analytical findings. Section 3.6 concludes.

3.2 The Portfolio Choice Problem

3.2.1 Assumptions and Notation

Suppose there are N risky assets and one riskless asset. Let the $N \times 1$ random vector $\mathbf{R}_t = (R_{1t}, \ldots, R_{Nt})$ denote the returns of risky assets at time t and R_f denote the return of riskless asset. Define the excess return as $\mathbf{r}_t = R_t - R_f \iota$ where ι denotes an $N-$dimensional vector of ones.

In practice, the underlying distribution of excess returns is unknown and has to be estimated from historical data. Assume at the beginning of the investment period there are T observed excess returns, $\{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$, available. Based on such information set, the investor would like to construct a portfolio for period $T+1$. Similar to Okhrin & Schmid (2006), we make the following assumptions[3]:

Assumption 3.2.1.

A1) The excess returns are serially independent.

A2) There are at least 3 assets, i.e. $N \geq 3$.

A3) There is single structural break in the return distribution at $T_1 < T$ such that $r_t \overset{\text{iid}}{\sim} \mathrm{N}(\mu_1, \Sigma_1)$ for $t = 1, \ldots, T_1$ and $r_t \overset{\text{iid}}{\sim} \mathrm{N}(\mu_2, \Sigma_2)$ for $t = T_1 + 1, \ldots, T+1$.

[3]Similar assumptions are also made by other analytical studies, e.g. Kan & Zhou (2007), Frahm & Memmel (2010) and Liu & Pohlmeier (2013).

A4) The break point is known.

A5) The sizes of both subsamples are sufficiently large, more precisely $T_1 \geq N+4$ *and* $T_2 = T - T_1 \geq N+4$.

In this paper, we also make the normality assumption to obtain analytical results for the finite sample properties of the estimated portfolio weight. But different from the previous studies, we explicitly account for a possible structural break in the return distribution. We will see later that our results can be easily generalized to the multiple break case with unknown break points.

Under assumption that the break point T_1 is known and excess returns are i.i.d normally distributed before and after the structural break, the population means, μ_1 and μ_2, can be unbiasedly estimated by their sample counterparts from the pre-break and post-break samples respectively:

$$\begin{aligned} \bar{r}_{T_1} &= \frac{1}{T_1}\sum_{t=1}^{T_1} r_t \sim \mathrm{N}\left(\mu_1, \Sigma_1/T_1\right), \\ \bar{r}_{T_2} &= \frac{1}{T_2}\sum_{t=T_1+1}^{T} r_t \sim \mathrm{N}\left(\mu_2, \Sigma_2/T_2\right). \end{aligned}$$

Similarly, using pre-break or the post-break observations respectively, Σ_1 and Σ_2 can be estimated by:

$$\begin{aligned} S_{T_1} &= \frac{1}{T_1-1}\sum_{t=1}^{T_1}\left(r_t-\bar{r}_{T_1}\right)\left(r_t-\bar{r}_{T_1}\right)' \sim W_N(T_1-1,\Sigma_1)/T_1-1, \\ S_{T_2} &= \frac{1}{T_2-1}\sum_{t=T_1+1}^{T}\left(r_t-\bar{r}_{T_2}\right)\left(r_t-\bar{r}_{T_2}\right)' \sim W_N(T_2-1,\Sigma_2)/T_2-1, \end{aligned}$$

where $W_N(T_i-1, \Sigma_i)$ denotes the Wishart distribution with T_i-1 degrees of freedom and covariance matrix Σ_i, for $i=1,2$. If the structural break at time T_1 is ignored, we can obtain the following sample estimates of the mean and the covariance matrix based on the entire sample:

$$\begin{aligned} \bar{r}_T &= \frac{1}{T}\sum_{t=1}^{T} r_t \\ S_T &= \frac{1}{T-1}\sum_{t=1}^{T}\left(r_t-\bar{r}\right)\left(r_t-\bar{r}\right)'. \end{aligned}$$

Under the assumptions $A1-A3$, the sample mean $\bar{r}_T$ is still normally distributed with the mean and the covariance matrix given by:

$$\begin{aligned}\mu_T &= \mathrm{E}\left[\bar{r}_T\right] = \frac{T_1}{T}\mu_1 + \frac{T_2}{T}\mu_2 \\ \Sigma_T &= \mathrm{V}\left[\bar{r}_T\right] = \frac{T_1}{T^2}\Sigma_1 + \frac{T_2}{T^2}\Sigma_2,\end{aligned}$$

respectively. However, the distribution of the sample covariance based on the whole sample is intractable in the presence of structural break, which can be easily seen by the following decomposition:

$$\begin{aligned}S_T &= \frac{1}{T-1}\sum_{t=1}^{T}\left(r_t - \bar{r}\right)\left(r_t - \bar{r}\right)' \\ &= \frac{1}{T-1}\left[\sum_{t=1}^{T} r_t r_t' - \frac{1}{T}\left(\sum_{t=1}^{T} r_t\right)\left(\sum_{t=1}^{T} r_t\right)'\right] \\ &= \frac{1}{T-1}\left[(T_1-1)S_{T_1} + (T_2-1)S_{T_2} + \frac{T_1T_2}{T}\left(\bar{r}_{T_1} - \bar{r}_{T_2}\right)\left(\bar{r}_{T_1} - \bar{r}_{T_2}\right)'\right]\end{aligned}$$

Obviously, the sample covariance S_T is the sum of three random matrices, and no longer Wishart if $\Sigma_1 \neq \Sigma_2$ or $\mu_1 \neq \mu_2$. Therefore, in the presence of structural breaks, analytical inference about the portfolio estimated by using S_T is not available.

In this paper, we derive analytical results of using $\bar{r}_T$ instead of $\bar{r}_{T_2}$ to show the potential benefit of pre-break information. To reduce the estimation risk in the covariance matrix, we propose a linear combination of portfolios estimated from S_{T_1} and S_{T_2}, and show analytically that this combined portfolio outperforms the portfolio based on S_{T_2}. Although the analytical inference about S_T is not available, we perform a simulation study to show that the portfolio constructed from S_T performs very similarly to the combined portfolio and thus outperforms the portfolio constructed by the post-break sample covariance matrix.

3.2.2 The Mean-Variance Approach

According to the standard mean-variance framework, in the presence of riskless asset, the investor maximize his certainty equivalent (CE) return of the portfolio at time

T for the period $T+1$[4]:

$$\max_w CE(w) = \max_w \{\mu_2' w - \frac{\gamma}{2} w' \Sigma_2 w\}, \tag{3.2.1}$$

where the parameter $\gamma \in (0, \infty]$ reflects the investor's risk aversion level and w is the portfolio weight allocated to the risky assets. The solution of the optimization problem (3.2.1) is usually called as tangency portfolio (see e.g. Okhrin & Schmid (2006) or DeMiguel, Garlappi & Uppal (2009)) whose weight is given by:

$$w_{tp}^* = w_{tp}(\mu_2, \Sigma_2) = \frac{1}{\gamma} \cdot \Sigma_2^{-1} \cdot \mu_2, \tag{3.2.2}$$

with the CE given by (see e.g. Kan & Zhou (2007)):

$$CE(w_{tp}^*) = \frac{1}{2\gamma} \mu_2' \Sigma_2^{-1} \mu_2. \tag{3.2.3}$$

Obviously, the weight of the tangency portfolio does not necessarily add up to one. In some particular situations, the investor can only allocate wealth to the risky assets and need a portfolio satisfying the adding-up constraint. In this case the investor faces the following optimization problem:

$$\max_w CE(w) = \max_w \{\mu_2' w - \frac{\gamma}{2} w' \Sigma_2 w\}, \quad \text{s.t.} \quad \iota' w = 1 \tag{3.2.4}$$

The closed form solution of (3.2.4) is called efficient portfolio with the weight given by:

$$w_{ep}^* = w_{ep}(\mu_2, \Sigma_2) = w_{gmv,2} + \frac{1}{\gamma} \cdot A_2 \cdot \mu_2, \tag{3.2.5}$$

where

$$A_2 = \Sigma_2^{-1} - \frac{\Sigma_2^{-1} \iota \iota' \Sigma_2^{-1}}{\iota' \Sigma_2^{-1} \iota} \tag{3.2.6}$$

is a semi-positive definite matrix, and the weight vector $w_{gmv,2} = \Sigma_2^{-1}\iota/(\iota'\Sigma_2^{-1}\iota)$ refers to the global minimum variance portfolio (GMVP) as the solution of

$$\min_w w' \Sigma_2 w \quad s.t \quad \iota' w = 1.$$

The CE of the efficient portfolio is given by (see Liu & Pohlmeier (2013)):

$$CE(w_{tp}^*) = \frac{\mu_2' A_2 \mu_2}{2\gamma} + CE(w_{gmv,2}). \tag{3.2.7}$$

[4]Note that the excess return r_{T+1} is normally distributed with the mean μ_2 and the variance Σ_2

where $CE(w_{gmv,2})$ is the CE of the GMVP.

Usually, if there exists a structural break in the return distribution and the time of structural break is known, only the post-break observations will be used in the estimation. Replacing the unknown μ_2 and Σ_2 with their sample estimates gives the traditional plug-in estimators of optimal portfolio weights.

3.2.3 Expected CE Loss and Elementary Results

As argued by Kan & Zhou (2007) and Liu & Pohlmeier (2013), it is natural to use the certainty equivalent, which is the objective function of the mean-variance optimization problem, to compare the performance of different portfolios strategies. By definition, the tangency portfolio has the highest CE and dominates any other portfolio, i.e. $CE(w_{tp}^*) \geq CE(\hat{w}_{tp})$ where $\hat{w}_{tp}$ is a weight vector obtained by some other arbitrary portfolio selection rule. Thus, we can use the difference between the CE of w_{tp}^* and $\hat{w}_{tp}$ to define the loss function of using $\hat{w}_{tp}$:

$$\mathcal{L}(\hat{w}_{tp}, w_{tp}^*) \equiv CE(w_{tp}^*) - CE(\hat{w}_{tp}) \geq 0 \tag{3.2.8}$$

The expectation over the loss function defines the risk function[5] of using $\hat{w}_{tp}$:

$$\mathcal{R}(\hat{w}_{tp}|w_{tp}^*) \equiv \mathrm{E}\left[\mathcal{L}(\hat{w}_{tp}, w_{tp}^*)\right] = CE(w_{tp}^*) - \mathrm{E}\left[CE(\hat{w}_{tp})\right] \geq 0. \tag{3.2.9}$$

Frahm (2010) shows that:

$$\mathcal{L}(\hat{w}_{tp}, w_{tp}^*) = \frac{\gamma}{2}(w_{tp}^* - \hat{w}_{tp})'\Sigma_2(w_{tp}^* - \hat{w}_{tp}), \tag{3.2.10}$$

and the risk of $\hat{w}_{tp}$ is given by:

$$\mathcal{R}(\hat{w}_{tp}|w_{tp}^*) = \frac{\gamma}{2} tr\left(\Sigma_2 \cdot \left[\mathrm{Cov}[\hat{w}_{tp}] + Bias(\hat{w}_{tp})^2\right]\right), \tag{3.2.11}$$

where $Bias(\hat{w}_{tp})^2 = \left(\mathrm{E}\left[\hat{w}_{tp}\right] - w_{tp}^*\right)\left(\mathrm{E}\left[\hat{w}_{tp}\right] - w_{tp}^*\right)'$. If we assume that the true covariance matrix is known, the risk function of plug-in portfolio $w_{tp}(\gamma, \hat{\mu}_2, \Sigma_2)$ can be expressed as:

$$\mathcal{R}(\hat{w}_{tp}|w_{tp}^*) = \frac{1}{2\gamma} tr\left(\Sigma_2^{-1} \cdot \left[\mathrm{Cov}[\hat{\mu}_2] + Bias(\hat{\mu}_2)^2\right]\right), \tag{3.2.12}$$

where $Bias(\hat{\mu}_2)^2 = \left(\mathrm{E}\left[\hat{\mu}_2\right] - \mu_2\right)\left(\mathrm{E}\left[\hat{\mu}_2\right] - \mu_2\right)'$.
Obviously, for the same investor with a given risk aversion level, the efficient portfolio

[5] Note that the loss and risk functions are defined for the given risk aversion level γ.

is also dominated by the tangency portfolio but dominates any other portfolios satisfying the adding-up constraints, i.e. $CE(w_{ep}^*) \geq CE(\hat{w}_{ep})$, where $\hat{w}_{ep}$ is the weight vector of an arbitrary portfolio with $\iota'\hat{w}_{ep} = 1$. Based on the same comparison approach, the loss and risk function of using $\hat{w}_{ep}$ can be defined as:

$$\mathcal{L}(\hat{w}_{ep}, w_{ep}^*) \equiv CE(w_{ep}^*) - CE(\hat{w}_{ep}) \geq 0 \tag{3.2.13}$$

and

$$\mathcal{R}(\hat{w}_{ep}|w_{ep}^*) \equiv \mathrm{E}\left[\mathcal{L}(\hat{w}_{ep}, w_{ep}^*)\right] = CE(w_{ep}^*) - \mathrm{E}\left[CE(\hat{w}_{ep})\right] > 0, \tag{3.2.14}$$

respectively. Similar to Frahm (2010), Liu & Pohlmeier (2013) show that:

$$\mathcal{L}(\hat{w}_{ep}, w_{ep}^*) = CE(w_{ep}^*) - CE(\hat{w}_{ep}) = \frac{\gamma}{2}(w_{ep}^* - \hat{w}_{ep})'\Sigma_2(w_{ep}^* - \hat{w}_{ep}),$$

and

$$\mathcal{R}(\hat{w}_{ep}|w_{ep}^*) = CE(w_{ep}^*) - \mathrm{E}\left[CE(\hat{w}_{ep})\right] = \frac{\gamma}{2} tr\left(\Sigma_2 \cdot \left[\mathrm{Cov}[\hat{w}_{ep}] + Bias(\hat{w}_{ep})^2\right]\right),$$

where $Bias(\hat{w})^{2=} \left(\mathrm{E}\left[\hat{w}\right] - w_{ep}^*\right)\left(\mathrm{E}\left[\hat{w}\right] - w_{ep}^*\right)'$. If we assume that the true covariance matrix is known, the risk function of using the portfolio strategy $w_{ep}(\hat{\mu}_2, \Sigma_2)$ can be expressed as:

$$\mathcal{R}(\hat{w}_{ep}|w_{ep}^*) = \frac{1}{2\gamma} tr\left(A_2 \cdot \left[\mathrm{Cov}\,\hat{\mu}_2 + Bias(\hat{\mu}_2)^2\right]\right), \tag{3.2.15}$$

where $Bias(\hat{\mu}_2)^2 = \left(\mathrm{E}\left[\hat{\mu}_2\right] - \mu_2\right)\left(\mathrm{E}\left[\hat{\mu}_2\right] - \mu_2\right)'$.

3.3 Estimation with Pre-Break Data

3.3.1 Mean Estimation

3.3.1.1 True Covariance Matrix

To compare the portfolios using extra information from pre-break window with the post-break portfolio strategy, we first consider the most simple case where the true population covariance matrix is known. In this case, only the mean is estimated from the sample by using either $\bar{r}_T$ or $\bar{r}_{T_2}$, i.e. either simply ignoring the structural break or using exclusively post-break observations to estimate the mean. Fixing the value of the covariance matrix simplifies the analysis and helps us to better understand the trade-off between the variance and the bias of estimated portfolios in

the presence of a structural break. Although the assumption of the true population covariance matrix Σ_2 being known is unrealistic, it is still commonly made by many studies in the literature due to the fact that the mean and the covariance matrix are usually estimated separately. In addition, as pointed out by Merton (1980), while it requires a long history of returns to produce a precise estimate of the mean, a more precise estimate of the variance can be obtained by increasing the sampling frequency[6]. Therefore, it seems to be more reasonable to include pre-break observations in mean estimation than in estimation of the covariance matrix.
The normality assumption allows us to derive an exact comparison of the performance of the portfolios based on either $\bar{r}_T$ or $\bar{r}_{T_2}$:

Proposition 3.3.1.
Assume that the true covariance matrix Σ_2 is known. Under the assumptions $A1-A5$, the expected CE difference between two estimated portfolios using $\bar{r}_T$ and $\bar{r}_{T_2}$ is:

$$\begin{aligned}\Delta_{tp}\left(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2\right) &= \mathrm{E}\left[CE(w_{tp}(\bar{r}_T, \Sigma_2)) - CE(w_{tp}(\bar{r}_{T_2}, \Sigma_2))\right] \\ &= \mathcal{R}(w_{tp}(\bar{r}_{T_2}, \Sigma_2)|w^*_{tp}) - \mathcal{R}(w_{tp}(\bar{r}_T, \Sigma_2)|w^*_{tp}) \\ &= \frac{T_1^2}{2\gamma T^2}\left[N\left(\frac{1}{T_1} + \frac{1}{T_2}\right) - \eta_{tp} - \frac{1}{T_1}\delta_{tp}\right] \qquad (3.3.1)\end{aligned}$$

where

$$\eta_{tp} = (\mu_1 - \mu_2)' \, \Sigma_2^{-1} (\mu_1 - \mu_2) \quad \text{and} \quad \delta_{tp} = tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right).$$

Proof 3.3.1. *See Appendix.*

The quantity η_{tp} represents the "bias in mean" caused by the structural break. It is clear that, in the case of large N but small T_2, using $\bar{r}_T$ can significantly reduce the risk of empirical portfolio when the magnitude of the break is moderate. As T_2 increases, using post-break observations provides more and more precise estimate of the weight, and therefore, the benefit of using $\bar{r}_T$ decreases. As T_1 increases, the variance of the estimated mean decreases, but the bias caused by the structural break increases. Thus, the benefit of using pre-break observations is not monotonic in T_1 and there is an optimal T_1 which maximizes the CE of the estimated portfolio. However, this optimal sampling window depends not only on the size of the post-break sample, but also on the population quantities η_{tp} and δ_{tp} which have to be estimated. The estimation method of these population quantities is presented later

[6]See e.g., Best & Grauer (1991), Jorion (1986) and Frahm (2010). Antoine (2012) argues that "ignoring the estimation risk of the variance appears to be a reasonable simplifying assumption as long as the number of risky assets relative to the sample size is kept small, typically 1/6 in the above comparative study".

in Section 3.4 of this paper.

In addition, it can be seen that the CE difference is proportional to the inverse of risk aversion parameter which implies that the ranking of two tangency portfolios based on $\bar{r}_T$ and $\bar{r}_{T_2}$ is independent with the risk aversion level.

Similar results can be obtained for the efficient portfolio:

Proposition 3.3.2.
Assume that the true covariance matrix Σ_2 is known. Under the assumptions $A1-A5$, the expected CE difference between two estimated portfolios using $\bar{r}_T$ and $\bar{r}_{T_2}$ is:

$$\begin{aligned}\Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2\right) &= \mathrm{E}\left[CE(w_{ep}(\bar{r}_T, \Sigma_2)) - CE(w_{ep}(\bar{r}_{T_2}, \Sigma_2))\right] \\ &= \mathcal{R}(w_{ep}(\bar{r}_{T_2}, \Sigma_2)|w_{ep}^*) - \mathcal{R}(w_{ep}(\bar{r}_T, \Sigma_2)|w_{ep}^*) \\ &= \frac{T_1^2}{2\gamma T^2}\left[(N-1)\left(\frac{1}{T_1}+\frac{1}{T_2}\right) - \eta_{ep} - \frac{1}{T_1}\delta_{ep}\right] \qquad (3.3.2)\end{aligned}$$

where

$$\eta_{ep} = (\mu_1-\mu_2)' A_2 (\mu_1-\mu_2) \quad \text{and} \quad \delta_{ep} = tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right).$$

Proof 3.3.2. *See Appendix.*

For the efficient portfolio, biases in the mean and in the covariance matrix are weighted by the matrix A_2 instead of Σ_2^{-1}. Because $A_2 \cdot \Sigma_2 \cdot A_2 = A_2$, Σ_2^{-1} is the generalized inverse of the matrix A_2. Thus, A_2 plays a similar role as Σ_2^{-1} but it has one rank less than Σ_2^{-1} because of the adding-up constraint. When the portfolio CE using post-break sample is compared to the portfolio CE using both pre- and post-break sample, one can expect that the comparison result for efficient portfolios should be consistent with the result for tangency portfolios.

3.3.1.2 Sample Covariance Matrix

In practice, the covariance matrix is also unknown and has to be estimated from the sample. As argued in the previous section, analytical inferences about the distribution of S_T is not available in the presence of structural breaks in the return distribution. Therefore, to obtain an analytical comparison of the portfolio performance, we adopt the post-break sample estimate of the covariance matrix S_{T_2} and derive the following result:

Proposition 3.3.3.
Let the covariance matrix Σ_2 be estimated by S_{T_2}. Under the assumptions $A1-A5$,

the expected CE difference between the tangency portfolio estimated by using $\bar{r}_T$ and the post-break tangency portfolio based on $\bar{r}_{T_2}$ is:

$$\begin{aligned}
& \Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|S_{T_2}) \\
= \; & \mathrm{E}\left[CE(w_{ep}(\bar{r}_T, S_{T_2})) - CE(w_{tp}(\bar{r}_{T_2}, S_{T_2}))\right] \\
= \; & \left(\frac{T}{T_1}\alpha_1 - \frac{T_2}{T_1}\alpha_2\right) \cdot \Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2) + \frac{1}{2\gamma}\frac{T_1}{T}\left(\alpha_2 - \alpha_1\right)\Delta_{tp}^{\mu} \qquad (3.3.3)
\end{aligned}$$

where

$$\alpha_1 = \frac{T_2 - 1}{T_2 - N - 2} \qquad (3.3.4)$$

$$\alpha_2 = \frac{(T_2 - 1)^2(T_2 - 2)}{(T_2 - N - 1)(T_2 - N - 2)(T_2 - N - 4)} \qquad (3.3.5)$$

$$\Delta_{tp}^{\mu} = \frac{N}{T_1} + \frac{N}{T_2} - (\theta_{tp,1} - \theta_{tp,2}) - \frac{1}{T_1}\delta_{tp} \qquad (3.3.6)$$

with

$$\delta_{tp} = tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right) \quad \text{and} \quad \theta_{tp,i} = \mu_i'\Sigma_2^{-1}\mu_i \text{ for } i = 1, 2. \qquad (3.3.7)$$

Proof 3.3.3. *See Appendix.*

As shown in the Appendix, α_1 and α_2 are determined by the first two moments of a standardized inverse Wishart distribution, respectively[7]. It is easy to see that $\alpha_2 > \alpha_1^2 > \alpha_1$. Kan & Zhou (2007) show that the estimation risks in the sample mean and the sample covariance matrix are not additive but interact with each other. This interaction effect is reflected by α_2 and could be very large in the large dimensional case. Table 3.1 gives some values of α_1 and α_2 with respect to different dimensions and sample sizes.

[7] This analytical result can be easily generalized to other estimates of covariances matrix if they are Wishart-distributed.

Table 3.1: Values of α_1 and α_2 for different T (number of observations) and N (number of assets).

$T_2 \backslash N$	5	10	15	20	25	30
			Panel A: *Values of* α_1			
36	1.21	1.46	1.84	2.50	3.89	8.75
60	1.11	1.23	1.37	1.55	1.79	2.11
84	1.08	1.15	1.24	1.34	1.46	1.60
108	1.06	1.11	1.18	1.24	1.32	1.41
132	1.05	1.09	1.14	1.19	1.25	1.31
			Panel B: *Values of* α_2			
36	1.77	3.16	6.45	16.53	66.11	1041.25
60	1.38	1.87	2.6	3.78	5.8	9.56
84	1.25	1.54	1.91	2.41	3.11	4.10
108	1.19	1.39	1.63	1.93	2.31	2.80
132	1.15	1.30	1.48	1.69	1.95	2.25

Assuming that $\theta_{tp,1} - \theta_{tp,2} = \eta_{tp}$, we can easily obtain:

$$\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|S_{T_2}) = \alpha_2 \Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2)$$

Therefore, if the two measures of the "bias in mean", $\theta_{tp,1} - \theta_{tp,2}$ and η_{ep}, are the same, incorporating the pre-break data in mean estimation improves the portfolio performance more in the case where the covariance matrix is unknown and estimated by S_{T_2}. This addition improvement is associated with α_2 and could be huge if the sample size is small but the portfolio dimension is large.

The following proposition compares the expected CE of efficient portfolios:

Proposition 3.3.4.
Assume that the true covariance matrix Σ_2 is unknown and estimated by S_{T_2}. Let the mean be estimated either by $\bar{r}_T$ or $\bar{r}_{T_2}$. Under the assumptions $A1 - A5$ and based on the sample covariance matrix, S_{T_2}, the expected CE difference between two estimated tangency portfolios using $\bar{r}_T$ and $\bar{r}_{T_2}$ is:

$$\begin{aligned} & \Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}|S_{T_2}\right) \\ = \;& \mathrm{E}\left[CE(w_{ep}(\bar{r}_T, S_{T_2})) - CE(w_{ep}(\bar{r}_{T_2}, S_{T_2}))\right] \\ = \;& \left(\frac{T}{T_1}\beta_1 - \frac{T_2}{T_1}\beta_2\right) \Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2\right) + \frac{1}{2\gamma}\frac{T_1}{T}\left(\beta_2 - \beta_1\right)\Delta^{\mu}_{tp} \end{aligned} \tag{3.3.8}$$

where

$$\beta_1 = \frac{T_2 - 1}{T_2 - N - 1} \tag{3.3.9}$$

$$\beta_2 = \frac{(T_2 - 1)^2 (T_2 - 2)}{(T_2 - N - 1)(T_2 - N)(T_2 - N - 3)} \tag{3.3.10}$$

$$\Delta^{\mu}_{ep} = \frac{N-1}{T_1} + \frac{N-1}{T_2} - (\theta_{ep,1} - \theta_{ep,2}) - \frac{1}{T_1}\delta_{ep} \tag{3.3.11}$$

with

$$\delta_{ep} = tr\left(A_2\left[\Sigma_1 - \Sigma_2\right]\right) \quad \text{and} \quad \theta_{ep,i} = \mu_i' A_2 \mu_i \quad \text{for for } i = 1, 2 \tag{3.3.12}$$

Proof 3.3.4. *See Appendix.*

The roles of β_1 and β_2 for the efficient portfolio are almost the same as the roles of α_1 and α_2 in the tangency portfolio case. β_2 reflects the interaction between the sample mean and the sample covariance matrix. Some values of β_1 and β_2 with respect to different dimensions and sample sizes are given in Table 3.2. It can be seen that β_2 also increases dramatically with increasing dimension, but has a much smaller magnitude than α_2. Therefore, in the small sample case, the empirical tangency portfolio has much larger estimation risk than the empirical efficient portfolio and the benefit of using pre-break data in tangency portfolio estimation can be expected to be larger.

Table 3.2: Values of β_1 and β_2 for different T (number of observations) and N (number of assets).

$T_2 \setminus N$	5	10	15	20	25	30
			Panel A: *Values of* β_1			
36	1.17	1.40	1.75	2.33	3.50	7.00
60	1.09	1.20	1.34	1.51	1.74	2.03
84	1.06	1.14	1.22	1.32	1.43	1.57
108	1.05	1.10	1.16	1.23	1.30	1.39
132	1.04	1.08	1.13	1.18	1.24	1.30
			Panel B: *Values of* β_2			
36	1.6	2.79	5.51	13.35	47.33	462.78
60	1.31	1.75	2.43	3.5	5.3	8.60
84	1.21	1.47	1.82	2.3	2.95	3.87
108	1.16	1.34	1.58	1.86	2.23	2.69
132	1.12	1.27	1.44	1.65	1.89	2.19

3.3.2 Portfolio Combination

3.3.2.1 Estimation Error in Means

Obviously, the $\bar{r}_T$ is a linear combination of $\bar{r}_{T_1}$ and $\bar{r}_{T_2}$. Assume that there is no structural break in the covariance matrix. Consider the general form of linear combination between $\bar{r}_{T_1}$ and $\bar{r}_{T_2}$:

$$\bar{r}_c = c\bar{r}_{T_1} + (1-c)\bar{r}_{T_2}.$$

which has mean $\mu_c = c\mu_1 + (1-c)\mu_2$ and variance $\mathrm{V}[\bar{r}_c] = (\frac{c^2}{T_1} + \frac{(1-c)^2}{T_2})\Sigma_2$. The expected CE loss of the tangency portfolio estimated using such $\bar{r}_c$ can be obtained by:

$$\begin{aligned}
& \mathcal{R}(w_{tp}(\bar{r}_c, \Sigma_2)|w_{tp}^*) \\
= \; & \frac{1}{2\gamma} tr\left(\Sigma_2^{-1}\left[\mathrm{V}[\bar{r}_c] + Bias(\bar{r}_c)^2\right]\right) \\
= \; & \frac{1}{2\gamma} tr\left(\Sigma_2^{-1}\left[(\frac{c^2}{T_1} + \frac{(1-c)^2}{T_2})\Sigma_2 + c^2(\mu_1 - \mu)(\mu_1 - \mu)'\right]\right) \\
= \; & \frac{1}{2\gamma}\left(N\left(\frac{c^2}{T_1} + \frac{(1-c)^2}{T_2}\right) + c^2(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)\right) \qquad (3.3.13)
\end{aligned}$$

It is not difficult to show that $c^* = \frac{T_1}{T}$ minimize $\frac{c^2}{T_1} + \frac{(1-c)^2}{T_2}$, i.e. the expected CE loss caused by the variance of $\bar{r}_c$, but does not take the potential bias into account. Thus, simply ignoring the structural break and using $\bar{r}_T$ is not the most efficient way to combine the pre- and post-break information. The optimal combination weight c can be obtained by solving the FOC of (3.3.13):

$$c = \frac{T_1}{T + \frac{T_1 T_2}{N}(\mu_1 - \mu_2)'\Sigma_2^{-1}(\mu_1 - \mu_2)}$$

which is strictly smaller than $\frac{T_1}{T}$.

3.3.2.2 Estimation Error in Covariances

If we assume that there is no change in the covariance matrix before and after the structural break, pre-break observations should contain useful information for estimating the covariance matrix. As shown in the previous section, unfortunately, it is impossible to obtain an analytical result for the tangency portfolio estimated by S_T. Therefore we consider another way of using pre-break observations to reduce the CE loss caused by the estimated covariance matrix. Instead of estimating the covariance matrix, we construct a linear combination of portfolios based on pre- and post-break observations.

The following proposition compares the expected CE of the combined portfolios with those estimated from S_{T_2}.

Proposition 3.3.5.
Let $\hat{\mu}$ be an arbitrary estimator of the mean return. Consider the combined tangency portfolio and the efficient portfolio constructed as following

$$\hat{w}^c_{tp} = c \cdot w_{tp}(\hat{\mu}, S_{T_1}) + (1-c) w_{tp}(\hat{\mu}, S_{T_2}) \tag{3.3.14}$$

and

$$\hat{w}^c_{ep} = c \cdot w_{ep}(\hat{\mu}, S_{T_1}) + (1-c) w_{ep}(\hat{\mu}, S_{T_2}). \tag{3.3.15}$$

Suppose that the assumptions A1−A5 are satisfied. Additionally assume that $T_1 = T_2$ and $\Sigma_1 = \Sigma_2$. Then combined portfolios outperform uncombined portfolios estimated from post-break sample covariance matrix for all $c \in (0,1)$, more precisely,

$$E\left[CE\left(\hat{w}^c_{tp}\right) - CE\left(w_{tp}(\hat{\mu}, S_{T_2})\right)\right] > 0$$

for all $c \in (0,1)$, and

$$E\left[CE\left(\hat{w}^c_{ep}\right) - CE\left(w_{ep}(\hat{\mu}, S_{T_2})\right)\right] > 0$$

for all $c \in (0,1)$.

Proof 3.3.5. *See Appendix.*

Here, we impose two additional assumptions that the sample sizes are the same before and after the structural break, and there is no structural break in the covariance matrix. Under such assumptions, Proposition 3.3.5 shows that, the combined portfolio always outperforms the portfolio estimated only by using S_{T_2}, and the maximum CE of combined portfolio can be achieved at $c^* = 0.5$[8]. Of course, if these two assumptions are not fulfilled, the performance of the combined portfolio will depend on the size of structural breaks in both the mean and the covariance matrix. As the number of pre-break observations included in estimation increases, the variance of portfolio weights is reduced further on the one hand and the bias of estimated portfolio weights caused by the structural break increases on the other hand. However it seems difficult to obtain an analytical comparison with a meaningful interpretation for such general cases.

[8]See Appendix

Our combination approach can be easily generalized to the multiple break case with unknown break points. Once we can identify the subsamples in which there is no structural break, we can first estimate the mean from the subsamples and combine them. Then, based on the combined mean estimate, we can construct a combination of portfolios constructed from the covariance matrices which are estimated from different subsamples. In this approach, we only need to ensure the absence of structural breaks in our subsamples, but not need to know the exact dates of structural breaks.

3.4 Estimation

In the previous section, we derive the conditions under which pre-break observations can be used to improve performances of the empirical tangency and the efficient portfolio. Obviously, these conditions depend on true population characteristics of the return distribution which have to be estimated in practice.
It is well known, under the i.i.d normality assumption , the sample mean and the sample variance are independent, and the sample covariance matrix $S_{T_2} \sim W_N(T_1 - 1, \Sigma_2)/T_2 - 1$ and therefore the expectation of $S_{T_2}^{-1}$ is:

$$\mathrm{E}\left[S_{T_2}^{-1}\right] = \frac{T_2 - 1}{T_2 - N - 2}\Sigma_2^{-1}.$$

Furthermore, based on the properties of inverse Wishart distribution, it is shown that the expectation of the plug-in estimator of A_2 is given by[9]:

$$\mathrm{E}\left[\hat{A}_2\right] = \mathrm{E}\left[S_{T_2}^{-1} - \frac{S_{T_2}^{-1}\iota\iota' S_{T_2}^{-1}}{\iota' S_{T_2}^{-1}\iota}\right] = \frac{T_2 - 1}{T_2 - N - 1}A_2.$$

It can be observed that most population quantities used in previous sections, e.g. η_{tp} in Proposition 3.3.1, are of quadratic forms weighted by Σ_2^{-1} or A_2. Consider an arbitrary random vector y which is independent with S_{T_2} and A_2. It is easy to obtain the expectation of the following weighted squares of y:

$$\mathrm{E}\left[y' S_{T_2}^{-1} y\right] = \frac{T_2 - 1}{T_2 - N - 2}\mathrm{E}\left[y'\Sigma_2^{-1}y\right] = \frac{T_2 - 1}{T_2 - N - 2}\left[tr\left(\Sigma_2^{-1}\mathrm{V}\left[y\right]\right) + \mathrm{E}\left[y\right]'\Sigma_2^{-1}\mathrm{E}\left[y\right]\right],$$

and

$$\mathrm{E}\left[y' A_2 y\right] = \frac{T_2 - 1}{T_2 - N - 1}\mathrm{E}\left[y' A_2 y\right] = \frac{T_2 - 1}{T_2 - N - 1}\left[tr\left(\Sigma_2^{-1}\mathrm{V}\left[y\right]\right) + \mathrm{E}\left[y\right]' A_2 \mathrm{E}\left[y\right]\right].$$

For instance, let $y = \bar{r}_{T_1} - \bar{r}_{T_2}$ which has the expectation $\mathrm{E}\left[y\right] = \mu_1 - \mu_2$ and the

[9] see e.g. Okhrin & Schmid (2006) or Mori (2004)

variance:

$$\mathrm{V}[y] = \mathrm{V}[\bar{r}_{T_1} - \bar{r}_{T_2}] = \frac{1}{T_1}\Sigma_1 + \frac{1}{T_2}\Sigma_2.$$

Then the expectation of the plug-in estimator $\hat{\eta}_{tp}$ can be easily obtained by:

$$\begin{aligned} & \mathrm{E}\left[(\bar{r}_{T_1} - \bar{r}_{T_2})' S_{T_2}^{-1} (\bar{r}_{T_1} - \bar{r}_{T_2})\right] \\ = \; & \frac{T_2 - 1}{T_2 - N - 2}\left[tr\left(\Sigma_2^{-1}\left[\frac{1}{T_1}\Sigma_1 + \frac{1}{T_2}\Sigma_2\right]\right) + (\mu_1 - \mu_2)' \Sigma_2^{-1} (\mu_1 - \mu_2)\right]. \end{aligned}$$

Therefore, if there is no structural break in the covariance matrix, we can estimate the true population quantities used in Section 3.3 unbiasedly as following:

$$\begin{aligned} \hat{\eta}_{tp}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} (\bar{r}_{T_1} - \bar{r}_{T_2})' S_{T_2}^{-1} (\bar{r}_{T_1} - \bar{r}_{T_2}) - \frac{TN}{T_1 T_2} \\ \hat{\theta}_{tp,i}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} \bar{r}_{T_i}' S_{T_2}^{-1} \bar{r}_{T_i} - \frac{N}{T_i} \\ \hat{\eta}_{ep}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} (\bar{r}_{T_1} - \bar{r}_{T_2})' \hat{A}_2 (\bar{r}_{T_1} - \bar{r}_{T_2}) - \frac{T(N-1)}{T_1 T_2} \\ \hat{\theta}_{ep,i}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} \bar{r}_{T_i}' \hat{A}_2 \bar{r}_{T_i} - \frac{N-1}{T_i} \end{aligned}$$

for $i = 1, 2$. If there exist structural breaks in both the mean and the covariance matrix, we have to adjust the bias in S_{T_2} and obtain the following unbiased estimators:

$$\begin{aligned} \hat{\eta}_{tp}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} \left[(\bar{r}_{T_1} - \bar{r}_{T_2})' S_{T_2}^{-1} (\bar{r}_{T_1} - \bar{r}_{T_2}) - \frac{1}{T_1} tr\left(S_{T_2}^{-1} S_{T_1}\right)\right] - \frac{N}{T_2} \\ \hat{\theta}_{tp,1}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} \left[\bar{r}_{T_1}' S_{T_2}^{-1} \bar{r}_{T_1} - \frac{1}{T_1} tr\left(S_{T_2}^{-1} S_{T_1}\right)\right] \\ \hat{\theta}_{tp,2}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} \bar{r}_{T_2}' S_{T_2}^{-1} \bar{r}_{T_2} - \frac{N}{T_2} \\ \hat{\delta}_{tp}^{u} &= \frac{(T_2 - N - 2)}{(T_2 - 1)} tr\left(S_{T_2}^{-1} S_{T_1}\right) - N \\ \hat{\eta}_{ep}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} \left[(\bar{r}_{T_1} - \bar{r}_{T_2})' \hat{A}_2 (\bar{r}_{T_1} - \bar{r}_{T_2}) - \frac{1}{T_1} tr\left(\hat{A}_2 S_{T_1}\right)\right] - \frac{N-1}{T_2} \\ \hat{\theta}_{ep,1}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} \left[\bar{r}_{T_1}' \hat{A}_2 \bar{r}_{T_1} - \frac{1}{T_1} tr\left(\hat{A}_2 S_{T_1}\right)\right] \\ \hat{\theta}_{ep,1}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} \bar{r}_{T_2}' \hat{A}_2 \bar{r}_{T_2} - \frac{N-1}{T_2} \\ \hat{\delta}_{ep}^{u} &= \frac{(T_2 - N - 1)}{(T_2 - 1)} tr\left(\hat{A}_2 S_{T_1}\right) - (N-1) \end{aligned}$$

In the case of large N, it is possible that using the estimation method described above yields negative values of the weighted squares, e.g. $\hat{\eta}_{tp}^{u} < 0$. To get reasonable inference, we restrict all estimated values of the above weighted squares to be nonnegative.

3.5 Numerical Results

The propositions in the Section 3.3 have provided analytical expressions for the exact performance difference between the portfolios incorporating pre-break information and the pure post-break portfolio strategy. In the following, the performance of such portfolios is compared for different dimensions, sizes of pre- and post-break samples as well as return characteristics that are realistic for financial data. More specifically, we consider three different data sets containing monthly returns of 1) 5 industry portfolios (5PF); 2) 10 industry portfolios (10PF); 3) 25 industry portfolios (25PF) published on Kenneth French's Web site.[10] The sample period is 01/2002 - 12/2011.

It has been shown in the structural-break literature that testing the break in finite samples can lead to invalid inference about the date and the magnitude of the break[11]. Thus, instead of testing potential breaks in our selected data sets, we simply vary the date of the break T_1 within the period 07/2007-06/2009 which includes the most subprime-crisis periods considered by other studies. For each date in this specific crisis-window, the unknown parameters associated with the population return distribution in Propositions 3.3.1 - 3.3.4 are estimated by their unbiased estimators given in Section 3.4 under the normality assumption. Then we plug the estimates into Propositions 3.3.1 - 3.3.4 to calculate the CE differences between the portfolios formed by using pre-break information and the portfolio based on exclusively post-break data.

Table 3.3 presents the minimum values of the CE improvement using pre-break information and the corresponding date of these minima. The results in the 5PF case are relatively informative for detecting the date of the break. It should occur in the middle or the second half year of 2008 and the results for tangency and efficient portfolios are roughly coincide. In the large dimensional cases where $N = 10$ or $N = 25$, it is impossible to draw inference about the break point, but the results clearly demonstrate the usefulness of including pre-break data into the portfolio estimation. Because in large dimensional case, the variances of the estimated portfolios are huge, the minimum CE difference tend to occur at beginning of the considered crisis window where the variance of the estimated weight can be significantly reduced at

[10] *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

[11] See e.g., Perron (2006)

the small cost of the weight bias.

Table 3.3: Minimum values of CE differences between the pure post-break portfolio strategies and the portfolios incorporating pre-break information.

	$\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}\|\Sigma_2)$			$\Delta_{ep}(\bar{r}_T, \bar{r}_{T_2}\|\Sigma_2)$		
	5PF	10PF	25PF	5PF	10PF	25PF
Date	11/2008	07/2007	07/2007	06/2008	07/2007	07/2007
Minimum	0.0006	0.0224	0.0687	0.0019	0.0203	0.0651
	$\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2})$			$\Delta_{ep}(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2})$		
	5PF	10PF	25PF	5PF	10PF	25PF
Date	07/2008	11/2007	07/2007	06/2008	07/2007	07/2007
Minimum	0.0191	0.0504	0.5621	0.0110	0.0455	0.4940

3.5.1 Selection of Sampling Window

To determine the possible date of breaks in the three data sets, we estimate the weighted squared differences η_{tp} and η_{ep} using the unbiased estimator given in the previous section and use the maxima of $\hat{\eta}_{tp}$ and $\hat{\eta}_{ep}$ to define the date of the break. The results are presented in the Table 3.4. The reason for choosing η_{tp} and η_{ep} to determine the break date is that they are always positive by definition and higher values of η_{tp} and η_{ep} imply larger cost of ignoring the structural break in portfolio construction. Nevertheless, other quantities representing the break magnitude, e.g. the bias in the covariance matrix, could be negative, and thus could essentially increase the CE of portfolios using pre-break data.

Table 3.4: Possible dates of structural breaks based on estimated η_{tp} and η_{ep}.

	Panel A: Tangency Portfolio								
	Date	T_1	T_2	$\hat{\eta}^u_{tp}$	$\hat{\theta}^u_{tp,1}$	$\hat{\theta}^u_{tp,2}$	$\hat{\delta}^u_{tp}$	$\Delta_{tp}\left(\bar{r}_T, \bar{r}_{T_2}\|\Sigma_2\right)$	$\Delta_{tp}\left(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2}\right)$
5PF	06/2008*	78	42	0.1208	0.0894	0.1235	0.7803	0.0055	0.0202
10PF	06/2008	78	42	0.0428	0.1444	0.0331	3.1242	0.0299	0.0638
25PF	10/2008	82	38	0.1148	0.0000	0.0000	5.7829	0.0908	4.5127
	Panel B: Efficient Portfolio								
	Date	T_1	T_2	$\hat{\eta}^u_{ep}$	$\hat{\theta}^u_{ep,1}$	$\hat{\theta}^u_{ep,2}$	$\hat{\delta}^u_{ep}$	$\Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}\|\Sigma_2\right)$	$\Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2}\right)$
5PF	06/2008	78	42	0.1130	0.0714	0.0000	1.1762	0.0019	0.0110
10PF	06/2008	78	42	0.0889	0.1214	0.0000	2.9289	0.0215	0.0539
25PF	10/2008	82	38	0.2628	0.0000	0.0697	4.6282	0.0706	3.9928

It can be seen that the detected dates of structural break based on $\hat{\eta}^u_{tp}$ are broadly consistent with that based on $\hat{\eta}^u_{ep}$ except the 5PF case. In the 5PF case, the maximum

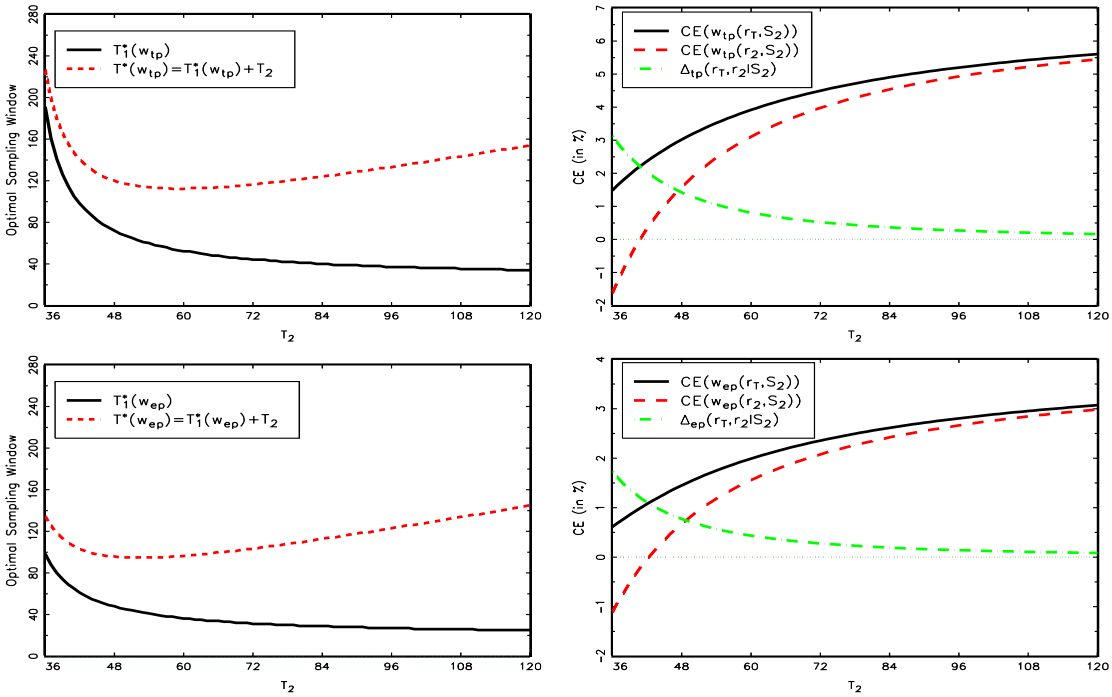


Figure 3.1: Optimal sampling windows and the expected CE of portfolio based on optimal sampling window for different post-break sample sizes in the 5PF case. Left Panel: Optimal pre-break sampling window T_1^* , as well as the entire sample (dashed line) $T^* = T_1^* + T_2$ for the tangency portfolio (left upper panel) and the efficient portfolio (left lower panel). Right Panel: CE (in %) of tangency portfolios(upper panel) and efficient portfolio (lower panel) based on optimal sampling windows (red dashed line), on the post-break samples (solid line), as well as their differences (green dashed line). $\gamma = 2$.

of estimated $\hat{\eta}_{tp}$ occurs at 11/2008 but the second maximal value of $\hat{\eta}_{tp}^u$ occurs at 06/2008. Hence, 06/2008 is taken as the break point for 5PF in the further study. In 5PF case, since the size of the post-break sample is relative large for the dimension, the detected date of the structural break is also roughly consistent with the previous results about the minimum improvement of using pre-break data. Additionally, we can observe that the detected break dates are consistent across the 5PF and 10PF cases. This is not surprising because both of them are industry portfolios formed according to their SIC codes. But the 25PF is constructed based on economic properties of the firms and therefore responses differently to market shocks. In the following, the estimated parameter values in Table 3.4 are taken as true population parameters to obtain further numerical results.

According to the of parameter values given in Table 3.4, we can calculate the optimal size of sampling windows which leads to highest expected CE of the estimated portfolio[12]. The results for the 5PF case are reported in Figure 3.1[13]. It is clear that, as the post-break sample size increases, the variance of post-break portfolio strategies are dramatically reduced, and thus the benefit of using pre-break also decreases. But even for $T = 120$ which means that ten year post-break observations are available, including some pre-break observations can still helps.

Table 3.5: CE improvement (in %) for different pre- and post-break sample sizes.

		$\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2})$			$\Delta_{ep}(\bar{r}_T, \bar{r}_{T_2}\|S_{T_2})$		
	$T_2 \backslash T_1$	36	48	60	36	48	60
	36	2.60	2.79	2.90	1.55	1.64	1.68
5PF	48	1.31	1.38	1.42	0.76	0.77	0.76
	60	0.78	0.81	0.81	0.44	0.42	0.39
	36	6.95	8.08	8.96	6.15	7.00	7.64
10PF	48	2.99	3.53	3.96	2.72	3.12	3.42
	60	1.66	1.98	2.24	1.52	1.75	1.92
	36	549.37	630.94	692.63	477.25	538.95	584.16
25PF	48	61.44	71.74	79.77	66.15	75.68	82.80
	60	21.38	25.25	28.34	24.53	28.32	31.17

Table 3.5 reports the improvement of portfolio performance with respect to some realistic pre- and post break sample size for the three data sets. Comparing to the magnitude of CE increment in the 5PF and 10PF cases, we can observe a dramatic improvement of the portfolio performance in the 25PF case when the pre-break observations are included. There are at least two reasons for this result. First, the ratio of dimension over the size of post-break sample is large, and hence the variance of estimated post-break portfolio is large. Second, since the 25 portfolios are formed on the firm characteristics, namely the firm size and its book to market ratio, and updated at the end of each June, they are less sensitive to the structural break and the pre-break data are much more informative than other data sets.

3.5.2 Combined Portfolios and Portfolio Based on S_T

Section 3.3 shows that, under the assumption that there is no structural change in the covariance matrix and $T_1 = T_2$, the combined portfolios always outperform

[12] It is obvious that the CE differences in propositions from Section 3.3 are non-linear functions of T_1 and have non-unique maxima. The optimal T_1 reported here is the first date before which the CE differences are always increasing.

[13] The results for 10PF and 25PF cases are not presented here, because for resealable post-sample sizes T_2, the resulted optimal sizes of pre-break samples are too large and do not make any sense for the practical implementation.

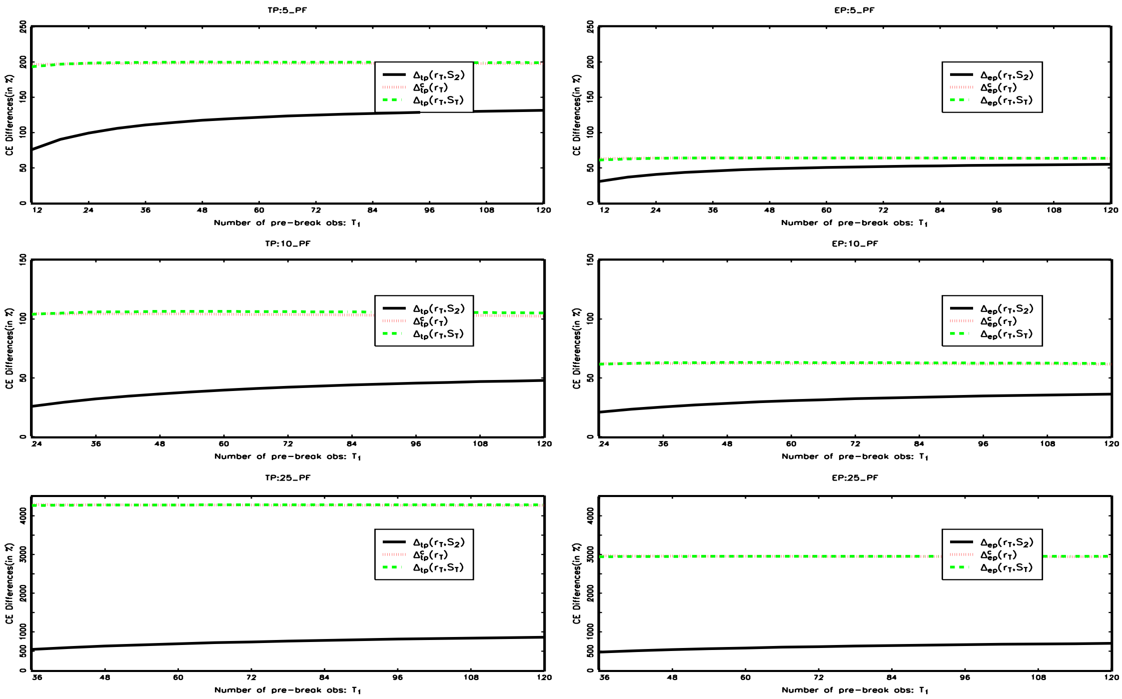


Figure 3.2: CE improvement of combined portfolio (red dotted line), portfolio based on r_T and S_T (green dash line), as well as portfolio based on r_T and S_{T_2} (black solid line) relative to the pure post-break portfolio. Left panel: tangency portfolio case. Right panel: efficient portfolio case. $\gamma = 2$.

uncombined portfolios, and it is not difficult to show that the optimal combination weights is $c = 1/2$. However, when such assumptions are violated, the optimal combination is affected by too many factors related to the return distribution as well as the date and magnitude of breaks, and it seems very difficult to draw any inference about the dominance of the portfolios without particular inputs of return characteristics.

In the following, we conduct a simulation study to compare our combined portfolio strategy, w^c_{tp} and w^c_{ep} with the uncombined portfolio based on r_T and S_{T_2}. Since the optimal combination weights in equations (3.3.14) and (3.3.15) depend on the return distributions before and after the structural break, we take the natural weight $c = \frac{T_1}{T}$. In addition, although the distributional properties of the sample covariance matrix S_T are intractable, the simulation study allows us to include the portfolio estimated from S_T into the comparison, and to see whether whether simply ignoring all breaks in the return distribution still provides acceptable portfolio performance. Based on the specified means and covariance matrices, we simulate multivariate i.i.d. normally distributed returns $10,000$ times. The post-break sample size T_2 is equal to: 1) 12 for 5PF; 2) 24 for 10PF; 3) 36 for 25PF, and pre-break sample size T_1 varies from T_2 to 120. For each simulated return series, we estimate the weights of the combined portfolio, the portfolio based on r_T and S_{T_2} as well as the portfolio based on r_T and S_T. We calculate their CEs based on the true mean and true covariance matrix, and then plot their expected CE improvement relative to the pure post-break portfolio in Figure 3.2. It can be seen that the combined portfolio and portfolio based on S_T perform very similarly but significantly better than the portfolio estimated from r_T and S_{T_2}. The similar performance of the combined portfolio and portfolio based on S_T could be explained by the fact that the combination weight $c = \frac{T_1}{T}$, and hence the contribution of pre- and post data to the combined portfolio is similar to their contribution to S_T.

3.6 Conclusion

This paper utilizes the spirit of Pesaran & Timmermann (2007) and analytically studies the possible benefit of the pre-break information for the post-break portfolio selection problem. Our theoretical results reveal the crucial effect of the trade-off between the bias and the variance of empirical portfolios on the ultimate portfolio performance and clearly show why and when the pre-break data can help.

Since the sample mean is obviously a linear combination of pre- and post-break sample estimated of the corresponding means, we generalize this idea to combine the portfolios estimated from pre- and post-break data. Under simplified assumptions, we show analytically that the combined portfolios can significantly reduce the estimation risk in the covariance matrix and outperforms the portfolio estimated from the post-break sample covariance matrix. Relaxing the assumption made in theoretical derivation, we perform a simulation study to confirm our theoretical results for more general scenarios. In the simulation study, we also show that simply ignoring the structural break in the mean and covariance matrix leads to similar performance as the combined portfolio with the combination weight calculated from the ratio of pre-break sample size over the size of the entire sample.

In this paper, we only consider the case of the presence of a single structural break. But it is straightforward to generalize our combination approach to the multiple structural break cases. In addition, since the combination weight calculated from the sample size ratio is definitively not optimal, feasible methods for computing the optimal combination weight as well as the optimal sampling windows are desired in future studies.

Bibliography

Antoine, B. (2012), 'Portfolio selection with estimation risk: A test-based approach', *Journal of Financial Econometrics* **10**(1), 164–197.

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

Britten-Jones, M. (1999), 'The sampling error in estimates of mean-variance efficient portfolio weights', *The Journal of Finance* **54**(2), 655–671.

Brodie, J., Daubechies, I., Molc, C. D., Giannone, D. & Loris, I. (2009), 'Sparse and stable markowitz portfolios', *PNAS* **106**(30), 12267–12272.

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

Engle, R. & Colacito, R. (2006), 'Testing and valuing dynamic correlations for asset allocation', *Journal of Business & Economic Statistics* **24**(2), 238–253.

Frahm, G. (2010), An analytical investigation of estimators for expected asset returns from the perspective of optimal asset allocation. University of Cologne, Seminar of Economic and Social Statistics, Discussion Paper 1/10.

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Frahm, G., Wickern, T. & Wiechers, C. (2012), 'Multiple tests for the performance of different investment strategies', *AStA Advances in Statistical Analysis* **96**(3), 343–383.

Frahm, G. & Wiechers, C. (2011), On the diversification of portfolios of risky assets. University of Cologne, Seminar of Economic and Social Statistics, Discussion Paper 2/11.

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

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

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

Liu, H. & Pohlmeier, W. (2013), Risk preferences and estimation risk in portfolio choice. Working Paper.

Merton, R. C. (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**(4), 323–361.

Mori, H. (2004), 'Finite sample properties of estimators for the optimal portfolio weight', *Journal of the Japan Statistical Society* **34**, 27–46.

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

Perron, P. (2006), Dealing with structural breaks, *in* K.Patterson & T. C.Mills, eds, 'Palgrave Handbook of Econometrics', Vol. 1, London : Palgrave Macmillan.

Pesaran, M. H. & Timmermann, A. (2005), 'Small sample properties of forecasts from autoregressive models under structural breaks', *Journal of Econometrics* **129**, 183–217.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

3.7 Appendix

Proof 3.7.1 (*Proposition 3.3.1*)**.** *Kan & Zhou (2007) show that the expected CE loss using the sample mean* $\bar{r}_{T_2}$ *and the true population covariance matrix is:*

$$\mathcal{R}(w_{tp}(\bar{r}_{T_2}, \Sigma_2)|w_{tp}^*) = \frac{1}{2\gamma}\frac{N}{T_2}. \tag{3.7.1}$$

Based on the risk function given by equation (3.2.12), the expected CE loss of the estimated tangency portfolio using $\bar{r}_T$ *can be obtain by:*

$$\begin{aligned} & \mathcal{R}(w_{tp}(\bar{r}_T, \Sigma_2)|w_{tp}^*) \\ = & \frac{1}{2\gamma}tr\left(\Sigma_2^{-1}\left[\mathrm{V}\left[\bar{r}_T\right] + Bias(\bar{r}_T)^2\right]\right) \\ = & \frac{1}{2\gamma}tr\left(\Sigma_2^{-1}\left[\frac{T_1}{T^2}\Sigma_1 + \frac{T_2}{T^2}\Sigma_2 + (\mu_T - \mu)(\mu_T - \mu)'\right]\right) \\ = & \frac{1}{2\gamma}\left[\frac{N}{T} + \frac{T_1}{T^2}tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right) + \frac{T_1^2}{T^2}(\mu_1 - \mu_2)'\Sigma_2^{-1}(\mu_1 - \mu_2)\right] \end{aligned} \tag{3.7.2}$$

The difference between expected CE losses of empirical tangency portfolios based on $\bar{r}_T$ *and* $\bar{r}_{T_2}$ *is:*

$$\begin{aligned} & \Delta_{tp}\left(\bar{r}_T, \bar{r}_{T_2}|\Sigma\right) \\ = & \mathrm{E}\left[CE(w_{tp}(\bar{r}_T, \Sigma_2)) - CE(w_{tp}(\bar{r}_{T_2}, \Sigma_2))\right] \\ = & \mathcal{R}(w_{tp}(\bar{r}_{T_2}, \Sigma)|w_{tp}^*) - \mathcal{R}(w_{tp}(\bar{r}_T, \Sigma)|w_{tp}^*) \\ = & \frac{T_1^2}{2\gamma T^2}\left[N\left(\frac{1}{T_1} + \frac{1}{T_2}\right) - (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{T_1}tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right)\right] \end{aligned}$$

Proof 3.7.2 (*Proposition 3.3.2*)**.** *Similar to the tangency portfolio case, Liu & Pohlmeier (2013) show that the expected CE loss of the estimated efficient portfolio using the sample mean* $\bar{r}_{T_2}$ *and the true population covariance matrix is:*

$$\mathcal{R}(w_{ep}(\bar{r}_{T_2}, \Sigma_2)|w_{ep}^*) = \frac{1}{2\gamma}\frac{N-1}{T_2}. \tag{3.7.3}$$

Using risk function (3.2.15), it can be easily shown that

$$\begin{aligned} & \mathcal{R}(w_{ep}(\bar{r}_T, \Sigma_2)|w_{tp}^*) \\ = & \frac{1}{2\gamma}tr\left(A_2\left[\mathrm{V}\left[\bar{r}_T\right] + Bias(\bar{r}_T)^2\right]\right) \\ = & \frac{1}{2\gamma}tr\left(A_2\left[\frac{T_1}{T^2}\Sigma_1 + \frac{T_2}{T^2}\Sigma_2 + (\mu_T - \mu)(\mu_T - \mu)'\right]\right) \\ = & \frac{1}{2\gamma}\left[\frac{N-1}{T} + \frac{T_1}{T^2}tr\left(A_2\left[\Sigma_1 - \Sigma_2\right]\right) + \frac{T_1^2}{T^2}(\mu_1 - \mu_2)'A_2(\mu_1 - \mu_2)\right] \end{aligned} \tag{3.7.4}$$

The last equality holds because $tr(A_2\Sigma_2) = N - 1$[14]. *Therefore, the difference between expected CE losses of empirical efficient portfolios based on* $\bar{r}_T$ *and* $\bar{r}_{T_2}$ *can be expressed as:*

$$
\begin{aligned}
& \Delta_{ep}\left(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2\right) \\
= \; & \mathrm{E}\left[CE(w_{ep}(\bar{r}_T, \Sigma_2)) - CE(w_{ep}(\bar{r}_{T_2}, \Sigma_2))\right] \\
= \; & \mathcal{R}(w_{ep}(\bar{r}_{T_2}, \Sigma_2)|w^*_{ep}) - \mathcal{R}(w_{tp}(\bar{r}_T, \Sigma_2)|w^*_{ep}) \\
= \; & \frac{T_1^2}{2\gamma T^2}\left[\left(\frac{N-1}{T_1} + \frac{N-1}{T_2}\right) - (\mu_1 - \mu_2)' A_2 (\mu_1 - \mu_2) - \frac{1}{T_1} tr\left(A_2\left[\Sigma_1 - \Sigma_2\right]\right)\right]
\end{aligned}
$$

Proof 3.7.3 (*Proposition 3.3.3*)**.** *As argued by Okhrin & Schmid (2006), in the case where* $N \geq 2$*, it is difficult to derive the distributional properties of tangency portfolio weights estimated from the sample mean and the sample covariance matrix even if returns are i.i.d normal. Therefore, we cannot use the risk function (3.2.12) to obtain an analytical comparison result.*
Let $\hat{\mu}$ *denote an estimate of the mean vector. According to Kan & Zhou (2007), let* $W = \Sigma_2^{-\frac{1}{2}} S_{T_2} \Sigma_2^{-\frac{1}{2}} \sim W_N\left((T_2 - 1), I_N\right)/(T_2 - 1)$ *and its first two inverse moments are given by:*

$$
\begin{aligned}
E\left[W^{-1}\right] &= \frac{T_2 - 1}{T_2 - N - 2} I_n =: \alpha_1 \cdot I_N \\
E\left[W^{-2}\right] &= \frac{(T_2 - 1)^2 (T_2 - 2)}{(T_2 - N - 1)(T_2 - N - 2)(T_2 - N - 4)} I_N =: \alpha_2 \cdot I_N
\end{aligned}
$$

It is easy to see that $\alpha_2 > \alpha_1^2 > \alpha_1$*. Kan & Zhou (2007) show that, conditional on* $\hat{\mu}$*, the expected CE of the estimated tangency portfolio using the sample covariance matrix is:*

$$
E\left[CE\left(w_{tp}(\hat{\mu}, S_{T_2})\right)|\hat{\mu}\right] = \frac{\alpha_1}{\gamma}\hat{\mu}'\Sigma^{-1}\mu - \frac{\alpha_2}{2\gamma}\hat{\mu}'\Sigma^{-1}\hat{\mu}
$$

Setting $\tilde{\gamma} = \frac{\alpha_2}{\alpha_1^2}\gamma$ *and* $\tilde{\mu} = \frac{\alpha_2}{\alpha_1}\hat{\mu}$*, we can rewrite the expected CE of the empirical tangency portfolio as:*

$$
\mathrm{E}\left[CE\left(w_{tp}(\hat{\mu}, S_{T_2})\right)\right] = \mathrm{E}\left[\frac{1}{\tilde{\gamma}}\tilde{\mu}'\Sigma_2^{-1}\mu - \frac{1}{2\tilde{\gamma}}\tilde{\mu}'\Sigma_2^{-1}\tilde{\mu}\right] = \mathrm{E}\left[CE_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \tilde{\mu}, \Sigma_2)\right)\right]
$$

where the tangency portfolio weight

$$
w_{tp}(\tilde{\gamma}, \tilde{\mu}, \Sigma_2) = \frac{1}{\tilde{\gamma}}\Sigma_2^{-1}\tilde{\mu}
$$

and $CE_{\tilde{\gamma}}$ *represents the CE evaluated with the new risk aversion parameter* $\tilde{\gamma}$*.*
Therefore, we can compare the performance of $w_{tp}(\bar{r}_{T_2}, S_{T_2})$ *and* $w_{tp}(\bar{r}_T, S_{T_2})$ *in terms*

[14] see e.g. Liu & Pohlmeier (2013)

of CE with the new risk aversion parameter $\tilde{\gamma}$:

$$\begin{aligned}\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|S_{T_2}) &= \mathrm{E}\left[CE(w_{tp}(\bar{r}_T, S_{T_2})) - CE(w_{tp}(\bar{r}_{T_2}, S_{T_2}))\right] \\ &= \Big[CE_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)\right) - \mathrm{E}\left[CE_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \tilde{r}_2, \Sigma_2)\right)\right]\Big] \\ &\quad -\Big[CE_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)\right) - \mathrm{E}\left[CE_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \tilde{r}_2, \Sigma_2)\right)\right]\Big] \\ &= \mathcal{R}_{\tilde{\gamma}}(w_{tp}(\tilde{\gamma}, \tilde{r}_2, \Sigma_2)|w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)) - \mathcal{R}_{\tilde{\gamma}}(w_{tp}(\tilde{\gamma}, \tilde{r}_T, \Sigma_2)|w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2))\end{aligned}$$

where $\tilde{r}_2 = \frac{\alpha_2}{\alpha_1}\bar{r}_{T_2}$, $\tilde{r}_T = \frac{\alpha_2}{\alpha_1}\bar{r}_T$, $\mathcal{R}_{\tilde{\gamma}}$ is the risk function defined in terms of $CE_{\tilde{\gamma}}$ and $w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)$ is the optimal tangency portfolio based on the new risk aversion parameter $\tilde{\gamma}$, the true mean and the true covariance matrix. Using risk function given in (3.2.12), the expected CE loss of the tangency portfolio estimated using post-break observations can be obtained by:

$$\begin{aligned}&\mathcal{R}_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \tilde{r}_2, \Sigma_2)|w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)\right) \\ &= \frac{1}{2\tilde{\gamma}} tr\left(\Sigma_2^{-1}\left[\left(\frac{\alpha_2}{\alpha_1}\right)^2 \frac{1}{T_2}\Sigma_2 + \left[\left(\frac{\alpha_2}{\alpha_1}\right)\mu_2 - \mu_2\right]\left[\left(\frac{\alpha_2}{\alpha_1}\right)\mu_2 - \mu_2\right]'\right]\right) \\ &= \frac{1}{2\gamma}\alpha_2\left[\frac{N}{T_2} + \left(\frac{\alpha_2 - \alpha_1}{\alpha_2}\right)^2 \mu_2'\Sigma_2^{-1}\mu_2\right]\end{aligned}$$

and the expected loss of the estimated tangency portfolio based on $\bar{r}_T$ is:

$$\begin{aligned}&\mathcal{R}_{\tilde{\gamma}}\left(w_{tp}(\tilde{\gamma}, \tilde{r}_T, \Sigma_2)|w_{tp}(\tilde{\gamma}, \mu_2, \Sigma_2)\right) \\ &= \frac{1}{2\tilde{\gamma}} tr\left(\Sigma_2^{-1}\left[\left(\frac{\alpha_2}{\alpha_1}\right)^2 \Sigma_T + \left(\frac{\alpha_2}{\alpha_1}\mu_T - \mu\right)\left(\frac{\alpha_2}{\alpha_1}\mu_T - \mu\right)'\right]\right) \\ &= \frac{\alpha_2}{2\gamma}\left[\frac{N}{T} + \frac{T_1}{T^2} tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right)\right. \\ &\quad \left. + \left(\frac{T_1}{T}(\mu_1 - \mu_2) + \frac{\alpha_2 - \alpha_1}{\alpha_2}\mu_2\right)'\Sigma_2^{-1}\left(\frac{T_1}{T}(\mu_1 - \mu_2) + \frac{\alpha_2 - \alpha_1}{\alpha_2}\mu_2\right)\right]\end{aligned}$$

Therefore, based on the sample covariance matrix S_{T_2}, the expected CE difference *between the estimated tangency portfolios using $\bar{r}_T$ and $\bar{r}_{T_2}$ is:*

$$\begin{aligned}
&\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|S_{T_2}) \\
= \;& \frac{\alpha_2}{2\gamma}\left[\frac{N}{T_2} - \frac{N}{T} - \frac{T_1}{T^2}tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right) - \frac{T_1^2}{T^2}\left(\mu_1 - \mu_2\right)'\Sigma_2^{-1}\left(\mu_1 - \mu_2\right)\right] \\
& - \frac{1}{2\gamma}\frac{T_1}{T}\left(\alpha_2 - \alpha_1\right)\left[2\left(\mu_1 - \mu_2\right)\Sigma_2^{-1}\mu_2\right] \\
= \;& \frac{\alpha_2}{2\gamma}\frac{T_1^2}{T^2}\left[\frac{NT}{T_1T_2} - \frac{1}{T_1}tr\left(\Sigma_2^{-1}\left[\Sigma_1 - \Sigma_2\right]\right) - \left(\mu_1 - \mu_2\right)'\Sigma_2^{-1}\left(\mu_1 - \mu_2\right)\right] \\
& - \frac{1}{2\gamma}\frac{T_1}{T}\left(\alpha_2 - \alpha_1\right)\left(\mu_1'\Sigma_2^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2 - \left(\mu_1 - \mu_2\right)'\Sigma_2^{-1}\left(\mu_1 - \mu_2\right)\right) \\
= \;& \frac{1}{2\gamma}\left(\frac{T}{T_1}\alpha_1 - \alpha_2\frac{T_2}{T_1}\right)\cdot\frac{T_1^2}{T^2}\left[\frac{NT}{T_1T_2} - \frac{1}{T_1}tr\left(\Sigma_2\left[\Sigma_1 - \Sigma_2\right]\right) - \left(\mu_1 - \mu_2\right)'\Sigma_2^{-1}\left(\mu_1 - \mu_2\right)\right] \\
& + \frac{1}{2\gamma}\frac{T_1}{T}\left(\alpha_2 - \alpha_1\right)\left[\frac{NT}{T_1T_2} - \frac{1}{T_1}tr\left(\Sigma_2\left[\Sigma_1 - \Sigma_2\right]\right) - \left(\mu_1'\Sigma_2^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2\right)\right] \\
= \;& \left(\frac{T}{T_1}\alpha_1 - \frac{T_2}{T_1}\alpha_2\right)\cdot\Delta_{tp}(\bar{r}_T, \bar{r}_{T_2}|\Sigma_2) \\
& + \frac{1}{2\gamma}\frac{T_1}{T}\left(\alpha_2 - \alpha_1\right)\left[\frac{N}{T_1} + \frac{N}{T_2} - \frac{1}{T_1}tr\left(\Sigma_2\left[\Sigma_1 - \Sigma_2\right]\right) - \left(\mu_1'\Sigma_2^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2\right)\right]
\end{aligned}$$

Proof 3.7.4 (*Proposition 3.3.4*)**.** *Let $\hat{\mu}$ be an estimate of the mean vector. Okhrin & Schmid (2006) show that, conditional on $\hat{\mu}$, the first two moments of the efficient portfolio weight estimated from the the sample covariance matrix, S_{T_2}, can be obtained by:*

$$\begin{aligned}
\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right] &= \frac{1}{T_2 - N - 1}\frac{A_2}{\iota'\Sigma_2^{-1}\iota} + \frac{1}{\gamma^2}\left(c_1A_2\cdot\hat{\mu}\hat{\mu}'\cdot A_2 + c_2\hat{\mu}'\cdot A_2\cdot\hat{\mu}\cdot A_2\right) \\
\mathrm{E}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right] &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma_2^{-1}\iota} + \frac{T_2 - 1}{T_2 - N - 1}\cdot\frac{1}{\gamma}A_2\cdot\hat{\mu} = \frac{\Sigma^{-1}\iota}{\iota'\Sigma_2^{-1}\iota} + \frac{\beta_1}{\gamma}\cdot A_2\cdot\hat{\mu}
\end{aligned}$$

where

$$c_1 = \frac{\left(T_2 - 1\right)^2\left(T_2 - N + 1\right)}{\left(T_2 - N\right)\left(T_2 - N - 1\right)^2\left(T_2 - N - 3\right)} \quad \text{and} \quad c_2 = \frac{\left(T_2 - 1\right)^2}{\left(T_2 - N\right)\left(T_2 - N - 1\right)\left(T_2 - N - 3\right)}.$$

It is known that the variance of the estimated portfolio weight can be decomposed as:

$$\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})\right] = \mathrm{E}\left[\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right]\right] + \mathrm{V}\left[\mathrm{E}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right]\right]$$

According to Liu & Pohlmeier (2013), the expected CE loss of the estimated portfolio *based on $\hat{\mu}$ and S_{T_2} can be given by:*

$$
\begin{aligned}
& \mathcal{R}\left(w_{ep}(\hat{\mu}, S_{T_2})|w^*_{ep}\right) \\
= & \frac{\gamma}{2} tr\left(\Sigma_2 \cdot \left[\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})\right] + Bias(w_{ep}(\hat{\mu}, S_{T_2}))^2\right]\right) \\
= & \frac{\gamma}{2} tr\Bigg(\Sigma_2 \cdot \Bigg[\mathrm{E}\left[\frac{1}{T_2 - N - 1}\frac{A_2}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma^2}\left(c_1 A_2 \cdot \hat{\mu}\hat{\mu}' \cdot A_2 + c_2 \hat{\mu}' \cdot A_2 \cdot \hat{\mu} \cdot A_2\right)\right] \\
& + \mathrm{V}\left[\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{\beta_1}{\gamma} \cdot A_2 \cdot \hat{\mu}\right] + \left(\frac{\beta_1}{\gamma} \cdot A_2 \cdot \mathrm{E}\left[\hat{\mu}\right] - \frac{1}{\gamma} A_2 \cdot \mu_2\right)\left(\frac{\beta_1}{\gamma} \cdot A_2 \cdot \mathrm{E}\left[\hat{\mu}\right] - \frac{1}{\gamma} A_2 \cdot \mu_2\right)'\Bigg]\Bigg) \\
= & \frac{\gamma}{2}\frac{N-1}{T_2 - N - 1}\frac{1}{\iota'\Sigma_2^{-1}\iota} + \frac{1}{2\gamma}\left(c_1 + c_2(N-1)\right) tr\left(A_2 \cdot \mathrm{E}\left[\hat{\mu}\hat{\mu}'\right]\right) + \frac{\beta_1^2}{2\gamma} tr\left(A_2 \cdot \mathrm{V}\left[\hat{\mu}\right]\right) \\
& + \frac{1}{2\gamma} A_2 \cdot \left(\beta_1 \cdot \mathrm{E}\left[\hat{\mu}\right] - \mu_2\right)\left(\beta_1 \cdot \mathrm{E}\left[\hat{\mu}\right] - \mu_2\right)' \\
= & \frac{\gamma}{2}\frac{N-1}{T_2 - N - 1}\frac{1}{\iota'\Sigma^{-1}\iota} + \frac{1}{2\gamma}\left(c_1 + c_2(N-1)\right) tr\left(A_2 \,\mathrm{E}\left[\hat{\mu}\right]\mathrm{E}\left[\hat{\mu}\right]'\right) \\
& + \frac{1}{2\gamma}\left[\beta_1^2 + c_1 + c_2(N-1)\right] tr\left(A_2 \cdot \mathrm{V}\left[\hat{\mu}\right]\right) + \frac{1}{2\gamma} tr\left[A_2 \cdot \left(\beta_1 \cdot \mathrm{E}\left[\hat{\mu}\right] - \mu_2\right)\left(\beta_1 \,\mathrm{E}\left[\hat{\mu}\right] - \mu_2\right)'\right] \\
= & \frac{\gamma}{2}\frac{N-1}{T_2 - N - 1}\sigma^2_{gmv} + \frac{\beta_2}{2\gamma} tr\left(A_2 \cdot \mathrm{V}\left[\hat{\mu}\right]\right) + \frac{\beta_2}{2\gamma}\left(\mathrm{E}\left[\hat{\mu}\right] - \frac{1}{\beta_1}\mu_2\right)' A_2 \left(\mathrm{E}\left[\hat{\mu}\right] - \frac{1}{\beta_1}\mu_2\right) \\
& + \frac{1}{2\gamma}\left(\beta_2 - \beta_1^2\right)\left(\frac{2}{\beta_1}\mathrm{E}\left[\hat{\mu}\right]' A_2 \mu_2 - \frac{1}{\beta_1^2}\mu_2' A_2 \mu_2\right)
\end{aligned}
$$

where $\sigma^2_{gmv} = \frac{1}{\iota'\Sigma^{-1}\iota}$ is the variance of the global minimum variance portfolio based on the true population covariance matrix, Σ_2. The last equality holds because[15]*:*

$$
\beta_2 = \frac{(T_2-1)^2(T_2-2)}{(T_2-N-1)(T_2-N)(T_2-N-3)} = c_1 + c_2\left(N-1\right) + \beta_1^2
$$

As shown by Liu & Pohlmeier (2013), β_1 is determined by the first moment of $\hat{A}_2$, and similar to the role of α_2 in the tangency portfolio case, β_2 captures the interaction effect between the sample mean and the sample covariance matrix.
In the case where $\hat{\mu} = \bar{r}_{T_2}$, we can obtain the expected CE loss of the estimated efficient portfolio based on exclusively post-break observation:

$$
\begin{aligned}
& \mathcal{R}\left(w_{ep}(\bar{r}_{T_2}, S_{T_2})|w^*_{ep}\right) \\
= & \frac{\gamma}{2}\frac{N-1}{T_2 - N - 1}\sigma^2_{gmv} + \frac{\beta_2}{2\gamma}\frac{N-1}{T_2} \\
& + \frac{1}{2\gamma}\left[\beta_2\left(\frac{\beta_1 - 1}{\beta_1}\right)^2 \mu_2' A_2 \mu_2 + \left(\beta_2 - \beta_1^2\right)\left(\frac{2}{\beta_1}\mu_2' A_2 \mu_2 - \frac{1}{\beta_1^2}\mu_2' A_2 \mu_2\right)\right] \\
= & \frac{\gamma}{2}\frac{N-1}{T_2 - N - 1}\sigma^2_{gmv} + \frac{\beta_2}{2\gamma}\frac{N-1}{T_2} + \frac{\mu_2' A_2 \mu_2}{2\gamma}\left(\beta_2 - \beta_1^2 + (\beta_1 - 1)^2\right)
\end{aligned}
$$

This result is consistent with the one derived by Liu & Pohlmeier (2013). Plugging

[15] See Liu & Pohlmeier (2013)

$\hat{\mu} = \bar{r}_T$ into the risk function described above, we can obtain the expected CE loss of the estimated efficient portfolio based on $\bar{r}_T$ and S_{T_2}:

$$\begin{aligned}
&\mathcal{R}\left(w_{ep}(\bar{r}_T, S_{T_2})|w^*_{ep}\right) \\
= \;& \frac{\gamma}{2}\frac{N-1}{T_2-N-1}\sigma^2_{gmv} + \frac{\beta_2}{2\gamma}\left(\frac{N-1}{T} + \frac{T_1}{T^2}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right)\right) \\
& + \frac{\beta_2}{2\gamma}\left(\frac{T_1}{T}\left(\mu_1-\mu_2\right) + \frac{\beta_1-1}{\beta_1}\mu_2\right)' A_2 \left(\frac{T_1}{T}\left(\mu_1-\mu_2\right) + \frac{\beta_1-1}{\beta_1}\mu_2\right) \\
& + \frac{\beta_2-\beta_1^2}{2\gamma}\left[\frac{2}{\beta_1}\left(\frac{T_1}{T}\left(\mu_1-\mu_2\right)+\mu_2\right)' A_2\mu_2 - \frac{1}{\beta_1^2}\mu_2' A_2 \mu_2\right]
\end{aligned}$$

Therefore, the portfolio performance of using $\bar{r}_{T_2}$ and $\bar{r}_T$ can be compared by taking the difference of their risks. Based on the sample covariance matrix S_{T_2}, the expected CE difference between the estimated efficient portfolios using $\bar{r}_T$ and $\bar{r}_{T_2}$ is:

$$\begin{aligned}
&\mathrm{E}\left[CE(w_{ep}(\bar{r}_T, S_{T_2})) - CE(w_{ep}(\bar{r}_{T_2}, S_{T_2}))\right] \\
= \;& \frac{\beta_2}{2\gamma}\left[\frac{N-1}{T_2} - \frac{N-1}{T} - \frac{T_1}{T^2}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right)\right] \\
& - \frac{\beta_2}{2\gamma}\left[\frac{T_1^2}{T^2}\left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right) + \frac{2T_1}{T}\cdot\frac{\beta_1-1}{\beta_1}\cdot\left(\mu_1-\mu_2\right)' A_2\mu_2\right] \\
& - \frac{\beta_2-\beta_1^2}{2\gamma}\cdot\frac{2}{\beta_1}\cdot\frac{T_1}{T}\cdot\left(\mu_1-\mu_2\right)' A_2\mu_2 \\
= \;& \frac{\beta_2}{2\gamma}\left[\frac{N-1}{T_2} - \frac{N-1}{T} - \frac{T_1}{T^2}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right) - \frac{T_1^2}{T^2}\left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right)\right] \\
& - \frac{1}{2\gamma}\frac{T_1}{T}\left(\frac{\beta_2\beta_1-\beta_2}{\beta_1} + \frac{\beta_2-\beta_1^2}{\beta_1}\right)\left(2\left(\mu_1-\mu_2\right)' A_2\mu_2\right) \\
= \;& \frac{\beta_2}{2\gamma}\left[\frac{N-1}{T_2} - \frac{N-1}{T} - \frac{T_1}{T^2}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right) - \frac{T_1^2}{T^2}\left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right)\right] \\
& - \frac{1}{2\gamma}\frac{T_1}{T}\left(\beta_2-\beta_1\right)\left(2\left(\mu_1-\mu_2\right)' A_2\mu_2\right) \\
= \;& \frac{\beta_2}{2\gamma}\cdot\frac{T_1^2}{T^2}\cdot\left[\frac{\left(N-1\right)T}{T_1T_2} - \frac{1}{T_1}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right) - \left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right)\right] \\
& - \frac{1}{2\gamma}\cdot\frac{T_1}{T}\left(\beta_2-\beta_1\right)\cdot\left(\mu_1' A_2\mu_1 - \mu_2' A_2\mu_2 - \left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right)\right) \\
= \;& \frac{1}{2\gamma}\left(\frac{T}{T_1}\beta_1 - \frac{T_2}{T_1}\beta_2\right)\cdot\frac{T_1^2}{T^2}\cdot\left[\frac{N-1}{T_1} + \frac{N-1}{T_2} - \frac{1}{T_1}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right) - \left(\mu_1-\mu_2\right)' A_2\left(\mu_1-\mu_2\right)\right] \\
& + \frac{1}{2\gamma}\frac{T_1}{T}\left(\beta_2-\beta_1\right)\left(\frac{N-1}{T_1} + \frac{N-1}{T_2} - \left(\mu_1' A_2\mu_1 - \mu_2' A_2\mu_2\right) - \frac{1}{T_1}tr\left(A_2\left[\Sigma_1-\Sigma_2\right]\right)\right)
\end{aligned}$$

Proof 3.7.5 (*Proposition 3.3.5*)**.** *Because it is assumed that $T_1 = T_2$ and $\Sigma_1 = \Sigma_2$,* *conditional on $\hat{\mu}$, the expected CE of $\hat{w}^c_{tp}$ can be obtained by:*

$$\begin{aligned}
& \mathrm{E}\left[CE\left(\hat{w}^c_{tp}\right)\middle|\,\hat{\mu}\right] \\
= \; & \mathrm{E}\left[\mu_2'\hat{w}^c_{tp} - \frac{\gamma}{2}\hat{w}^{c\prime}_{tp}\Sigma_2^{-1}w^c_{tp}\middle|\,\hat{\mu}\right] \\
= \; & \frac{\alpha_1}{\gamma}\hat{\mu}'\Sigma_2^{-1}\mu - \frac{1}{2\gamma}\left((c^2+(1-c)^2)\alpha_2 - 2c(1-c)\alpha_1^2\right)\hat{\mu}'\Sigma_2^{-1}\hat{\mu} \\
= \; & \frac{\alpha_1}{\gamma}\hat{\mu}'\Sigma^{-1}\mu - \frac{1}{2\gamma}\left(\alpha_1^2 + \left(c^2+(1-c)^2\right)(\alpha_2-\alpha_1^2)\right)\hat{\mu}'\Sigma^{-1}\hat{\mu}.
\end{aligned}$$

It is shown that, conditional on $\hat{\mu}$, the expected CE of $w_{tp}(\hat{\mu}, S_{T_2})$ is:

$$E\left[CE\left(w_{tp}(\hat{\mu}, S_{T_2})\right)|\hat{\mu}\right] = \frac{\alpha_1}{\gamma}\hat{\mu}'\Sigma^{-1}\mu - \frac{\alpha_2}{2\gamma}\hat{\mu}'\Sigma^{-1}\hat{\mu}.$$

Therefore, the expected CE difference between the combined portfolio $\hat{w}^c_{tp}$ and the uncombined portfolio $w_{tp}(\hat{\mu}, S_{T_2})$ is

$$\begin{aligned}
& E\left[CE\left(w^c_{tp}\right) - CE\left(w_{tp}(\hat{\mu}, S_{T_2})\right)\right] \\
= \; & \frac{1}{2\gamma}\left(\alpha_2 - \alpha_1^2 - \left(c^2+(1-c)^2\right)(\alpha_2-\alpha_1^2)\right)\mathrm{E}\left[\hat{\mu}'\Sigma^{-1}\hat{\mu}\right] \\
= \; & \frac{c(1-c)}{\gamma}\left(\alpha_2-\alpha_1^2\right)\mathrm{E}\left[\hat{\mu}'\Sigma^{-1}\hat{\mu}\right] \\
> \; & 0.
\end{aligned}$$

In the case of efficient portfolios, It can be easily shown that:

$$\begin{aligned}
\mathrm{E}\left[\hat{w}^c_{ep}\middle|\,\hat{\mu}\right] &= \mathrm{E}\left[w_{ep}(\hat{\mu}, S_{T_2})\middle|\,\hat{\mu}\right] \\
\mathrm{V}\left[\hat{w}^c_{ep}\middle|\,\hat{\mu}\right] &= \left(c^2+(1-c)^2\right)\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})\middle|\,\hat{\mu}\right]
\end{aligned}$$

Therefore, using the risk function (3.2.3), we have

$$\begin{aligned}
& E\left[CE\left(\hat{w}^c_{ep}\right) - CE\left(w_{ep}(\hat{\mu}, S_{T_2})\right)\right] \\
= \; & \mathcal{R}\left(w_{ep}(\hat{\mu}, S_{T_2})|w^*_{ep}\right) - \mathcal{R}\left(\hat{w}^c_{ep}|w^*_{ep}\right) \\
= \; & \frac{\gamma}{2}tr\left(\Sigma_2\left[\mathrm{E}\left[\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right]\right] + \mathrm{V}\left[\mathrm{E}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right]\right] + Bias^2\left(w_{ep}(\hat{\mu}, S_{T_2})\right)\right]\right) \\
& -\frac{\gamma}{2}tr\left(\Sigma_2\left[\mathrm{E}\left[\mathrm{V}\left[\hat{w}^c_{ep}\middle|\,\hat{\mu}\right]\right] + \mathrm{V}\left[\mathrm{E}\left[\hat{w}^c_{ep}\middle|\,\hat{\mu}\right]\right] + Bias^2\left(\hat{w}^c_{ep}\right)\right]\right) \\
= \; & \frac{\gamma}{2}tr\left(\Sigma_2\left[\left(1-c^2-(1-c)^2\right)\mathrm{E}\left[\mathrm{V}\left[w_{ep}(\hat{\mu}, S_{T_2})|\,\hat{\mu}\right]\right]\right]\right) \\
= \; & c\cdot(1-c)\cdot\left(\gamma\cdot\frac{N-1}{T_2-N-1}\sigma^2_{gmv} + \frac{1}{\gamma}\left(\beta_2-\beta_1^2\right)\mathrm{E}\left[\hat{\mu}'A_2\hat{\mu}\right]\right) \\
\geq \; & 0
\end{aligned}$$

Complete Bibliography

Antoine, B. (2012), 'Portfolio selection with estimation risk: A test-based approach', *Journal of Financial Econometrics* **10**(1), 164–197.

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

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

Britten-Jones, M. (1999), 'The sampling error in estimates of mean-variance efficient portfolio weights', *The Journal of Finance* **54**(2), 655–671.

Brodie, J., Daubechies, I., Molc, C. D., Giannone, D. & Loris, I. (2009), 'Sparse and stable markowitz portfolios', *PNAS* **106**(30), 12267–12272.

Cho, D. D. (2010), Estimation risk in covariance. Journal of Asset Management, (forthcoming).

Chopra, V. K. & Ziemba, W. T. (1993), 'The effect of errors in means, variances, and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**(2), 6–11.

DeMiguel, V., Garlappi, L., Nogales, F. J. & Uppal, R. (2009), 'A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms', *Management science* **55**(5), 798–812.

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

Engle, R. & Colacito, R. (2006), 'Testing and valuing dynamic correlations for asset allocation', *Journal of Business & Economic Statistics* **24**(2), 238–253.

Fan, J., Zhang, J. & Yu, K. (2012), 'Vast portfolio selection with gross-exposure constraints', *Journal of the American Statistical Association* **107**(498), 592–606.

Frahm, G. (2010), An analytical investigation of estimators for expected asset returns from the perspective of optimal asset allocation. University of Cologne, Seminar of Economic and Social Statistics, Discussion Paper 1/10.

Frahm, G. & Memmel, C. (2010), 'Dominating estimators for minimum-variance portfolios', *Journal of Econometrics* **159**(2), 289–302.

Frahm, G., Wickern, T. & Wiechers, C. (2012), 'Multiple tests for the performance of different investment strategies', *AStA Advances in Statistical Analysis* **96**(3), 343–383.

Frahm, G. & Wiechers, C. (2011), On the diversification of portfolios of risky assets. University of Cologne, Seminar of Economic and Social Statistics, Discussion Paper 2/11.

Green, R. C. & Hollifield, B. (1992), 'When will mean-variance efficient portfolios be well diversified?', *Journal of Finance* **47**(5), 1785–1809.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

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

Kan, R. & Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.

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

Liu, H. & Pohlmeier, W. (2013), Risk preferences and estimation risk in portfolio choice. Working Paper.

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

Markowitz, H. M. (1987), *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell.

Merton, R. C. (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**(4), 323–361.

Mori, H. (2004), 'Finite sample properties of estimators for the optimal portfolio weight', *Journal of the Japan Statistical Society* **34**, 27–46.

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

Perron, P. (2006), Dealing with structural breaks, *in* K.Patterson & T. C.Mills, eds, 'Palgrave Handbook of Econometrics', Vol. 1, London : Palgrave Macmillan.

Pesaran, M. H. & Timmermann, A. (2005), 'Small sample properties of forecasts from autoregressive models under structural breaks', *Journal of Econometrics* **129**, 183–217.

Pesaran, M. H. & Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema

Three Essays on Robust Optimization of Efficient Portfolios

ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt.[16] Die Arbeit wurde bisher weder im In- noch Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, den 15. Mai 2013

(Hao Liu)

[16]Siehe hierzu die Abgrenzung zu Kapiteln 1 auf der folgenden Seite.

Abgrenzung

Kapitel 1 entstammt einer gemeinsamen Arbeit mit Herrn Prof. Dr. Winfried Pohlmeier (Universität Konstanz). Meine individuelle Leistung bei der Erstellung dieser Arbeit beträgt 80%.

Ich versichere hiermit, dass ich Kapitel 2 und Kapitel 3 der vorliegenden Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.