technische universität
dortmund

Essays in Finance: Aggregating Distributions and Detecting Structural Breaks

Dissertation zur Erlangung des akademischen Grades
Doctor rerum politicarum
der Fakultät für Wirtschaftswissenschaften
der Technischen Universität Dortmund

Daniel Ullmann

October 16, 2017

Contents

List of Figures

List of Tables

Acknowledgment

I would first like to thank Professor Peter N. Posch. His door was always open, whenever I ran into a trouble spot or had other questions. He steered me in the right direction, whenever it was necessary.

I would also like to thank the whole research group (Johannes, Janis, Timotheos, Philipp, Vladislav and Gerrit) for their patience, criticism and support during the whole time as a Ph.D. student, as well as our assistants Anton, Marius and Nils.

Finally, I would like to thank my family who supported me in many aspects and smoothed the way I have taken. Most of all, I thank my partner for life Helene for her unconditional support during all the ups and especially during all the downs of this thesis.

Thank you very much.

1 Introduction

Covariance and correlation are a central topic in finance. One example is the risk correction in asset prices. The basic pricing equation from the consumption-based model in asset pricing (Cochrane 2000):

$$p_t = \mathbb{E}_t \left(\beta \frac{u'(c_{t+1})}{u'(c_t)} \cdot x_{t+1} \right) \tag{1.1}$$

defines the price p_t of an asset with future payoff x_{t+1} in a two period model, which maximizes an investor's utility. The optimization problem is the choice of consuming today or investing today and additionally consuming the payoff of the asset tomorrow. The stochastic discount factor m_{t+1} is the product of the subjective discount factor β and the marginal rate of inter-temporal consumption utility given the investor's utility function u: One can show that prices of assets are a function of the covariance of the stochastic discount factor and the asset's payoff (Cochrane 2000, p. 22ff). Consequently their volatility determines the magnitude of the price correction, while the correlation determines the sign. Assets with a positive correlation with the stochastic discount factor show a higher price and vice versa. Turing from the stochastic discount factor to the consumption itself, this effects switches the sign, due to the (assumed) concavity of the utility function u. Assets with a negative correlation with future consumption show a higher price and vice versa:

$$p_t = \frac{\mathbb{E}(x_{t+1})}{R_f} + \frac{cov\left[\beta u'(c_{t+1}), x_{t+1}\right]}{u'(c_t)} \tag{1.2}$$

where R_f denotes the risk free gross return. Formulated in terms of gross returns R, one can construct an upper bound for the absolute value of any assets excess return:

$$\frac{\left|\mathbb{E}\left(R_{t+1}\right) - R_f\right|}{\sigma_{R_{t+1}}} \leq \frac{\sigma_{m_{t+1}}}{\mathbb{E}\left(m_{t+1}\right)} \tag{1.3}$$

This inequality is due to the bounded correlation coefficient. It defines the slope of the mean-standard deviation frontier in which all single assets and all possible portfolios lie

In practice one has to deal with estimates of both the mean and the covariance. Of course, one would like to have a precise, well-conditioned estimate. In the case of the covariance matrix an invertible matrix is highly desirable. An invertible covariance matrix is a necessity for the analytical solution of the minimization problem from above (cf. Cochrane 2000, p. 81ff). For the covariance's sample estimator to be nonsingular/invertible more observations than dimensions of the covariance matrix are needed (Ledoit and Wolf 2004a). There exist two main approaches to deal with this problem, assuming factor models for the returns or using shrinking methods. In chapter 2 examines if improvements in covariance estimations translate to correlation estimations. Here, the differences between the estimated and true matrices are measured and statistically analyzed, using a Monte Carlo Simulation. Several data-generating distributions and underlying covariance matrices are implemented to determine the magnitude and significance of non-normally distributed returns upon estimators using exactly the normality assumption and distribution free estimators. The results show that the improvements in covariance estimation precision translate to correlation matrices estimation precision in general. The second goal of this chapter is to add economic considerations. Does an improvement in estimation precision lead to a statistical out-performance of portfolios based on this estimate? Chapter 2 uses the covariance estimates to determine the corresponding minimum-variance portfolios. Due to a better estimation, Sharpe ratios can be increased by a factor of about two, which corresponds to a volatility reduction of a factor of about two. The results are robust when changing the data-generating distribution and the underlying covariance type respectively.

In all these considerations a constant covariance/correlation matrix is assumed. If this is not the case, the time series show a structural break within their relational structure. There exist several tests for structural breaks, starting from the (Chow 1960) test for time series themselves. In brief, the test divides the sample at the a priori known break and regressions in the subsamples are tested to be statistically different. Chapter 3 deals with a nonparametric test for breaks in the correlation structure of returns. Using such a test allows identifying periods with a statistically constant correlation structure and thereby periods in which, for example, portfolio optimization is valid. The crux in such tests is the need of many observations relative to the number of assets in the observed portfolio. Chapter 3 proposes to reduce the dimension of the portfolio and test the 'reduced version' of the portfolio for structural breaks. The Principal Component Analysis and the Hierarchical Clustering techniques are introduced and the validity of the limiting distribution, power and size are discussed using Monte Carlo Simulations. The aptitude of this two stage process is demonstrated by two empirical cases. In the first case, the scenario described in Wied (2015) is extended to a longer time horizon and the full portfolio of the EuroStoxx-50 and comparable break dates to the original test are shown. In the second case, the NASDAQ-100 index is analyzed to detect multiple break dates using the algorithm proposed in Galeano and Wied (2014) and find matching time-points for breaks in several reduced dimensions.

Thus far, this thesis looks at methods for improved estimations in the mean-squared error sense and the detection of structural breaks in the estimated aggregate of a high dimensional distribution. While the latter takes the aggregate for granted, the former tries to minimize the estimation error, while keeping the aggregating function mostly unaffected. The second part of the thesis is devoted to the aggregating process itself. Following the concept of entropy shifts (Bowden 2016a) the starting points are conditional moments, unlike the mean or variance of a set of observations. An internal point of view is taken where one asks the question: am I in a better situation than the others in the sample?

Chapter 4 answers this question in the context of incomes and thereby examines the inequality associated with the income distribution. The estimate of the complete distribution is based on data of the EU-SILC database, which enables a granularity on the level of families. Different from Eltetö and Frigyes (1968) not a single 'representative agent' (using the mean/median of the observations) is analyzed, but all possible reference points are taken into account. A relative advantage/asymmetry metric is created, which shows a duality property: a change of sign results in a metric for spread. Phase planes, based on these two metrics, visualize the trajectories of European countries from 2005 to 2013 and enable a grouping into three categories of inequality risers, fallers and a mixed group. As a result, the ambiguity of the Gini coefficient in terms of asymmetry and spread is resolved. In a regression analysis chapter 4 analyzes the effects of variables, stated to be a determinant of the Gini index, on the metrics proposed here. Since we find a strong correlation between the Gini and the d metric, one can transfer the results of the latter to the Gini and analyze European effects on inequality based on rather new and high resolution data.

Structural similarities of the standardized dispersion metric and the Sharpe ratio were the starting point for the final research question of this thesis: How does a performance metric based on the internal view of a fund manager look like and would fund managers be ranked differently when using different performance metrics? Chapter 5 proposes a performance measure based on the metrics of the previous chapter. It is analyzed how the asymmetry of the return distributions affects the metric's sign and magnitude compared to the Sharpe ratio. Afterwards the rank correlations with other performance metrics are shown based on four different asset classes. Stocks, bonds and foreign exchange rates are used and as a portfolio of many different asset classes hedge funds. High and statistically significant correlations for the other metrics are found (c.f Eling and Schuhmacher (2007)). The ones proposed here do not show these high correlations, but are nearly uncorrelated. The conclusion is, that these aggregate the distributions in such a manner that new information is created.

Summarizing, this thesis focuses on the aggregating process of high dimensional distributions and the detection of structural breaks within those. The latter is addressed in chapter 3, where an extension to higher dimensional portfolios is shown. The former is the topic of chapter 4 and chapter 5 and indirect in chapter 2, since not the aggregating process itself is the topic but rather an a posteriori improvement of the estimation.

1.1 Publication details

Paper I (chapter 2):
ESTIMATION OF LARGE CORRELATION MATRICES USING SHRINKING METHODS

Authors:
Peter N. Posch, Daniel Ullmann

Abstract:
Quantitative estimations are subject to estimation errors, due to limited observations. One way of minimizing them is the shrinking approach. This paper studies shrinking estimators with and without the normality assumption of returns together with several data-generating distributions and different types of underlying covariance matrices. The transferability of covariance estimation improvements to the correlation matrix and to economic improvements is analyzed. Significant out-performance of some shrinking estimators in the Sharpe ratio and the volatility of minimum-variance portfolios is found, which seem to depend weakly on the normality assumption. Using a scaled identity matrix as shrinking target, the Sharpe ratio increases by a factor of about two.

Publication details:
Working paper.

Paper II (chapter 3):

Detecting Structural Breaks in large Portfolios

Authors:

Peter N. Posch, Daniel Ullmann, Dominik Wied

Abstract:

Model free tests for constant parameters often fail to detect structural changes in high dimensions. In practice this corresponds to a portfolio with many assets and a reasonable long time series. We reduce the dimensionality of the problem by looking at a compressed panel of time series obtained by cluster analysis and the principal components of the data. With this procedure we can extend tests for constant correlation matrix from a sub portfolio to whole indices, which we exemplify using a major stock index.

Publication details:

Revised and Resubmitted to Empirical Economics

Paper III (chapter 4):
INCOME DISTRIBUTION IN TROUBLED TIMES: DISADVANTAGE AND DISPERSION DYNAMICS IN EUROPE 2005-2013

Authors:
Roger J. Bowden, Peter N. Posch, Daniel Ullmann

Abstract:
Income distribution has been a longstanding focus of social and economic interest, but never more so than in recent times. New metrics for disadvantage and spread enable a more precise differentiation of directional asymmetry and dispersion, drawing on an internal contextual perspective. The dual metrics for asymmetry and spread can be plotted over time into a phase plane, enabling comparative social welfare perspectives over time and between countries. The methods are utilized to study the dramatic changes that took place in Europe prior to and after the global financial crisis. Major differences are revealed. In terms of asymmetry and spread, some countries have been fallers (lower in both) while other countries are risers.

Publication details:
Finance Research Letters, Forthcoming 2017

Paper IV (chapter 5):
Asymmetry and performance metrics for financial returns

Authors:
Roger J. Bowden, Peter N. Posch, Daniel Ullmann

Abstract:
The assumption of symmetric asset returns together with globally risk averse utility functions is unappealing for fund managers and other activist investors, whose preferences switch between risk aversion on the downside and risk seeking on the upside. A performance return criterion is originated that is more consistent with the implicit Friedman-Savage utility ordering. Adapted from recent developments in the income distribution literature, the metric w weights the lower versus upper conditional expected returns, while a dual spread or dispersion metric d also exists. The resulting performance metric is easy to compute. A point of departure is the conventional Sharpe performance ratio, with the empirical comparisons extending to a range of existing performance criteria. In contrast to existing metrics, the proposed performance metric W results in different and more embracing rankings.

Publication details:
Working paper

2 Estimation of Large Correlation Matrices using Shrinking Methods

The following is based on Posch and Ullmann (2017).

For many purposes an accurate estimation of a covariance or a correlation matrix given historical returns is of high interest. One example is the Markowitz (1952) portfolio optimization. Here one seeks to find assets which negatively correlate to the existing portfolio to gain a diversification effect. Especially when testing for the validity of the optimization, the correlation matrix becomes the center of interest and tests for a constant correlation matrix outperform the ones for a constant covariance matrix, cf. Wied, Kraemer, and Dehling (2012).

Formally, one estimates a covariance matrix Σ from N given time series of length T, which is equivalent to looking at T N-dimensional random vectors X_t. The sample/empirical covariance matrix estimator is then calculated as:

$$S = \frac{1}{T-1} \sum_{t=1}^{T} \left(X_t - \bar{X}\right) \left(X_t - \bar{X}\right)', \tag{2.1}$$

where $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t \in \mathbb{R}^N$ denotes the vector of first moments.

Of course this estimator is only appropriate if the sample is representative, which means that no structural break did occur; cf. Posch, Ullmann, and Wied (2017) for testing for structural breaks in high dimensions. Furthermore this estimator is limited to situations where the number of observations is sufficiently large ($N \ll T$). In other situations, the estimated matrix is no longer positive definite (Kwan 2010), nor is it invertible (Ledoit and Wolf 2004a), thus only usable to a limited extent in the context of portfolio optimization. One could argue that this is only a theoretical problem as in practice stocks are traded daily and sufficient data should be available. But even portfolios of a single asset class, like stocks, are affected by this problem as the number of positions increase. For example fully replicating the S&P 500 index leads to $N \approx 500$ and thus to a need of over two years of daily data for all constituents without any structural break present. When looking at multi-asset class portfolios, the joint information available for all assets is typically much shorter. Consider, for example, a simple portfolio with an interest rate swap for which market values are calculated only from inception or a future which rolls every three months. Apart from the problem of sufficient data availability it is desirable that the estimator is well conditioned, i.e. the estimation errors do not amplify by inverting the estimated matrix, which is not the case for the sample covariance estimator in equation (2.1) (Ledoit and Wolf 2004b).

In principle there exist two main approaches to this problem: The first is to impose some structure on the covariance matrix and use a factor model (Fan, Fan, and Lv 2008). The second is to use a method called shrinkage (Stigler 1990), which is an application of the Stein paradox (Efron and Morris 1977). The idea is that the sample covariance matrix S is transformed via a linear combination with another, well-conditioned target matrix T_M

$$\Sigma^* = (1-\lambda)S + \lambda T_M, \tag{2.2}$$

where well-conditioned is defined as follows: let A be a regular matrix and $\|\cdot\|$ a matrix-norm. The condition of the matrix is then defined as

$$cond(A) := \|A\| \left\|A^{-1}\right\| \in (1, \infty), \tag{2.3}$$

where $cond(A) \approx 1$ corresponds to a well-conditioned matrix and ill-conditioned else (Bronstein 2012; Touloumis 2015). The weight is chosen by minimizing the estimation error of the shrinking matrix and the population matrix, measured by the mean-squared errors. As a result extreme values shrink towards more central values and thus the mean-squared error becomes smaller (Ledoit and Wolf 2004b).

Contrary to the standard procedure in the literature, where only the out-performance over other estimators is shown, this analysis considers three different quantities, based on the covariance matrix, additionally. We assess the overall performance of the different estimators for different purposes, in the sense of a 'one for all' solution. Two types of estimators for the covariance are used, which are derived under different assumptions.

Not only the distance to the true covariance matrix is measured, but also the distance of the resulting correlation matrix, defined as:

$$\rho = (\rho_{i,j}) \in \mathbb{R}^{N \times N} \quad \text{with} \quad \rho_{i,j} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} \tag{2.4}$$

Both quantities, the correlation and the variances define the covariance. So it is not obvious, if a more precise estimation of the covariance is due to a more precise estimate of the correlation or a more precise estimation of the variances.

A second point of view is the economic significance of the findings from above. Does the rather abstract smaller distance to the true matrix result in better investment decisions or do these benefits get lost during the information processing process? Minimum-variance portfolios, based on the covariance estimations are analyzed concerning their Sharpe ratio and annualized volatilities.

The scheme of the paper is as follows. The assumptions under which the different shrinking estimators are derived are described in section 2.1. Afterwards, we present our simulation environment, which is adapted from Touloumis (2015). Section 2.3 shows and discusses the results of our simulation study and the final section concludes.

2.1 Shrinking estimators

Starting from equation (2.2), there exist several possibilities for the target matrix. The most common ones are the identity matrix I, the scaled identity matrix $\mu^* I$ (μ^* denotes the average of the eigenvalues of the covariance matrix) and the diagonal matrix $D_\Sigma = diag(\Sigma)$ (Chen, Wiesel, and Hero 2009; Ledoit and Wolf 2004a).

Using the target matrix $\mu^* I$, Ledoit and Wolf (2004a) derive the optimal shrinking intensity λ as

$$\lambda = \frac{\mathbb{E}\left[\|S - \lambda\|^2\right]}{\mathbb{E}\left[\|S - \mu I\|^2\right]}. \tag{2.5}$$

Both Fisher and Sun (2011) and Touloumis (2015) use this result and propose estimators for the unknown quantities. They extend the possible target matrices from only the scaled identity matrix to the sole identity matrix I and the diagonal matrix of the empirical covariance matrix $D_S = \operatorname{diag}(S)$. To do so, Fisher and Sun (2011) assumes normally distributed returns and the following two assumptions:

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \Sigma^i < \infty \quad \forall i \in 1, ..., 4 \qquad \text{and} \qquad T-1 = O(N^k) \quad \text{for} \quad 0 < k \leq 1$$

Using the results of Srivastava (2005), he derives the optimal shrinking intensities $\lambda_\mu, \lambda_I, \lambda_D$ (c.f. equation (2.2)) as:

$$\lambda_\mu^F = \frac{\hat{a}_2 + N\hat{a}_1^2}{T\hat{a}_2 + (N-T+1)\hat{a}_1^2} \tag{2.6}$$

$$\lambda_I^F = \frac{\frac{1}{T-1}\hat{a}_2 + \frac{N}{T-1}\hat{a}_1^2}{\frac{T}{T-1}\hat{a}_2 + \frac{N}{T-1}\hat{a}_1^2 - 2\hat{a}_1 + 1} \tag{2.7}$$

$$\lambda_D^F = \frac{\hat{a}_2 + N\hat{a}_1^2 - 2a_2^*}{T\hat{a}_2 + N\hat{a}_1^2 - (T+1)\hat{a}_2^*} \tag{2.8}$$

with

$$\hat{a}_1 = \frac{\operatorname{tr} S}{N} = \nu \tag{2.9}$$

$$\hat{a}_2 = \frac{(T-1)^2}{T(T+1)} \frac{1}{N} \left[\operatorname{tr} S^2 - \frac{1}{N-1} \operatorname{tr}^2 S \right] \tag{2.10}$$

$$a_2^* = \frac{T-1}{T+1} \frac{\operatorname{tr} D_S^2}{N} \tag{2.11}$$

and tr$\cdot$ denoting the trace of the respective matrix.

Touloumis (2015) relaxes the assumption of normally distributed returns, to iid. returns. To restrict the dimension of the covariance matrix Σ he assumes

$$\lim_{T\to\infty} N \to \infty \qquad \text{and} \qquad \lim_{T\to\infty} \frac{\operatorname{tr} \Sigma^4}{\operatorname{tr}^2 \Sigma^2} \to t_1 \qquad \text{and} \qquad \lim_{T\to\infty} \frac{\operatorname{tr} \Sigma^2}{\operatorname{tr}^2 \Sigma} \to t_2$$

together with $0 \leq t_2 \leq t_1 \leq 1$.

While the first assumption here is equivalent to the second in Fisher and Sun (2011), the set of possible covariance matrices differs. In Fisher and Sun (2011) suppose finite arithmetic means of the eigenvalues for the first four moments, Touloumis (2015) restricts this set by assumptions two and three.

The shrinking intensities for the three different shrinking targets are given by:

$$\lambda_{\mu}^{T} = \frac{\hat{b}_2 + \hat{b}_1^2}{T\hat{b}_2 + \frac{N-T+1}{N}\hat{b}_1^2} \tag{2.12}$$

$$\lambda_{I}^{T} = \frac{\hat{b}_2 + \hat{b}_1^2}{T\hat{b}_2 + \hat{b}_1^2 - (T-1)(2\hat{b}_1 - N)} \tag{2.13}$$

$$\lambda_{D}^{T} = \frac{\hat{b}_2 + \hat{b}_1^2 - 2\hat{b}_3}{T\hat{b}_2 + \hat{b}_1^2 - (T-1)\hat{b}_3} \tag{2.14}$$

Using results from Himeno and Yamada (2014), $\hat{b}_1, \hat{b}_2$ and $\hat{b}_3$ can be expressed as follows:

$$\hat{b}_1 = N\hat{a}_1 = N \cdot \nu = \operatorname{tr} S \tag{2.15}$$

$$\hat{b}_2 = \frac{T-1}{T(T-2)(T-3)} \left[(T-1)(T-2)\operatorname{tr} S^2 + \operatorname{tr}^2 S - TQ\right] \tag{2.16}$$

$$\begin{aligned}
\hat{b}_3 = {} & \frac{1}{P_2^N} \sum_{a=1}^{N} \sum_{i \neq j}^{*} X_{ia}^2 X_{ja}^2 \\
& - 4\frac{1}{P_3^N} \left\{ \sum_{a=1}^{N} \left(\sum_{i=1}^{T} X_{ia}^2 \right) \left(\sum_{i=1}^{T-1} \sum_{j=i+1}^{T} X_{ia} X_{ja} \right) - \sum_{a=1}^{N} \sum_{i \neq j}^{*} X_{ia}^3 X_{ja} \right\} \\
& + \frac{2}{P_4^N} \left\{ 2 \sum_{a=1}^{N} \left(\sum_{i=1}^{T-1} \sum_{j=i+1}^{T} X_{ia} X_{ja} \right)^2 - \sum_{a=1}^{N} \sum_{i \neq j}^{*} X_{ia}^2 X_{ja}^2 \right\}
\end{aligned} \tag{2.17}$$

with $\sum^*$ denoting summation over mutually distinct indices. The quantities Q and P_t^s are defined as:

$$Q = \frac{1}{T-1} \sum_{t=1}^{T} \left[\left(X_t - \bar{X}\right)' \left(X_t - \bar{X}\right) \right]^2 \tag{2.18}$$

$$P_t^s = s!/(s-t)! \tag{2.19}$$

Touloumis (2015) shows that both estimator groups, the ones with the normality assumption of Fisher and Sun (2011) and his own, not using the normality assumption lead to comparable shrinking intensities, thus comparable estimates of the covariance matrix. He proposes a data-driven selection of the shrinking target: First one compares the shrinking intensities for all three targets and chooses the target matrix with the largest proposed intensity.

In the following section, we present our Monte Carlo simulation environment: the data-generating processes and different covariance types used together with the overall procedure.

2.2 Simulation environment

Our Monte Carlo simulation is similar to Touloumis (2015). We increase the number of simulations by a factor of 5 to 5,000. To generate random correlated numbers for a given covariance matrix Σ we:

1. Calculate the Cholesky Decomposition C, i.e. $CC' = \Sigma$.
2. Generate N random time-series X_i of length T where all entries are iid. with zero mean and unit variance, e.g. $x_{i,t} \sim \mathcal{N}(0,1)\ \forall i \in 1,...,N, t \in 1,T$.
3. Transforming the time-series via $r_{i,t} = \mu_i + CX_i := \mu_i + R_i$ with a given mean μ_i leads to the desired covariance structure:

$$\operatorname{Var}(R) = \mathbb{E}\left(CX\,(CX)'\right) = \mathbb{E}\left(CXX'C'\right) = \mathbb{E}\left(CC'\right) = \Sigma. \tag{2.20}$$

The S&P500 index from 2007 to 2017 is used for the estimation of this true covariance matrix and the average return, where due to data availability only 400 constituents are included. In the following analyses 1, 2 and 3 years of financial returns are mimicked through 250, 500 and 750 returns per stock.

Then resulting panel is used to estimate the quantities of interest: the covariance matrix, the correlation matrix and the weights of the minimum-variance portfolio. The difference of the covariance and the correlation matrix is measured using the Frobenius norm, which is defined as follows:

$$\|A\|_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} |a_{i,j}|^2}. \tag{2.21}$$

The weights of the minimum-variance portfolio are defined as the weights $w \in \mathbb{R}^N$, which result in in minimal variance of the portfolio. With the covariance matrix Σ_P of the portfolio's assets, the portfolio variance is defined as:

$$\sigma_{PF}^2 = w' \Sigma_P w. \tag{2.22}$$

The different covariance estimations result in different weights and those lead to different Sharpe ratios and different annualized volatilities.

For the true covariance matrix, three different approaches are used. The first one uses random eigenvalues $\lambda_i \in [1, 10]$, together with a randomly generated orthogonal matrix Ψ. The construction of the covariance matrix follows:

$$\Sigma = \Psi \cdot \operatorname{diag}(\lambda_1, \ldots, \lambda_N) \cdot \Psi'. \tag{2.23}$$

This procedure yields a positive definite matrix, when choosing the boundaries for the eigenvalues appropriately. The correlation matrix is calculated by dividing the matrix by the square root of the corresponding diagonal elements. The second approach uses a AR(1) model with the true covariance matrix defined as

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \ldots & \rho^{N-1} \\ \rho & 1 & \ldots & \rho^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \ldots & 1 \end{pmatrix}, \tag{2.24}$$

with $\rho \in [-1, 1]$ and a common variance σ^2 (Fisher and Sun 2011). It can be shown, that this definition of a covariance matrix results in a positive definite matrix as well (Pourahmadi 2013). Our last approach is to use the empirically estimated covariance matrix, described at the beginning of this section. We pretend that the 10 years of historical data are a all realizations from a constant covariance matrix. So the empirical estimate yields a positive definite matrix as well.

For the assets returns, we use a normal distribution, a t-distribution with 5 degrees of freedom and a gamma-distribution with parameters 4 and 0.5. Whereas the first two are well known and often used in modeling asset returns, the latter is used to ensure comparability with the results of Touloumis (2015).

2.3 Simulation results

Our first quantity of interest is the covariance matrix. Table 2.1 shows the scaled Frobenius norm of the different differences from estimators to the true matrix. For clarity, the results are scaled by a factor 10,000. We group by the data by generating process vertically into the 'normal', 't' and 'gamma' columns, for returns generated by the corresponding distribution. These columns are separated by the length of the time series.

The horizontal grouping in table 2.1 is based on the different estimators. The first three rows correspond to the empirical estimation using equation (2.1), where the type of underlying covariance matrix ('AR' for the covariance matrix created by the AR-model (2.24), 'eigen' for the matrix created by random eigenvalues (2.23) and 'emp' for the empirically estimated covariance matrix) is denoted in the column 'type' (second from the left). The middle part of the table correspond to the estimators in Fisher and Sun (2011), first grouped by the covariance type, like in the upper part and grouped by the shrinking target (most left column named 'target'). The lower part corresponds to the estimators in Touloumis (2015) with the same grouping as before. T-statistics are not reported since all differences are highly significant different from zero.

The distance becomes smaller, the longer the time series is, independent of the estimator, the data generating distribution and the covariance type. The differences over different time horizons are statistically significant on the 0.1% level. For the identity matrix as shrinking target, the differences between the Fisher and Touloumis estimator are not significant on a 10% level and they do not show a significant improvement over the empirical estimator.

In the case of Fisher's estimator using the scaled identity matrix, we see a general improvement over the empirical estimator of $11.6 - 13.5\%$ and for the diagonal matrix an improvement of $83.9 - 87.3\%$, which is comparable to the Touloumis estimators. In their case, using the scaled identity matrix the improvement is even higher with $16.8 - 34.5\%$. All those are statistically significant on a 0.1%, as well as the differences for the diagonal shrinking target between Fisher's and Touloumis' estimators. The differences between different data generating processes are not significant on a 10% level, like in Touloumis (2015), as well as differences for different covariance models.

Conversely, in the following analyses, differences between estimators within the data generating processes or the covariance models are found to be significant on a 0.1% level if not else explicitly stated. So a compressed version of the results is shown in table 2.2 and the residual extended tables can be found in tables A.1 to A.3 in the appendix.

Table 2.1: Distances to the true covariance matrix. Results are grouped horizontally by the three data generating distributions and the number of observations. The vertical grouping is based on the estimator used and the underlying covariance type. For clarity, the differences are scaled by a factor of 10,000.

		250	500	750	250	500	750	250	500	750
		normal			t			gamma		
target	type				Empirical					
	AR	87.919	62.223	50.834	87.911	62.226	50.826	87.908	62.220	50.824
	eigen	87.936	62.202	50.859	87.929	62.206	50.826	87.932	62.201	50.851
	emp	87.915	62.211	50.836	87.906	62.195	50.839	87.909	62.190	50.839
					Fisher					
	AR	76.042	55.007	44.169	75.969	54.980	44.171	76.030	54.950	44.207
μ	eigen	76.056	54.982	44.211	76.068	54.974	44.205	76.044	55.004	44.174
	emp	75.980	55.003	44.150	75.986	55.000	44.217	76.074	54.973	44.157
	AR	87.919	62.223	50.834	87.911	62.226	50.826	87.908	62.220	50.824
I	eigen	87.936	62.202	50.859	87.929	62.206	50.826	87.932	62.201	50.851
	emp	87.915	62.211	50.836	87.906	62.195	50.839	87.909	62.190	50.839
	AR	11.111	9.017	8.170	11.028	9.008	8.193	11.049	9.000	8.165
D	eigen	11.059	9.024	8.190	11.039	8.990	8.163	11.046	8.997	8.203
	emp	11.062	8.998	8.186	11.046	9.013	8.159	11.041	9.001	8.139
					Touloumis					
	AR	57.565	48.200	42.300	57.520	48.214	42.291	57.547	48.197	42.314
μ	eigen	57.586	48.195	42.343	57.580	48.160	42.307	57.556	48.184	42.312
	emp	57.535	48.225	42.268	57.526	48.183	42.321	57.602	48.146	42.284
	AR	87.919	62.223	50.834	87.911	62.226	50.826	87.908	62.220	50.824
I	eigen	87.936	62.202	50.859	87.929	62.206	50.826	87.932	62.201	50.851
	emp	87.915	62.211	50.836	87.906	62.195	50.839	87.909	62.190	50.839
	AR	11.145	9.178	8.503	11.057	9.163	8.535	11.081	9.152	8.516
D	eigen	11.090	9.184	8.548	11.069	9.131	8.517	11.079	9.157	8.551
	emp	11.090	9.158	8.528	11.075	9.170	8.513	11.077	9.153	8.475

Note: For clarity, t-statistics are omitted. All differences are significant on the 0.1% level.

The left hand side of table 2.2 shows the distance to the true correlation matrix. Note that these distances are not scaled. A similar pattern to before is visible: The gained estimation precision using the identity matrix as target is the smallest of all shrinking target. It is found to be in the range of one-tenth of a percent and significant on the 0.1% level. Fisher's and Touloumis' estimator are statistically not distinguishable from one another on a 10% level.

The improvements using the diagonal matrix are in the range of $93.9 - 89.3\%$ where the improvement is higher for less observations. This observation holds for Touloumis' estimator using the scaled identity matrix (with improvements in the range $62.6-37.9\%$) and Fisher's estimator (with improvements in the range $93.9-51.2\%$), while the absolute distance is increasing for an increasing number of observations for the latter.

These results indicate an out-performance of the diagonal shrinking target over all other estimators. Although we use non-normally distributed returns we find marginal gains by using the Touloumis estimator compared to Fisher's, where the assumption of normality has been abandoned.

Table 2.2: Distances to the true correlation matrix on the left hand side, Sharpe ratios on the in the middle and volatilities (given in percent) on the right hand side for minimum-variance portfolios resulting from the different covariance estimators. The result correspond to the case of normally distributed returns and an AR(1) covariance model.

	250	500	750	250	500	750	250	500	750
	Correlation			Sharpe Ratios			Annualized Volas / %		
				Empirical					
emp	25.339	17.922	14.636	-	0.288	0.473	-	2.833	1.732
				Fisher					
μ	1.545	5.308	7.139	0.578	0.575	0.618	1.406	1.408	1.314
I	25.317	17.914	14.632	0.389	0.291	0.473	2.098	2.808	1.731
D	1.545	1.556	1.543	0.578	0.551	0.579	1.406	1.471	1.402
				Touloumis					
μ	9.479	9.488	9.080	0.593	0.567	0.611	1.369	1.429	1.331
I	25.317	17.914	14.632	0.389	0.291	0.473	2.098	2.808	1.731
D	1.569	1.576	1.562	0.579	0.554	0.583	1.400	1.462	1.389

Note: All values are greater than zero on the 0.1% significance level.

In the last part of our analysis these abstract distance quantity are transformed to economically interpretable quantities: the Sharpe ratio and annualized volatilities, shown in the middle and right part of table 2.2 respectively.

Due to non-invertibility of the empirical estimator based on 250 observations, the corresponding cells are left blank (-). For the Sharpe ratio all shrinking estimators show improvements to the empirical. Except for 750 observations in the case of the scaled identity matrix, Touloumis' estimator yields higher Sharpe ratios than Fisher's. The Sharpe ratio increases by a factor of $1.91 - 1.99$ in the case of 500 and by a factor of $1.22 - 1.31$ for 750 observations. Within the two shrinking targets improvements are statistically significant on a 0.1% level but marginal in size with only a few percent. As before, Fisher's and Touloumis estimator using the identity matrix target are not distinguishable on a 10% level and for the case of 750 observations end up with the same Sharpe ratio as the empirical estimator. The other two shrinking targets outperform the identity matrix in all lengths of the given time series by a factor of $1.48 - 1.52$. An interesting pattern appears when switching from 250 to 500 observations. We find a slight decrease (statistically significant on a 0.1% level) in the Sharpe ratio, although the analyses above show a more precise estimation of the covariance/correlation matrix in these cases. Going from 500 to 750 observations the Sharpe ratio increases, as well as the precision of the covariance/correlation estimation. This indicates that the general tendency of a more precise estimation of the covariance matrix leading to higher Sharpe ratio of the resulting minimum-variance portfolio does not hold in all cases. The question remains if the higher Sharpe ratios are caused by higher portfolio returns or by lower portfolio volatilities.

To answer this question, table 2.2 shows the corresponding annualized volatilities of the portfolios on the right hand side. The shrinking estimators using the identity matrix are not distinguishable on a 10% level and lower the volatility by about 1 percent in the case of 500 observations. The difference to the empirical estimator is statistically significant on a 0.1% level, whereas the difference for 750 observations is not. For the scaled identity matrix, we find roughly half and three quarters the volatility for 500 and 750 observations respectively.

The diagonal shrinking target shows about the same improvement sizes. Except for Fisher's scaled identity and diagonal target, we find the differences between the shrinking targets to be significant on the 0.1% level and lower volatilities for the scaled identity matrix as shrinking target. Differences between Fisher's and Touloumis estimators for the last two shrinking targets are all significant on a 0.1% level. A similar pattern as in the case of Sharpe ratios can be seen when increasing the length of the time series. The portfolio volatility increases from 250 to 500 observations and decreases when using 750 observations, which explains the observation in the Sharpe ratios. This increase by a factor of about 2 comes with a decrease of volatilities by a factor of about 2. Thus we conclude the improvements in the Sharpe ratios are caused mostly by the effect of a lower portfolio variance and rather stable portfolio returns.

2.4 Conclusion

An accurate estimate of a covariance matrix is highly desirable for many purposes, e.g. the portfolio optimization. Especially in the case of fewer observations than dimensions of the covariance matrix, shrinking estimators are one tool to get an accurate estimate. This study uses Monte-Carlo simulations to assess the improvement in estimation accuracy of shrinking methods compared to the empirical estimator. Several data generating processes are used to determine the influence of non-normally distributed returns on estimator's estimation precision. The differences between estimators using the normality assumption and the ones not using this are marginal but significant. Differences within a shrinking estimator concerning the data-generating distribution are not found to be significant. Also different types of covariance matrices, used to correlate the generic returns are used and differences in the accuracy are found to be not significant as well.

The general tendency of an out-performance of the scaled identity matrix and the diagonal matrix over the sole identity matrix, independent of the quantity of interest is shown. The latter performed comparably to the empirical estimator and thus does not seem to fulfill its purpose in the cases analyzed here. Going further we show that improved covariance accuracy also translates to improved correlation estimation, when scaling the covariance matrix by the variances, given on the diagonal.

By using the Sharpe ratio and the annualized volatility of the different minimum-variance portfolios we analyze if the abstract improvement of a smaller distance between the matrices translates to economically interpretable improvement. Here the important fact of an invertible covariance matrix for fewer observations than dimensions comes to play, since for the empirical estimator the minimum-variance portfolio is undefined. It is shown that the Sharpe ratios of the minimum-variance portfolio based on the corresponding covariance estimation can increase by about 100% compared to the empirical estimator. The annualized volatilities of these portfolios show that the benefit in the Sharpe ratio is due to a smaller volatility.

3 Detecting Structural Breaks in large Portfolios

The following is based on Posch, Ullmann, and Wied (2017)

Portfolio optimization is most often based on the empirical moments of the portfolio constituents' returns, where the diversification effect is based on some measure of pairwise co-movement between the constituents, e.g. correlation. Whenever the characteristics of either the individual moments or the correlation changes, the portfolio's optimality is affected. Thus, it is important to test for the occurrence of changes in these parameters and there exist tests for detecting retrograde *structural breaks*, (see Andreou and Ghysels (2009) and Aue and Horváth (2013) for an overview). In the last few years, there is a growing interest in the literature for detecting breakpoints in dependence measures, especially for the case that the time of potential breaks need not be known for applying the test. People look, e.g., at the whole copula of different random variables (Bücher, Axel and Kojadinovic, Ivan and Rohmer, Tom and Segers, Johan 2014), but also at the usual bivariate correlation (Wied et al. 2012). The motivation for such approaches comes from empirical evidence that correlations are in general time-varying (see Longin and Solnik (1995) for a seminal paper on this topic), but it is unclear whether this is true for both conditional and unconditional correlations.

In this paper, we investigate in detail potential changes in correlation using the nonparametric fluctuation test for a constant correlation matrix proposed by Wied (2015). This test, as many others constructed in similar fashion, needs a high number of time observations relative to the number of assets for sufficient size and power properties. In practice, a typical multi asset portfolio has several hundreds of assets under management, but the joint time series for all assets is considerably smaller.

So how can we test for a structural break of the correlation structure of a portfolio where the number of assets is large and possibly larger than the number of observations? Our approach is to reduce the dimensionality and then applying the Wied (2015) test to the reduced problem. We consider two classical approaches to reduce the dimension of the problem, in order to check, if they can be used, or not. In the latter, more advanced methods like the Multi-way principal components analysis (Wold, Esbensen, and Geladi 1987) could be used.

First, we employ cluster analysis, second principal component analysis, cf. Fodor (2002) for a discussion of their use in dimensionality reduction. Cluster analysis has a wide range of applications such as biology (Eisen et al. 1998), medicine (Haldar et al. 2008) and finance, resp. econophysics (Bonanno et al. 2004; Brida and Risso 2010; Mantegna and Stanley 1999; Mantegna 1999; Tumminello, Lillo, and Mantegna 2010), where it is also applied in portfolio optimization (Onnela et al. 2002, 2003; Tola et al. 2008). Yet, as far as we know, we are the first to combine clustering and tests for structural breaks, which might be an interesting contribution to the existing literature on portfolio optimization. The principal component analysis (PCA) dates back to Pearson (1901) and is "one of the most useful and most popular techniques of multivariate analysis" (Hallin, Paindaveine, and Verdebout 2014) with wide applications in finance, cf. Greene (2008) or Alexander (1999). It is suitable for reducing the dimensionality of a problem (Fodor 2002), since its central idea is the transformation and dimension reduction of the data, while keeping as much variance as possible (Jolliffe 2002).

3.1 Methodology

In this section we develop the test approach to detect structural breaks in a large portfolio's correlation structure. The basis for our analysis is the correlation matrix:

$$\rho = (\rho_{i,j})_{i,j=1,\dots,n} \text{ where } \rho_{i,j} = \frac{E\big((x_i - \mu_i)(x_j - \mu_j)\big)}{\sqrt{\sigma_j^2 \sigma_i^2}} \tag{3.1}$$

with μ_i as the first moment and σ_i^2 as the variance of the corresponding time series. For the estimator $\hat{\rho}$ we use the empirical average $\hat{\mu}_i$ and the empirical variance $\hat{\sigma}_i^2$. Since the estimated correlation matrix has to be positive definite (Kwan 2010), we need more observations than assets. Looking at n assets with T observations, a sequence of random vectors $X_t = (X_{1,t}, X_{2,t}, \dots, X_{n,t})$, we alculate the correlation matrix from the first k observations according to equation (3.1) and denoted as $\hat{\rho}_k$.

We set the hypotheses as

$$H_0 : \rho_t = \rho_{t+\tau} \quad \forall \quad t = 1, ..., T; \tau = 1, \dots, T - t \tag{3.2}$$

$$H_1 : \exists t, \tau : \rho_t \neq \rho_{t+\tau} \tag{3.3}$$

and define the difference to the correlation matrix from all T observations as

$$\hat{P}_{k,T} = \text{vech}\left(\hat{\rho}_k - \hat{\rho}_T\right), \tag{3.4}$$

where the operator $\text{vech}(A)$ denotes the half-vectorization:

$$vech(A) = \left(a_{i,j}\right)_{1 \leq i < j \leq dim(A)} \tag{3.5}$$

The test statistic is given by Wied (2015) as

$$\hat{A}_T := \max_{2 \leq k \leq T} \frac{k}{\sqrt{T}} \left\| \hat{E}^{-\frac{1}{2}} \hat{P}_{k,T} \right\|_1, \tag{3.6}$$

where $\| \cdot \|_1$ denotes the L1 norm. The null is rejected if $\hat{A}_T$ exceeds the threshold given by the 95% quantile of A with the following definition:

$$A := \sup_{0 \leq s \leq 1} \left\| B^{\frac{n(n-1)}{2}}(s) \right\|_1 \tag{3.7}$$

Here $B^k(s)$ is the vector of k independent standard Brownian Bridges. If the test statistic exceeds this threshold, we define the structural break date k as the following time point:

$$\arg\max_k \frac{k}{\sqrt{T}} \left| \hat{P}_{k,T} \right| \tag{3.8}$$

For the limiting distribution of the statistic we need a 'scaling' matrix $\hat{E}$, which can be obtained by bootstrapping in the following way:

- We define a block length l and divide the data into $T - l - 1$ overlapping blocks:

$$B_1 = \left(X_1, \dots, X_l\right), B_2 = \left(X_2, \dots, X_{l+1}\right), \dots$$

- In each repetition $b \in [1, B]$ for some large B, we sample $\lfloor \frac{T}{l} \rfloor$ times with replacement one of the blocks and merge them all to a time series of dimension $l \cdot \lfloor \frac{T}{l} \rfloor \times n$.

- For each bootstrapped time series we calculate the covariance matrix. We convert the scaled elements above the diagonal to the vector $\hat{v}_b = \sqrt{T} \left(\hat{\rho}_{b,T}^{i,j} \right)_{1 \leq i < j \leq n}$

- We denote the covariance of all these vectors with $\hat{E} := Cov(\hat{v}_1, \ldots, \hat{v}_B)$

For the original test to provide a good approximation of the limiting distribution the ratio of observations to assets needs to be much larger than one. A portfolio with a large number of assets and insufficient time length of observation thus cannot be analyzed with the present test for structural changes in its correlation matrix.

To quantify this problem, we simulate for 4, 6 and 10 assets time series with length 200, 500, 1000 and 2000 observations. To assess the power and size of the test, we simulate in one case with a constant correlation matrix and in another with a break in the middle of the time series. For the latter, we choose randomly with replacement $n/2$ correlations $(\rho_{i,j}^*)$ in the given correlation matrix randomly to a new correlation in [-1,1]. We transform the resulting matrix into a positive definite matrix, and consequently the resulting matrix differs from the original one in all cases, but most prominently in the randomly selected $\rho_{i,j}^*$.

The result of the simulation study is shown in table 3.1. Thereby we use the critical values 4.47, resp. 9.13, resp. 23.21 for the number of assets being equal to 4, resp. 6, resp. 10. Our simulations runs 2500 times. We find comparable results as in Wied (2015) for the size distortions. It seems that the level converges to the 5% level. Concerning the power, we find an increase in the number of observations and a decrease with the number of assets. A very extreme decrease can be found for 500 observations and the shift from 6 to 10 assets. Whereas in the former case we find a rejection rate of about 72% it drops in the latter case to only about 20%. Such drops seem to exist for all cases, but are not that prominent. The lower rejection rate in the case of 4 assets can be explained in the random choice of pairwise correlations. Since we draw with replacement, the chance of a given pair is drawn twice considerably higher than for 6 and 10 dimensions and thus the size of the break in the correlation structure is smaller when compared to the number of assets.

Table 3.1: Empirical rejection rate (in percent) of H0 of the test in Wied (2015) based on 2500 simulations.

	number of assets		
	4	6	10
number of observations	$\Delta\rho^*_{i,j} = 0$		
200	4.60	2.16	0.80
500	4.33	3.29	0.75
1000	4.07	4.01	1.30
2000	5.44	4.11	4.20
number of observations	$\Delta\rho^*_{i,j} \neq 0$		
200	39.48	31.77	4.73
500	70.31	71.98	19.73
1000	72.84	84.22	75.93
2000	79.64	96.49	95.13

In the following, we analyze how the use of dimension reduction techniques before applying the test changes these findings. We discuss two techniques, clustering and principal component analysis, to reduce the dimensionality of the problem and apply the test afterwards. A first approach is to use exogenous clusters such as e.g. industry sectors which, however, imposes an exogenous structure possibly not present in the data. Clustering endogenously based on the present correlation structure instead preserves this information (cf. Mantegna 1999), an approach which is applied widely, e.g. in financial markets by Brida and Risso (2010), in medicine by Eisen et al. (1998).

The first step is to transform the correlation into a distance metric d fulfilling the following four requirements (3.9)–(3.12) where in the application of a clustering algorithm (3.12) is replaced by the stronger (3.13), since an ultra-metric is used.

$$d(x, y) \geq 0 \tag{3.9}$$

$$d(x, y) = 0 \Leftrightarrow x = y \tag{3.10}$$

$$d(x, y) = d(y, x) \tag{3.11}$$

$$d(x, z) \leq d(x, y) + d(y, z) \tag{3.12}$$

$$d(x, z) = \max\{d(x, y), d(y, z)\} \tag{3.13}$$

Following Anderberg (2014) and Mantegna (1999) we use

$$d(x_i, x_j) = \sqrt{2(1 - \rho_{i,j})}, \tag{3.14}$$

which is the Euclidean distance between the standardized data points x_i and x_j. The metric is bounded in the interval $d \in [0, 2]$ and smaller values correspond to a smaller distance and thus to more similarity. The clustering algorithm itself runs as follows:

1. Find the pair i, j which satisfies: $\quad d(x_i, x_j) = \min_{m,n} d(x_n, x_m)$
2. Merge the pair i and j into a single cluster
3. Calculate the distance to the other clusters
4. Repeat steps 1 and 2 as often as desired

In the calculation of the distance in the third step we choose the following approaches

- *Complete linkage*: the distance to the other clusters is set to the maximum of the two distances of the merging clusters (Anderberg 2014). The maximal distance corresponds to a minimum in the absolute value of the correlation.
- *Single linkage*: the distance to the other clusters is determined by the minimum of the two sets to be merged (Anderberg 2014; Gower and Ross 1969). The minimal distance corresponds to a maximum in the absolute value of the correlation.
- *Average linkage*: the distance to the remaining clusters is calculated as the equally weighted average of all merged cluster elements (Anderberg 2014).

Furthermore we apply the ward algorithm, where the criterion for selecting two clusters to merge is such that the variance within them becomes minimal (Anderberg 2014; Murtagh and Legendre 2014; Ward Jr. 1963). This approach should lead to the most homogeneous clusters, while our three linkage algorithms are based on the correlation with the complete linkage reacting most conservative to correlation, the single linkage most aggressively and the average linkage providing a middle way between these two.

To calculate the distance in step 3, there exist several algorithms, which all result in different cluster constituents. In order to choose the 'best' one, we use the concept of equally sized clusters: If we seek to form m clusers out of n assets and one of them contains $n-m+1$ assets, we end up with the most unequal cluster size possible. In this extreme case the likelihood for a randomly chosen asset to be in the one large cluster is highest and its cluster weight lowest. In contrast clusters are uniformly distributed in size the sensitivity of the cluster formation is higher. To gain a homogeneously sized clustering we use the (Herfindahl) Hirschman index (Hirschman (1964, 1980)).

In our application, the ward algorithm, where the criterion for selecting two clusters to merge is such that the variance within them becomes minimal (Anderberg (2014), Murtagh and Legendre (2014), and Ward Jr. (1963)) is the one which satisfies best our specifications. In general, it leads to the most homogeneous clusters, due to the minimal variance criterion, while other linkage algorithms like *complete linkage*, *single linkage* or *average linkage* are based on the correlation. Complete linkage is reacting most conservative to correlation, the single linkage most aggressively and the average linkage providing a middle way between these two. All algorithms result in a hierarchical, which does not make any statement about the clusters itself. Instead they are formed by cutting the hierarchical tree horizontally at a height such that the desired amount of clusters is formed.

In a final step we transform the each cluster into a cluster-portfolio, which is a sub-portfolio of the initial portfolio of all assets. In general the weights needed are free parameters and a choice for their determination is needed. As examples one can use the market capitalization, like in the NASDAQ-100 index, weights according to the position in the portfolio observed or simple equal weights. Since the first two methods impose an external structure, which prefers some assets to others, we use the third option of equally weighting all assets within each cluster-portfolio.

This concludes the clustering approach. We now turn to the PCA as an alternative technique to reduce the dimensionality. Here the given data X in the high dimensional Cartesian coordinate system are projected onto another orthonormal coordinate system, which is based on the eigenvalue decomposition of the data's covariance matrix.

Let $\Sigma = \mathrm{Var}(X) \in \mathrm{R}^{n \times n}$ denote the empirical covariance matrix. Than there exists a transformation (Bronstein 2012) as:

$$\Sigma = P\,\Lambda\,P' \quad \Lambda \in \mathrm{R}^{n \times n}, \quad P \in \mathrm{R}^{n \times n} \tag{3.15}$$

where P is the matrix formed by the eigenvectors of Σ and Λ as a diagonal matrix with the eigenvalues λ_i on its diagonal. The matrix of principle components is calculated as $Z = X \cdot P$. This is the representation of the given data in the eigenvector coordinate system. For the variance of the "rotated" data we get

$$\begin{aligned}\operatorname{Var}(Z) &= \operatorname{Cov}\left((X\,P)', X\,P\right) = \operatorname{Cov}\left(P'X, X\,P\right) \\ &= P'\operatorname{Cov}\left(X', X\right)P = P'\Sigma\,P = \Lambda \end{aligned} \tag{3.16}$$

which is a diagonal matrix and therefore uncorrelated data in the rotated system.

By using only the k largest eigenvalues, we can now reduce the dimensions in the orthonormal base. The obtained data is an approximation of the original data but with only k dimensions in the orthonormal base and n dimensions in the Cartesian space, cf. Hair et al. (2006). But the following concern comes with the application of the PCA: As the dimensions increases, we have to neglect relatively more dimensions, since only a rather small number of time series in the basis of the eigenvectors. We are then left only with the high volatile dimensions and a structural break detected therein can be regarded as a rather big and thus most obvious break. In this light we like to address the error made in neglecting small volatile dimensions. In general, the variation in the small eigenvalues corresponds to noise in the estimation of the true correlation matrix, cf. Laloux et al. (1999, 2000) and Plerou et al. (2002). As a result one can obtain a more accurate estimate when intentionally not using these dimensions. Thus is not true that a reduction in the percentage of variation represented by the first k eigenvalues likewise reduces the likelihood of detecting a structural break. Improving the accuracy of the estimated correlation may even occur, since noise has been neglected. Concluding, we only have to worry about the information associated with eigenvalues, which are smaller than the fourth largest eigenvalue and larger then this threshold eigenvalue.

3.2 Limiting distribution and finite sample evidence

The two methods of clustering and PCA are two approaches to reduce the dimension of the problem at hand resulting in a linear transformation of the given data and forming actually indices or sub-portfolios, which are analyzed in the Euclidean or another

"Eigen" space. Both transformations are linear. Thus the limiting distribution should not change. An economic interpretation is the following: Suppose one chooses stocks from a given universe and creates portfolios thereof. This investor then could use a test to analyze the correlation structure, resp. breaks in the structure between her portfolios. No matter how the weights are constructed, e.g. all weights are unequal to zero (PCA case) or some of them are zero (hierarchical tree case), a valid financial time series is created.

To test this, we simulate 2500 realizations of 1000 assets with 1500 time points and a fixed random correlation matrix. We randomly choose the eigenvalues in the interval $[1, 10]$ and use the columns of a random orthogonal matrix as the corresponding eigenvectors. Figure 3.1 shows the histogram of the test statistic on the left hand side and the corresponding empirical p values on the right hand side.

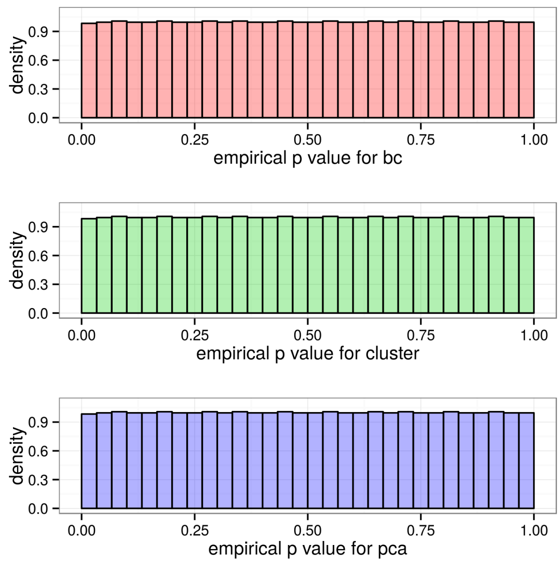

Figure 3.1: Empirical p values for the cluster and PCA approach. The figure shows the p values of the test statistic for both the cluster and the PCA approach. The original test statistic of Wied (2015) with 4 assets is denoted with bc.

We do not find deviations from the original limiting distribution in Wied (2015) for both methods. The p values seem to be uniformly distributed, so we conclude, that the limiting distribution does not change. Since these transformations are all linear, this result is what one can expect and the test seems to have the same validity as the original. Furthermore, we replace the generic correlation matrix from above with an empirical estimate from a sample of 50 stocks from the NASDAQ-100 index. Applying the same methodology as above, we find a very similar shape of the test statistic and of the distribution of the p-values.

The finite sample properties of the proposed test using a Monte-Carlo simulation clarifies what breaks are detected, resp. what it means, if the test statistic exceeds the critical value. We illustrate these for different number of observations per time series $t \in \{500, 1000, 2000, 4000\}$, and simulate such from a 100-dimensional normal distribution. In one case the correlation matrix is kept fix, in the other case we include a break in the middle of the time series. To do so, we change randomly $n_{\rho^*_{i,j}} \in \{100, 1000\}$ entries in the correlation matrix and transform it in a positive definite matrix. It is then used for the second half of the time series. We use 3, 4 and 5 clusters (with corresponding critical values of 2.27, 4.47 and 6.60) and check if the maximum of the test statistic is larger than the corresponding critical value. We repeat this process 2500 times. The result is shown in table 3.2. In all cases we find an increasing power in the number of reduced dimensions, but as a trade off the size distortion increases as well.

The upper part of the table shows the empirical rejection rate given H0 holds. For the PCA case, the size distortion seems to converge monotonously to the level of 5%. For the hierarchical tree, we find comparable sizes of size distortions but the convergence process seems not to be monotonous. The middle part corresponds to the case, where the break in the correlation structure of size $n_{\rho^*_{i,j}} = 100$ occurs. Keeping in mind that in a portfolio of 100 assets, a change in 100 pairwise correlations corresponds to roughly one asset changing its correlation to all other assets. So this is a rather small break in our setting. We find for nearly all cases a higher rejection rate for the hierarchical tree approach than for the PCA.

The lower part corresponds to the break size of $n_{\rho^*_{i,j}} = 1000$. Compared to the $100 \cdot 99/2 = 4950$ pairwise correlations overall, this is a rather big break. We find higher rejection rates compared to the small break before. This is what one expects, since the L1 norm increases and thus the probability of an exceedance of the critical value. In contrast to the case before, the rejection rates are higher for the PCA than for the hierarchical tree approach.

This suggest that no approach dominates the other and when using both approaches one can create two information sets, with a non-empty overlap. Compared with the original test, we find comparable rejection rates and size distortions for a break in

Table 3.2: Empirical rejection rate (in percent) of H0 given a portfolio of 100 assets and different number of clusters, given the 95% quantile.

	reduced dimensions PCA			reduced dimensions HT		
number of observations	3	4	5	3	4	5
	$\Delta\rho^*_{i,j} = 0$					
500	3.28	2.88	2.18	5.12	3.78	3.87
1000	4.58	3.81	2.66	4.54	4.67	3.13
2000	4.74	3.97	4.00	4.10	4.42	5.51
4000	4.80	4.67	4.14	4.32	4.93	3.97
	$\Delta\rho^*_{i,j} \neq 0, \quad n_{\rho^*_{i,j}} = 100$					
500	6.06	10.31	16.56	11.76	12.75	19.43
1000	5.30	10.81	26.00	12.88	27.18	23.18
2000	7.75	17.00	35.31	16.38	14.56	43.56
4000	13.69	23.56	50.88	22.50	42.50	50.69
	$\Delta\rho^*_{i,j} \neq 0, \quad n_{\rho^*_{i,j}} = 1000$					
500	75, 14	97.43	99.86	46.29	60.85	77.57
1000	74.86	98.57	100	68.14	92.86	96.43
2000	84.14	99.43	100	87.00	98.29	99.43
4000	92.43	99.86	100	88.43	100	100

1000 pairwise correlations. This suggests that when dealing with big breaks in a large portfolio the upstream clustering creates time series which can be treated as in the original test, such that the test has comparable finite sample properties.

3.3 Correlations within and between clusters

In the following we answer the question which correlation breaks we are able to detect. Using the notation of figure 3.2, we check for the correlation between clusters (ρ) and not for correlations within cluster ($\rho_{i,j}$) in the first place.

An implicit assumption is that the correlation changes within a cluster are transmitted to the clustered time series and thus are implicitly detectable. Consider the situation in figure 3.2. We have 4 assets (symbolized as circles) and form 2 clusters (symbolized as ellipses). Suppose the time series of the single assets are denoted as x_i for $i \in \{1, 2, 3, 4\}$

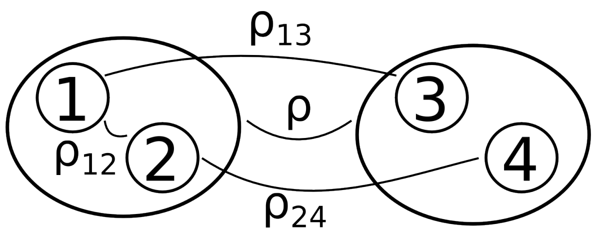

Figure 3.2: Notation for correlations within and between cluster. Shown are 4 assets (circles), where 2 assets are included in a cluster (ellipse). The correlation between clusters is denoted as ρ and the correlation between assets i, j as $\rho_{i,j}$

with zero mean and finite variance. Using the constant cluster weights w_i, the clustered time series y_i are calculated as follows:

$$y_1 = w_1 x_1 + w_2 x_2 \tag{3.17}$$

$$y_2 = w_3 x_3 + w_4 x_4 \tag{3.18}$$

For the correlation between the clusters, it holds:

$$\begin{aligned} \rho := \mathrm{Cor}(y_1, y_2) = w_1 w_3 \rho_{13} + w_1 w_4 \rho_{14} \\ + w_2 w_3 \rho_{23} + w_2 w_4 \rho_{24} \end{aligned} \tag{3.19}$$

Since the weights are constant over time, a change in this correlation is then a consequently a weighted sum of the single pairwise correlation changes:

$$\begin{aligned} \mathrm{d}\rho = \mathrm{dCor}(y_1, y_2) = w_1 w_3 \mathrm{d}\rho_{13} + w_1 w_4 \mathrm{d}\rho_{14} \\ + w_2 w_3 \mathrm{d}\rho_{23} + w_2 w_4 \mathrm{d}\rho_{24} \end{aligned} \tag{3.20}$$

It follows that the change of correlations within a cluster does not change the correlation between clusters per se, since $\mathrm{d}\rho$ is independent of e.g. $\mathrm{d}\rho_{1,2}$. Besides a very low probability of a change of one asset with solely one another within the cluster, there is also the condition of a positive definite correlation matrix of the whole x_i. So a change of e.g. $\rho_{1,2}$ cannot be independent of all other things, so that the other correlations are affected as well.

To check if the cluster formation is a stable process over time, we use a sample of 50 stocks over a timespan of 1500 observations and calculate the cluster members for each time point $t \in [100, 1500]$. We choose a lag of 100 data points to get a stable first estimate. We can conclude that cluster changes occur frequently especially at the beginning of the time interval which corresponds to the stability of the correlation matrix estimation. Also there exist many breaks in the correlation structure, which are so big, that assets would leave the cluster. In the following we simulate such breaks. We use 4 normal distributed variables correlated with a given correlation matrix with lengths of 500, 1000, 1500 and 2000 data points. To simulate a larger cluster, in which only one asset changes the within correlation, we set the weight w_2 randomly in the interval (i) $[0.1, 0.5]$, (ii) $[0.01, 0.05]$ and (iii) $[0.01, 0.5]$. So we can compare the behavior for small (i) and large (ii) clusters. Case (iii) serves as an overall assessment. We repeat this simulation 50000 times.

As a starting point we chose the matrix

$$\rho_0 = \begin{pmatrix} 1 & & & \\ 0.7717 & 1 & & \\ 0.2597 & 0.1328 & 1 & \\ -0.0589 & 0.1665 & 0.8523 & 1 \end{pmatrix}$$

which is positive semi definite and has a high correlation between assets 1 and 2 and assets 3 and 4. This matrix results in the desired clusters, so we create the situation shown in figure 3.2. This is our starting position in each simulation and we now distinguish the following cases:

(a) We change only ρ_{12}: This corresponds to fluctuations within a cluster that are not so large that the asset would leave the cluster. To do this, we set this value to a random number in the interval $[0.7, 1.0]$.

(b) We decrease ρ_{12} and increase ρ_{23}, ρ_{24}: This corresponds to a change of the cluster members in a way, that assets 2, 3 and 4 build a cluster. In order to simulate this, we set $\rho_{12} \in [-0.2, 0.2], \rho_{32} \in [0.6, 1.0], \rho_{42} \in [0.6, 1.0]$.

(c) We change the correlations of assets 2 and 3: This corresponds to a situation where a new clustering would result in a cluster of asset 1 and 3 and a cluster of asset 2 and 4. Thus we set $\rho_{12} \in [-0.2, 0.2], \rho_{13} \in [0.6, 1.0], \rho_{24} \in [0.6, 1.0], \rho_{34} \in [-0.2, 0.2]$.

In all cases we draw from the given interval until a positive semi definite matrix with the desired entries is obtained which is then used to change the time series' correlation at the middle of the simulated time span. Table 3.3 reports the detection of test, where 'BC' refers to the situation where we have just two equally weighted assets in each of the two clusters.

The table shows the cases (a) to (c) with the corresponding definitions from above. In general the size of the change in the correlation structure increases from (a) to (c). So we find a general tendency for higher rejection rates when moving from the left to the right hand side of the table.

The table is split up into four horizontal parts. The upper one, labeled with 'BC' is our base case, where we simulate only 2 assets in the two clusters. we find rejection rates of about $4\% - 5\%$ for the 95% significance level, which is a very poor power. When to size of break increases, the power converges to 1. also we find a general convergence with an increasing number of observations. For the second part, labeled with (i) we simulate a larger portfolio of 2-10 assets in the corresponding cluster. We find lower rejection rates as in the base case in general, which is expected, since the break size, relative to the number of other assets decreases. When dealing with a big break, we find for a sufficient number of observations high detection rates. The third part, labeled with (ii) simulates a portfolio of 20-100 clustered assets. Thus the relative size of the break decreases even more and the rejection rates drop further. Only in the case (c), where the break is such, that a re-clustering would lead to a switch of assets, rejection rates converge to 1 for a realistic length of the time series. The last part, labeled with (iii) we simulate a portfolio with 2-100 in the corresponding cluster. Thus it is located between (i) and (ii), with rejection rates in between (i) and (ii) for the corresponding observations and scenarios.

Concluding we find that the structural break test is not only able to detect breaks between clusters, but it may also be able to detect larger fluctuations within (unchanged) clusters. But this does not mean that finding no breaks between clusters means that there are no breaks within clusters.

Table 3.3: Empirical rejection rate (in percent) of H0 given two clusters of different assets. Base case has four assets, (i) 10-20 assets, (ii) 20-100 assets and (iii) 10-100 assets. The correlation structure was changed in the middle of the time series, depending on the scenario (a) with fluctuation within, (b) with an asset outflow or (c) with a switch of assets.

		(a)				(b)				(c)			
		observations				observations				observations			
	$\alpha/\%$	200	500	1000	2000	200	500	1000	2000	200	500	1000	2000
BC	90	8.55	8.93	9.04	9.66	87.15	99.81	100	100	87.61	99.96	100	100
	95	4.07	4.31	4.37	4.76	77.38	99.53	100	100	75.03	99.78	100	100
	99	0.65	0.70	0.81	0.94	47.62	97.18	100	100	37.16	97.78	99.99	100
(i)	90	8.49	8.86	9.13	9.49	39.57	60.87	71.93	80.33	76.59	98.32	99.93	100
	95	3.93	4.20	4.49	4.55	29.25	53.99	66.93	76.78	61.80	96.58	99.85	100
	99	0.62	0.79	0.85	0.87	12.72	40.15	57.75	70.04	28.28	88.78	99.32	100
(ii)	90	8.60	8.70	8.91	9.20	8.65	9.36	10.35	12.42	61.47	93.34	99.60	100
	95	3.93	4.15	4.33	4.36	4.09	4.52	5.30	6.77	46.86	88.54	99.05	100
	99	0.61	0.69	0.81	0.83	0.69	0.79	1.10	1.70	19.68	72.49	96.26	99.94
(iii)	90	8.58	8.90	9.03	9.56	39.31	60.91	71.92	80.16	76.50	98.32	99.93	100
	95	4.01	4.28	4.45	4.92	29.13	53.82	67.14	76.68	62.13	96.72	99.87	100
	99	0.70	0.78	0.85	0.95	12.60	39.94	57.71	69.90	28.14	88.83	99.33	100

3.4 Application to a real-life problem

We now turn to the application of the before mentioned methods. In a first step we want to check, if we can reproduce the break dates found in Wied (2015), although using a larger portfolio. To do so, we also use the EuroStoxx-50 index which is a widely used index on European stock markets, consisting of the 50 largest, by market capitalization, stocks traded publicly in European countries. In a second step, we use the algorithm described in Galeano and Wied (2014) to detect multiple breaks in the correlation matrix of the NASDAQ-100 index.

3.4.1 Detecting a single break

We obtain the index composition of early 2015 of the EuroStoxx-50 and individual stock prices for these constitutions from July 2005 – March 2015.

Based on this sample we form four portfolios:

(i) The portfolio of Wied (2015) which consists of four stocks from the EuroStoxx index: BASF, Sanofi, Siemens for the time span Jan 2007 – June 2012.

(ii) An extension of the portfolio in (i) with 21 other, randomly chosen stocks from the remaining index constitutions during the same period. This results in a portfolio of 25 stocks and thus half of the EuroStoxx50 index

(iii) The full EuroStoxx-50 over the same time period as in (i) and (ii).

(iv) The full EuroStoxx-50 index with all 50 stocks for the full time period 2005-15.

To mimic the setting of Wied (2015) we cluster the portfolios (ii), (iii) and (iv) into four cluster-portfolios as described in the previous section and shown in figure 3.3. The most homogenous clusters concerning the size are obtained with the ward algorithm.

To illustrate the difference to the exogenous sector-clustering provided by the index company we color each constituent with the corresponding sector color. The clusters do not become totally homogenous, concerning the industry sector, when increasing the number of assets monitored (figure 3.3). As an example, when monitoring all 50 assets, Deutsche Telekom and Telefonica are not placed in one cluster. This means, the inevitable error generated by the clustering would be bigger using the exogenous

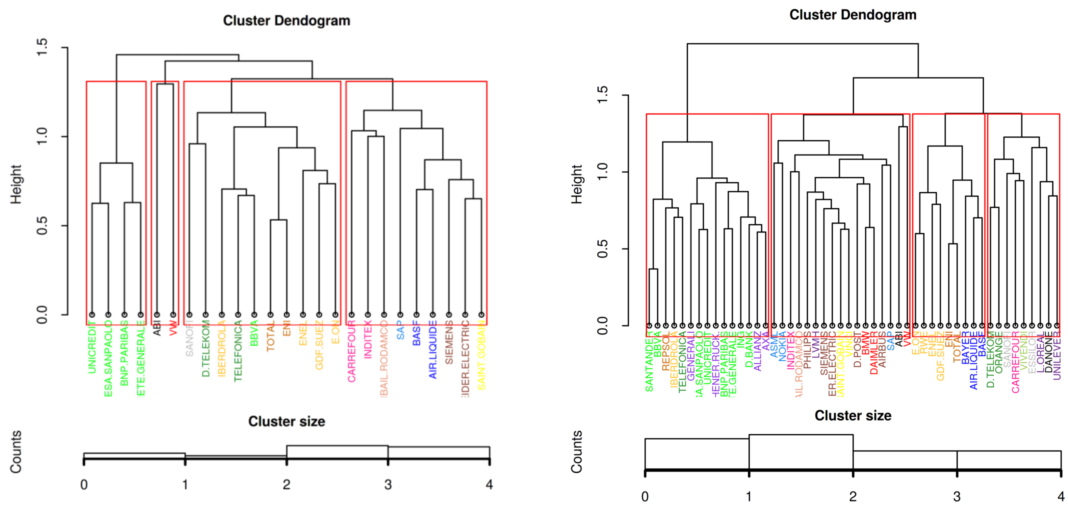

Figure 3.3: Hierarchical trees and cluster sizes. In the upper part, the figure shows the hierarchical trees for the portfolio (ii) and in the lower part for the portfolio (iii) and (iv). The trees are formed by the Ward algorithm. They are cut horizontally to form the four clusters. On the lower end of the trees we denoted the several companies and colored them according to their supersector. In the lower part of the figure, the number of cluster constituents is shown.

industry sectors, since companies from two different sectors may perform more similar than two in the same sector.

In the discussion of the results we use portfolio (i) as base case and show the result of the structural change test in figure 3.4. In the upper part, we see the actual test statistic for the clusters in black and the principal components in the red, dotted line. The horizontal line denotes the critical value 4.47, which corresponds to a confidence level of 95%. A structural change is detected if the test statistic exceeds the critical value, while the actual break date is given at the maximum of the absolute difference between the corresponding matrices. This difference is plotted in the lower graph scaled to one. Again, the black line shows the difference for the clusters and the dotted red line the differences for the principal components. We detect structural changes at the 443th data point which is 11th September 2008 in calendar date and the 460th data point, corresponding to 6th October 2008, in line with Wied (2015). Turning to portfolio (ii) we randomly add 21 stocks during the same period as an intermediate stage. The four resulting clusters are shown in figure 3.3.

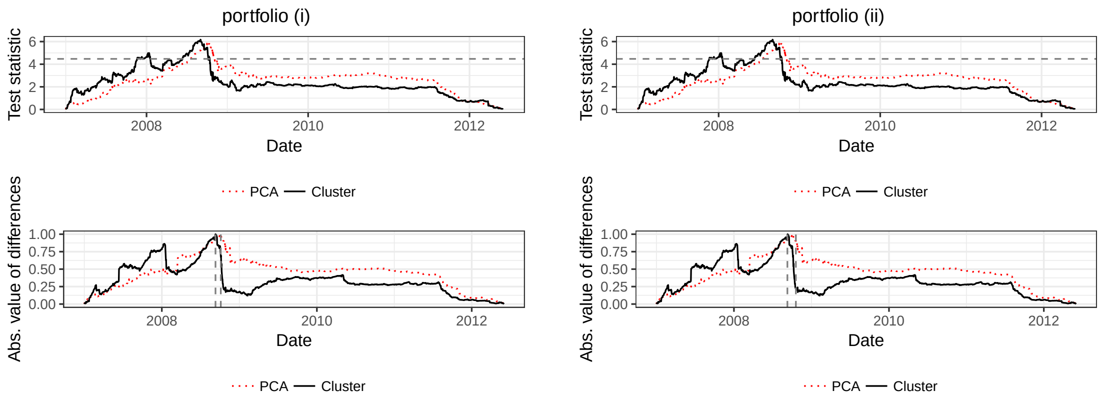

Figure 3.4: Test statistic and break detection for portfolio (i) and (ii). The figures show the resulting test statistics in the upper half and the scaled break detection in the lower half. The black solid lines correspond to the cluster approach and the red dashed lines to the PCA approach. The 95% confidence level is denoted by the black dotted line and the break dates are highlighted by the corresponding vertical lines.

The result of the structural change test is shown in Figure 3.4 left hand side with the previously described legend. The indicated break dates, as shown in the lower part of the figure, refer to the 11th September 2008 (443rd data point) and to the 21st October 2008 (471st data point). In the PCA case, we are able to preserve 73.6% of the overall variation. While it's break dates lag slightly by 15 days, the clusters' dates remain unchanged to the base case.

Turning to our full portfolio (iii) and (iv). The formation of the clusters is shown in figure 3.5 shows the cluster's composition. The difference between these two is the monitored time span. In (iii) we use the same period than before and in (iv) we use all given information.

When we look at the principal components, we are able to capture 65.8% of the overall variation in the data with the first four eigenvalues in both portfolios. The resulting test statistics of the structural break test are shown in figure 3.5. We see significant structural break both in the PCA and the clusters. The corresponding break dates are the 30th July 2008 (413th data point) and 2nd May 2010 (870th data point) in the case of portfolio (iii). We see a lag of nearly 2 years compared to portfolio (ii). It is shifted towards the one in portfolio (ii) when the time interval is increased to the whole data set in portfolio (iv). Here the corresponding break dates are the 29th July 2008 (800th data point) and the 3rd September 2008 (826th data point). Compared to

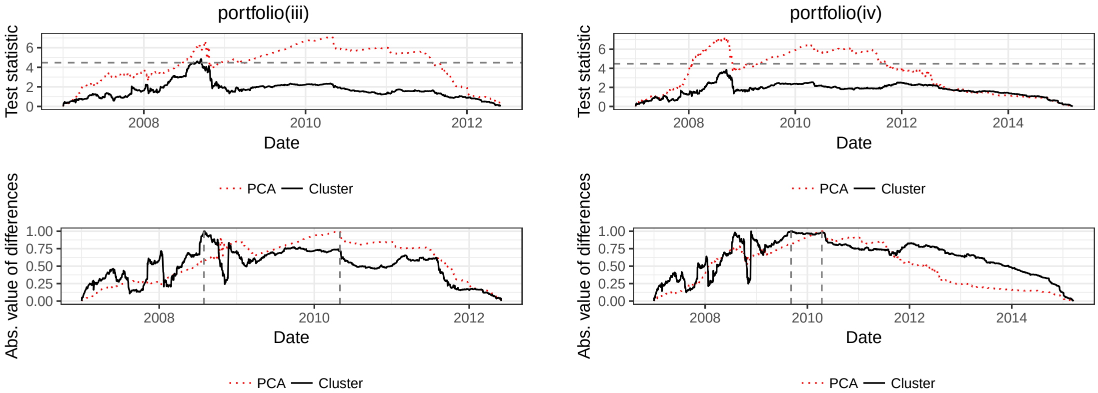

Figure 3.5: Test statistic and break detection for portfolio (iii) and (iv). The figures show the resulting test statistics in the upper half and the scaled break detection in the lower half. The black solid lines correspond to the cluster approach and the red dashed lines to the PCA approach. The 95% confidence level is denoted by the black dotted line and the break dates are highlighted by the corresponding vertical lines.

the results of portfolio (ii) we see that we now obtain leads which can be due the full information set used. With this sample we illustrated the use of a test for changes in the correlation matrix on a large portfolio where the number of assets does not allow a direct application of the test. Instead we propose two methods, clustering and PCA, to reduce the dimension prior to applying the test.

3.4.2 Detecting multiple breaks

We now turn to the application of the before mentioned methods. We use the stock prices of the constituent of the NASDAQ-100 index as of November 11th, 2015. Our observations span from January 2000 to December 2016. Thus we proxy a rather large portfolio, although it does not represent the NASDAQ-100 index at all times, since we do not adjust for fluctuations in the index constituents. We calculate log returns for all series and cluster them into 3, 4 and 5 clusters.

Figure 3.6 to 3.8 show the clustered time series for reduced dimension of 3, 4 and 5. The left hand side corresponds to the hierarchical clustering and the right hand side to the PCA. It can be seen that, while the volatility in the time series decreases in the PCA case with decreasing eigenvalues, this is not necessarily the case for the hierarchical tree. In the former case, we can preserve 40.8%, 44.1% and 46.8% percent of the total variation in the original data.

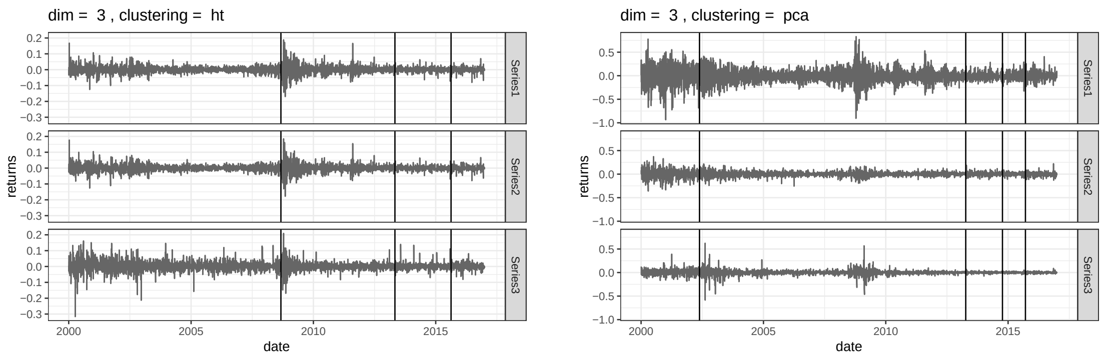

Figure 3.6: Break detection dates for 3 reduced dimensions. The figure shows the three clustered time series together with the associated break dates. The left hand side corresponds to the hierarchical clustering and the right hand side to the PCA.

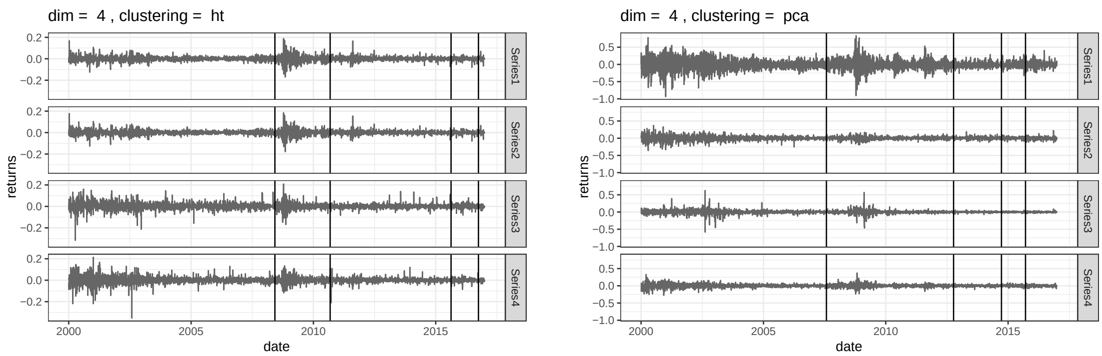

Figure 3.7: Break detection dates for 4 reduced dimensions. The figure shows the three clustered time series together with the associated break dates. The left hand side corresponds to the hierachical clusteing and the right hand side to the PCA.

At the beginning of our time interval from 2000 onwards to roughly 2003, we find very volatile time series. This could be connected to the Dot-Com Bubble which affected most the technology companies. Another obvious peak in volatility is located in the year 2008 and corresponds roughly to the default of Lehman Brothers, which is probably the best known starting point of the global financial crisis in public. As the NASDAQ index does not contain financial companies, figure 3.6 to figure 3.8 illustrate the distress in the whole market, not only in the financial sector.

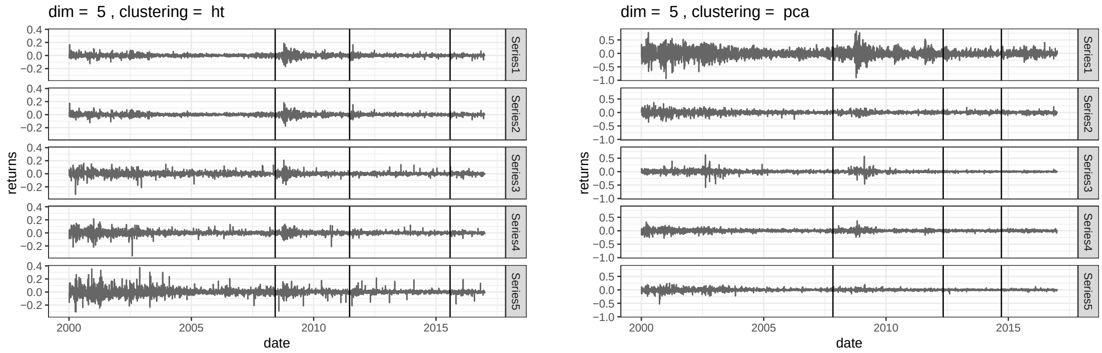

Figure 3.8: Break detection dates for 5 reduced dimensions. The figure shows the three clustered time series together with the associated break dates. The left hand side corresponds to the hierachical clusteing and the right hand side to the PCA.

The breaks in the correlation are indicated by the vertical lines in the figures and are listed in table 3.4.

Table 3.4: Detected structural breaks in the correaltion matrix of the NASDAQ-100 index according to different clustering algorithms and number of reduced dimensions.

reduced dimensions	cluster	break dates			
3	ht	3.9.2008	2.5.2013	18.8.2015	
	pca	22.5.2002	9.4.2013	9.10.2014	17.9.2015
4	ht	5.6.2008	8.9.2010	18.8.2015	30.9.2016
	pca	30.7.2007	9.10.2012	24.9.2014	18.9.2015
5	ht	5.6.2008	21.6.2011	30.7.2015	
	pca	31.10.2007	3.5.2012	19.9.2014	

In most cases, we find comparable results for the two clustering algorithms, for a given dimension. In two of the three cases, we find the same number of breaks, but we find deviations in the time location in general. A very short time difference is a break identified in 3 dimensions in 2013. Here the time difference is roughly one month when comparing the PCA with the hierarchical tree. As a more general example, for dimensions 4 and 5 the break date differs in about a year time difference. In the case of 3 dimensions, the first break is located in 2002 for PCA, whereas it is located in 2008 for the hierarchical tree. The latter case corresponds to 12 days before the bankruptcy of Lehman Brothers.

Keeping the clustering algorithm and looking at different dimensions, we find very similar break dates. For the hierarchical tree, a break in the second half of 2008 is indicated for all dimensions, as well as a break at the beginning, reps. middle of August 2015. For the PCA case, a break in the 3rd quarter of 2007, one in September 2014 and in September 2015 is located. Thus we are confident that these are actual breaks. Additionally there are four other breaks in the range from end 2010 to end 2012. The break in 2010 is only indicated once, as well as the one in 2011. The two breaks in 2012 for the PCA case in 4 and 5 reduced dimensions have a five month time difference.

On the other hand, we also find some outliers. As mentioned before, the first break in 2002 in the PCA case for 3 dimensions is not indicated by any of the other 5 tests. Another example is the break in 2016 detected by the hierarchical tree with 4 dimensions.

3.5 Conclusion

Monitoring the correlation matrix of a portfolio is a daily task in financial portfolio management and modern portfolio optimization heavily relies on the matrix in calculating the portfolio weights. Likewise triggers the occurrence of changes in this correlation structure, a structural break, a portfolio adjustment. While test for structural breaks are readily available and discussed in the literature, they can only deal with a low number of assets, e.g. Wied (2015) with only four assets. In practice, however, we most commonly have a situation in which the number of assets by far exceeds the available time series. Thus so far no test on structural changes in the correlation structure of typical portfolios was possible.

In order to tackle this dimensionality problem, we propose the use of a hierarchical tree as foundation for cluster formation. This technique has already applied to cluster financial markets, especially in the econophysics literature. As an alternative, the standard principal component analysis is used to form sub-portfolios. In a second stage a test on structural changes is then applied to the reduced problem.

Using a Monte Carlo simulation, we showed that we are able to extend the original test to a larger portfolio in the sense that we can reproduce the power with the corresponding size distortion as the original test for a sufficient big break in the correlation matrix for a large portfolio. Although we check for breaks between clusters we showed that, due to the fact of a positive semi definite correlation matrix, a change in the correlation structure within a cluster translates to some extend to the correlation between the clusters. We quantified this effect through a simulation as well.

Finally we applied both techniques to the case of constituents of the NASDAQ-100 index over a time period of 17 years. We do not check the index itself (since there are fluctuating constituents), but simulate a rather large portfolio compared to the 4 assets in the example in Wied (2015). We found some breaks in the correlation structure, which are independent of the reduced dimensionality and one, which is independent of the clustering approach at the first stage.

4 Income distribution in troubled times: Disadvantage and dispersion dynamics in Europe 2005-2013

The following is based on Bowden, Posch, and Ullmann (2017b)

Income inequality has been a perennial topic of economic and social interest, but never more so than the present, where dramatic changes have followed within just a short span of time. Performance driven management rewards seen as excessive have attracted much public attention. But other influences have been at least as pervasive. Technological displacement for middle management, import competition from cheaper emerging countries, free trade agreements, adverse fallouts from public spending bubbles, commodity price reversals, the global financial crisis, are all some of the causal influences, combining as the perfect storm in their fallouts for remuneration and employment down the line. The empirical content of the present paper, namely the ten years in Europe just prior to and succeeding the Global Financial Crisis (GFC), reflect changes of this sort. To what extent they represent welfare loss, and how to measure it, requires a methodology that is itself adapted to the economic and social context. Metrics are one thing; meaningful metrics within context are quite another. The agenda of the present paper is to develop and apply dual metrics for asymmetry and spread that can in turn be rationalized within a social welfare framework.

Metrics for income distributions are part of a wider body of knowledge into social welfare functions developed and debated over many years. But establishing a consensus has not been easy. At the most abstract level, Goodman and Arrow (1953) showed that the possibility of a consistent ordinal social utility, as earlier envisaged by Bergson (1938) and Samuelson (1947), was limited at best and certainly would not extend to universal agreement among the subjects themselves as to a single best metric for income inequality. Proposed axiomatic bases list desirable characteristics of possible metrics;

Cowell (2009) and Kraemer (1998) are useful surveys. But while criteria such as the Pigou-Dalton transfer principle (Dalton 1920; Pigou 1912) are widely accepted, others leave space for interpretations or disputes, e.g. as to whether absolute or relative scale invariance should be imposed.

The empirics have therefore focused upon metrics for income inequality that appeal in designated ways to the observer's own preconceptions as to fairness. Still the best known and most widely accepted metric of this kind is the Gini coefficient, which measures the non-alignment of the accumulated percentage of income with the progressive numbers of the people enjoying it. As the Gini coefficient is not able to represent different social preferences Atkinson (1970), Donaldson and Weymark (1980) and Yitzhaki (1983) propose a parametric extension, where the parameter can account for different preferences (Greselin and Zitikis 2015). Sen (1970) proposed a metric that combines in itself both the mean income and the complement (1-G) of the Gini index.

But the Gini has several downsides, both statistical and interpretive. One such difficulty is the ambiguity of the coefficient itself. Since it reflects the integral of the difference of the 45° line and the Lorenz curve, one can mirror the Lorenz curve on the orthogonal line that intersects the 45° line in the middle, and end up with just the same Gini. Apart from this, there is a potential ambiguity problem arising in intertemporal or cross country comparisons. Atkinson (1970) and Davies and Hoy (1995) point out that where two Lorenz curves intersect it may not be possible to rank the Gini coefficients, even using the same underlying social welfare function; only with further assumptions about the variance, this is possible. The problem is a general one and applies to all measures based on the Lorenz curves. As a possible alternative, the Theil index (Theil 1967) uses the distribution's Shannon entropy in order to measure the expected information content of the distribution, which is then interpreted as inequality. But as Sen 1970, p. 36 put it, '... the average of the logarithms of the reciprocals of income shares weighted by incomes shares [the Theil index] is not a measure that is exactly overflowing with intuitive sense.' See also Conceição and Ferreira (2000). On the other hand, the Theil index does obey the axiom of decomposability (Foster 1983; Kraemer 1998), which means that the measure for the whole distribution can be expressed as a weighted sum of the measures applied to subgroups. Bourguignon (1979) showed that the Gini coefficient, together with other proposed metrics, does not have this property.

But a more contextual problem with the Gini index is that it lacks direction as to the source of the inequality, and in this sense fails to separate out directional skewness from dispersion or spread (Bowden 2016b). Specifically, it does not properly pick up the kind of inequality that would concern most observers, namely positive distribution skewness, meaning too many people on low incomes. Following the precept of Dalton (1920) that any measurement of inequality has implicitly a social welfare reference, Atkinson (1970) also Foster, Greer, and Thorbecke (1984), address the direction problem by means of an observer chosen parameter that reflects a greater, or lesser, observer preoccupation with the positive skewness in the income distribution. Formally, the Atkinson index can be transformed to the Theil index and thus both indices belong to the class of the generalized entropy measures, which is a single parameter family (Shorrocks 1980; Cowell 1980). In turn, there exist generalizations from the single parameter family to a two parameter family (Foster and Shneyerov 1999).

An internal rather than external vantage point using conditional expectations and thus comparing 'poorer' with 'richer' people is given in Eltetö and Frigyes (1968). The reference point is set as the mean income of the population. The measures are given by fractions of the conditional expectation of incomes larger, resp. smaller than the mean income and the mean income itself. It can be shown that the three measures can be condensed to the relative mean deviation (Kondor 1971).

In a recent contribution by Bowden (2016a) all possible reference points are used in the calculation, one for each subject in the population. The v-index (standing for disadvantage or envy) looks at inequality from the point of view of the subjects themselves. On the average, do people think that others are better off, or worse off, than themselves? The resulting metric can be simply expressed in terms of the means of ancillary distributions, the left and right unit entropic shifts of the original. As an asymmetry measure, the v-metric adheres to all the standard axioms or prescriptions that have been proposed in the general statistics literature (Kraemer and Conring 2017). However, its special usefulness in the current context originates from two properties. The first is the connotation as a relative advantage metric. The second arises from an inherent duality property whereby a simple internal change of sign yields an accompanying metric for the spread of incomes. Dispersion of the income distribution can be viewed as a 'noticeability' property. Wide dispersion attracts more attention to very high or very low incomes, in this context relative to my own. The dual d-metric for spread therefore adds a further measure of social welfare. Taken together, the two metrics, v for directional asymmetry and d for dispersion, resolve the inherent ambiguity of the Gini coefficient in terms of asymmetry and the spread Bowden (2016a).

All this suggests that three key metrics are involved in social judgments about income distribution: the mean (or median), the directional skewness, and the dispersion (spread). A cardinal utility analysis would call on some optimum weighting of the three welfare dimensions: mean, asymmetry, and dispersion. But in an ordinal framework, one can adopt an indifference curve approach, reminiscent of the ordinal utility Bergson-Samuelson framework for a social welfare function. Longitudinal studies over time in the three key metrics can be used to study comparative welfare economics between countries. Comparisons of this kind are facilitated if the metrics for asymmetry and spread fall within a common framework. The duality between the v- and d-metrics ensures both dimensional conformity and a convenient phase plane diagram over time.

The empirical application utilizes such a framework to study changes that took place pre- and post- the Global Financial Crisis in Europe. This is not to say that the GFC and its aftermath were by any means the sole causal influence. These were also the times of structural change arising from import competition, the oil price rise and fall, and other contributing factors. The result was some dramatic changes in income distribution metrics and implicitly, the rise or fall in their latent social indifference level surfaces (section 4.1.3). Nor were these changes equally severe across the respective countries. The empirical work of the present paper translates the experiences of 15 of the major European countries in the form of dynamic phase planes between directed asymmetry and spread. It turns out that there are marked differences, which can be categorized into two groups, the risers ('bad') and the fallers ('good'). These tend to be uncompensated by relative difference in their respective mean income series.

4.1 Methodological review

In what follows the conceptual basis is reviewed, starting with the metrics for disadvantage as directed asymmetry (section 4.1.1), followed by spread (section 4.1.2). Mathematical details are kept to a minimum; a more comprehensive treatment can be found in Bowden (2016a, 2016b, 2012), also Kraemer and Conring (2017) for the axiomatic basis as an income inequality metric. Section 4.1.3 turns to consideration of the social welfare; the asymmetry and spread metrics for any given year can be viewed as analogous to points on a social welfare indifference curve. Over successive years this curve may itself change, with social welfare shifts as an outcome.

4.1.1 The asymmetry metric

Let Y be the income random variable, defined on $\mathbb{R}_+$ with cumulative distribution function F and density f. Suppose my own income is y. The metric for asymmetry derives from comparisons between incomes above mine and those below. The average of those below is the lower conditional expectation $\mu_l(y) = \mathbb{E}\left[Y|Y \leq y\right]$ and for those above, $\mu_r(y) = \mathbb{E}\left[Y|Y > y\right]$. Relative to those above and below, my net disadvantage is

$$v(y) = (\mu_r(y) - y) - (y - \mu_l(y)) . \tag{4.1}$$

Taking the expected value over the entire distribution $F(y)$ gives a scalar metric that represents the average degree of peer relative advantage:

$$v = \mathbb{E}\left(v(y)\right) = \mu_L + \mu_R - 2\mu. \tag{4.2}$$

Here μ is the mean of the original or natural income distribution. The two terms μ_L, μ_R also refer to means, but in this case of ancillary distributions with densities defined by

$$f_L(y) = \xi_L(y)f(y); \quad f_R(y) = \xi_R(y)f(y), \tag{4.3}$$

where $\xi_L(y) = -\ln F(y); \quad \xi_R(y) = -\ln\left(1 - F(y)\right)$ are factors (technically, Radon-Nikodym derivatives) that shift the original distribution respectively to the left and right. The corresponding distribution functions are given by

$$F_L(y) = F(y)\left(1 + \xi_L(y)\right); \quad 1 - F_R(y) = \left(1 - F(y)\right)\left(1 + \xi_R(y)\right) \tag{4.4}$$

In such terms, the metric 4.2 becomes easy to compute, starting with the original histogram for the income distribution. Expression 4.2 can be more compactly expressed as $v = 2\left(\mu_c - \mu\right)$, where μ_c is the mean of the centred or average shift $F_c(y) = \left(F_L(y) + F_R(y)\right)/2$. Noting that μ, μ_l, μ_r are all commensurate first moments, the standardisation $\tilde{v} = v/\mu$ is recommended for time series or cross section applications.

The resulting metric v is explicitly a directional asymmetry metric. A value $v > 0$ means that the distribution is positively skewed, so that $\mu > x_m$, the median, while if $v < 0$ the reverse is true: $\mu < x_m$. Unlike textbook skewness diagnostics such as the third order moment, the v-metric has contextual reference, in this case to how people feel about their comparative income. Additionally, since the existence of higher moments in income distributions cannot be granted in general (Kleiber 1997), it is possible to assess skewness for a more extended set of distributions.

Figure 4.1 illustrates with the Norwegian monthly disposable income data (truncated at 200,000 NOK). Over the time span 2005 - 2013 the distributions have become more symmetric, manifested by the corresponding v-metrics, though still with some way to go towards perfect symmetry ($v = 0$). It should further be noted that the mean has increased over the years. The standardised values $\tilde{v}$ are respectively 0.078, 0.064 and 0.036 (cf. table 4.3).

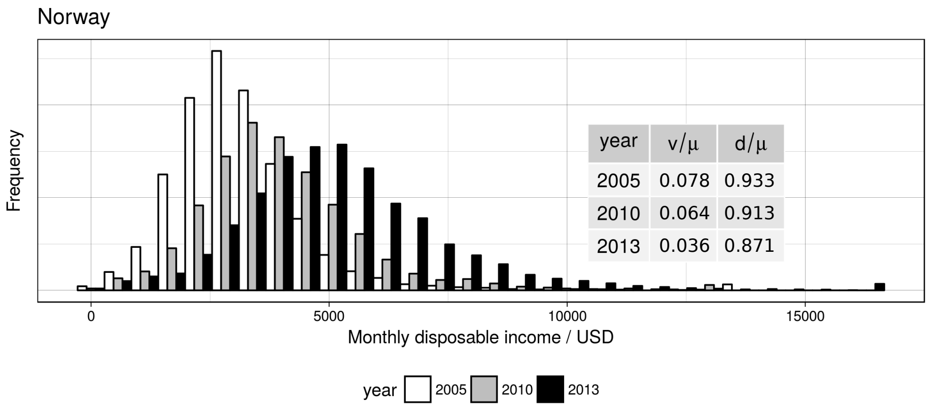

Figure 4.1: Distribution of the monthly disposable income in Norway. Shown is the histogram of the monthly disposable income of Norway for the years 2005, 2010 and 2013. The table shows the asymmetry and dispersion metrics relative to the mean income.

4.1.2 The dual dispersion metric

A metric for dispersion or distribution spread that is dual with respect to the v-metric is derived by simply changing the separating sign in expression 4.1. Thus the function

$$d(y) = (\mu_r(y) - y) + (y - \mu_l(y)) \tag{4.5}$$

can be interpreted as a dispersion penalty function. This is a two sided function rising away from the median in either direction. Its expected value is given by

$$d = \mathbb{E}\left(d(y)\right) = \mu_R - \mu_L. \tag{4.6}$$

Note that $d > 0$ and by construction, $d > v$. In turn, this can be standardised as $\tilde{d} = d/\mu$. Several rationales exist for the interpretation of d as a spread or dispersion metric. It is equal to the total area under the partition entropy function, which refers to the distribution of uncertainty as to whether $Y \leq y$ or $Y > y$ as y varies along its range. Bowden (2012, 2016b) contains further details as to this link. A related rationale runs in terms of the accumulation function of stochastic dominance theory. A distribution for which $F(y)$ slopes more gradually will always have a higher value for the metric d.

For the Norwegian example as in figure 4.1, the d measure has decreased over the given horizon. The effect is accentuated in the standardised versions $\tilde{d}$, where the values are 0.933, 0.913 and 0.871y (cf. table 4.3). So the net effect, taken in conjunction with $\tilde{d}$, is for a narrowing of the distribution spread together with diminishing asymmetry.

4.1.3 Social welfare aspects

The v-metric satisfies a number of standard criteria for income distribution metrics, such as invariance under translation and scale, or under internal transfers that do not affect rankings, the 'Robin Hood' principle (Kraemer 1998; Kraemer and Conring 2017). It also straightforward to show that the dispersion metric d is invariant under translations, proportional under scaling ($d(c \cdot y) = c \cdot d(y); c > 0$), and increases under a median preserving spread, i.e. is a dispersive ordering (Jeon, Kochar, and Park 2006).

In addition, the three metrics, namely v for asymmetry and d for spread, together with μ (or the median) for central tendency, provide potential cues for thinking how the average person might react to the publication of income data or an income distribution histogram. A positive skewness indicator v will indicate that the average person is net disadvantaged relative to those below and above him or her on the scale. This will be more noticeable where the distribution has a wider spread. And in both cases the reaction will be moderated when the average worker's income has itself risen.

The foregoing suggests an ordinal social welfare function, geared to the average worker, of the general form

$$SWF = \Psi\left(\mu, v, d\right), \tag{4.7}$$

with $\varphi_1 = \frac{\partial \Psi}{\partial \mu} > 0; \quad \varphi_2 = \frac{\partial \Psi}{\partial v} \lesseqgtr 0; \quad \varphi_3 = \frac{\partial \Psi}{\partial d} < 0$. As a further aid to interpretation, suppose that φ is separable between μ and v, d so:

$$SWF = \Psi\left(\mu, \varphi\left(\tilde{v}, \tilde{d}\right)\right). \tag{4.8}$$

This would certainly be the case if the parent function (4.7) was homogenous in its three arguments, but it is not necessary to introduce such a restriction.

As it stands, the form (4.8) could be interpreted as saying that the average worker looks first at the spread and asymmetry in relation to his or her own income, then modifies any reaction if the personal income is higher or lower during any given year. For any given income there is therefore a set of indifference curves (level surfaces) as between spread and skewness. However, these may not be uniformly concave or convex. For as skewness (v) becomes more positive, it requires progressively lower dispersion (d) to materially lessen the envy. By way of contrast, if v becomes more negative, it might require a progressively higher spread to preserve the same social utility. So the social indifference curves relating skewness to spread might therefore be sigmoid in shape.

4.2 The European experience: temporal phase plane

During the years 2005 - 2013 Europe passed through a cycle of boom and in some cases, bust. Although there was a common exposure to the GFC, recovery was far from universal across the zone. In part this was due to differing exposures to private and public sector debt. In addition, specific casual influences differed as between individual countries. Oil prices recovered quickly between 2009 and 2013, to the benefit of producers such as Norway, the Netherlands and the UK. A number of countries, including Poland and the Czech Republic joined the EU in the course of its 2004 enlargement. Some countries were impacted more than others by the rise of manufactured imports from China and Vietnam. Data availability precludes coverage of subsequent events such as the collapse of oil prices and more recently the refugee crisis. The empirical analysis that follows is based on data drawn from Eurostat, EU Statistics on Income and Living Conditions (2013). The survey for creating the dataset started in 2003 with six members of the EU and spans 32 countries in the year 2016 (Eurostat 2016b). It was the first dataset which provided a longitudinal and cross-sectional micro-level dataset Iacovou, Kaminska, and Levy (2012). Since we do not try to quantify effects on the income distribution separately, discussion is confined to the cross-sectional dataset. The subset selected for detailed analysis covers 15 European countries between 2004 and 2013. In our analysis, we use the disposable income on the household level. Consistent with Eurostat (2014), we correct for different household sizes ending up with an equivalised household income, which is assigned to the corresponding household. Following Kerm (2007) we winsorize our data, confining its scope to nonnegative values and employing the 99.5% quantile as an upper threshold. Incomes are converted from local currencies to comparable US Dollars based on the corresponding purchasing power parity ratios, provided by the Wold Bank (WorldBank 2016b).

4.2.1 Dynamics with the Gini index

A preliminary check is whether the Lorenz curves do intersect in our dataset. Table 4.1 shows the number of intersections for the Lorenz curves in the case of Germany. In only five out of 28 combinations the curves none are found. Since the assumptions about the relative variance required for intersecting curves are not met, it would therefore not be legitimate to compare the Gini coefficients between intersecting years. Furthermore when moving to a cross-sectional perspective, the number of comparisons of Lorenz curves will increase exponentially and with it the incommensurability of the Gini's within time or across countries.

Although comparison between different Gini values is non trivial, most analyses use the Gini, resp. it's temporal changes and try to explain these. That is why we want to

Table 4.1: Number of intersections of Lorenz curves for the disposable income in Germany

2006	2007	2008	2009	2010	2011	2012	2013	
1	1	1	1	5	1	4	5	2005
	5	3	3	2	2	2	3	2006
		5	0	2	2	0	1	2007
			2	0	4	0	2	2008
				2	10	2	3	2009
					1	2	5	2010
						0	3	2011
							3	2012

clarify the correlation between the Gini and the v and d metrics, independently from the fact from above. Formats in figure 4.2 show the temporal paths of the Gini measure (dashed, blue), the v metric (solid, red) and the d metric (dotted, green). On the first sight, we find a rather strong correlation between the d metric and the Gini whereas the correlation between the v metric and the Gini is not that strong.

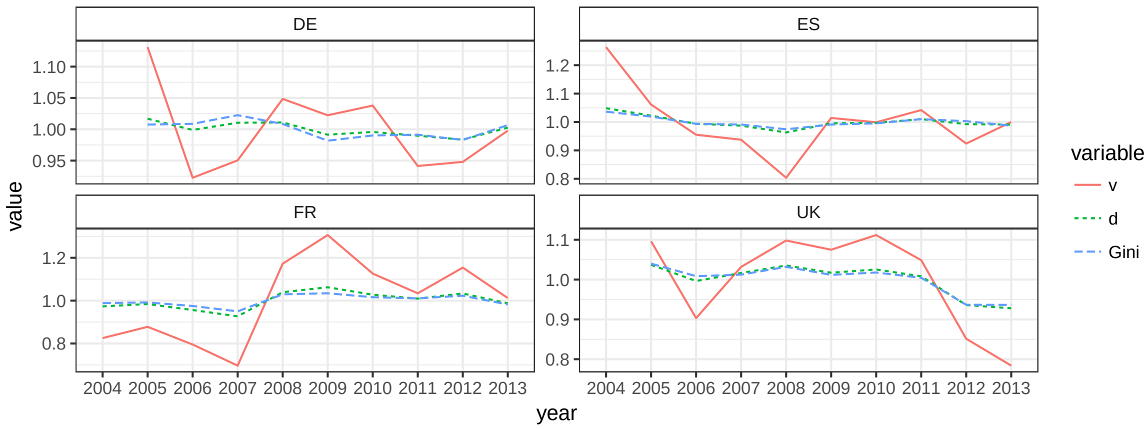

Figure 4.2: Temporal paths of inequality measures. The figures show scaled temporal paths of the v, d and the Gini measure for different countries. For clarity, the time series have been scaled to a unit mean.

To analyze the pattern further, we use a sample cross correlation function on the log returns of the inequality metrics (cf. Cryer and Chan 2008, p. 265ff). As an example, the cross correlation functions for Germany are plotted in figure 4.3. Shown are the lags $h \in \{-3, \ldots, 3\}$. The 5% significance level is denoted as the dotted blue line. At lag $h = 0$, only the right part shows a value, significantly different from zero. For all the other lags, we find a non-significant correlation.

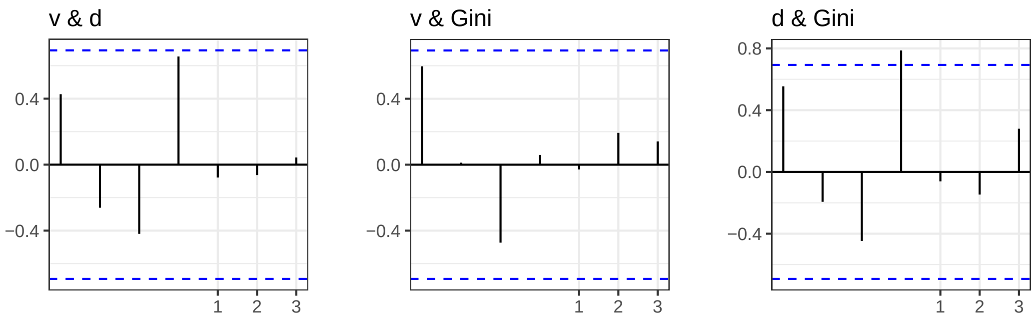

Figure 4.3: Sample cross-correlation functions for Germany. The figures shows the sample cross correlation functions for Germany. The left part shows v and d, the middle part v and Gini and the right part d and the Gini. The 5% confidence level is denoted by the blue, dashed line. Shown are lags $h \in \{-3, \ldots, 3\}$

We repeat this analysis for all 30 countries and out of these 30 cases, we find correlations significantly different from zero in 27 cases for the Gini and the d metric, 16 cases for the Gini and the v metric and 22 cases for the v and d metric. This supports our graphical interpretation of a strong correlation of d and Gini, whereas the correlation of v and the Gini is not that strong.

4.2.2 Dynamics in the v-d plane

A dynamic phase plane analysis turns out to be an effective way of highlighting both the similarities and differences. For space reasons the phase diagrams are reproduced only for the more major economics, or else those that were subject to particular stress. Diagrams for all the remaining European countries can be obtained from the authors.

Figure 4.4 plots the normalized v metrics against those for spread on the horizontal axis for the year 2013 for all countries, resp. figures and B.1 for each country from 2004-2013 in the appendix. No country has a negative v metric, hence the positive vertical half axis. The countries differ as to the observed range of the normalized asymmetry and spread metrics $\tilde{v}, \tilde{d}$, and this is reflected in the axis ranges of the respective diagrams. As a visual aid, directional links are employed to make the time sequence more identifiable. The respective time periods of data availability can be identified from the beginning and end dates of these sequences.

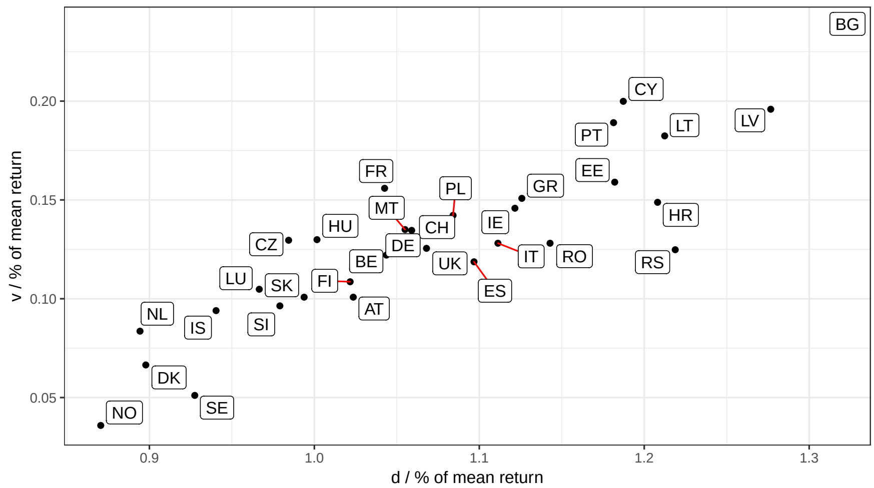

Figure 4.4: Phase diagram in 2013. Shown is the phase diagram for all countries in the year 2013

As a useful if somewhat imprecise generalization, the respective v-d dynamics can be categorized into three groups, provisionally labeled the 'inequality risers', the 'inequality fallers'; and 'mixed' where the trend is less clear. Table 4.2 lists the countries in this format. Further details about these groups and their constituents are provided in Table 4.3. There, the first column gives the average GDP per capita change over the available data horizon (WorldBank 2016a) for each country as the third dimension of welfare analysis (cf. expression 4.8,), together with the overall ranking of the growth. There is no clear association on this count between inequality risers and fallers. Also reported are the asymmetry and dispersion parameters before, at, and after the GFC together with the corresponding ranking within the 15 countries. Comparative rankings are not automatically responsive to individual metric changes, though there is some effect with inequality increases.

What the data do show is that the impact of economic events has varied between the European countries. For some of the countries in the upper right part of fig 4.4, bad economic times affected more at the low end of the income scale (such as Italy, Spain and France); while in Portugal the burden has been spread more uniformly. This is also true of Germany and the Netherlands, even though the economic impact was more

Table 4.2: Secular inequality trends over the sample period from 2004 - 2013

Inequality risers v, d rise over time	Inequality fallers v, d fall over time	Mixed
Austria	Czech Republic	Belgium
France	Germany	Hungary
Greece	Ireland	Italy
Netherlands	Norway	Spain
	Poland	Sweden
	Portugal	
	United Kingdom	

muted. Newer members of the European Union have fared better in this respect. The Czech Republic, Hungary, and Poland, all admitted in the 2004 Enlargement, share a common experience of falling income inequality and a healthy average GDP growth rate, although Hungary is labeled as a "mixed" country (cf. fig B.1 for details).

Table 4.3: Results of the proposed new inequality measure with ranking[α]

	Av. Change GDP/ % p. cap.	Rank	v/μ 2005	v/μ 2010	v/μ 2013	Rank v/μ 2005	Rank v/μ 2010	Rank v/μ 2013	d/μ 2005	d/μ 2010	d/μ 2013	Rank d/μ 2005	Rank d/μ 2010	Rank d/μ 2013
							Risers							
AT	4.025	5	0.1228	0.1207	0.1008	4	7	4	1.0068	1.0663	1.0236	4	8	7
FR	2.757	9	0.1352	0.1735	0.1559	7	13	15	1.0391	1.0851	1.0426	8	9	8
GR	-0.393	16	0.1554	0.1743	0.1508	13	14	14	1.1762	1.1736	1.1258	13	14	15
NL	2.961	8	0.0606	0.0905	0.0836	2	3	3	0.848	0.9045	0.8943	1	2	2
							Fallers							
CZ	6.097	3	0.1386	0.1139	0.1296	10	5	10	1.0143	0.9768	0.9844	5	5	4
DE	3.928	6	0.1384	0.127	0.1221	9	9	7	1.0585	1.0366	1.0437	9	7	9
IE	0.246	15	0.1901	0.1361	0.1458	15	11	13	1.2241	1.108	1.1216	15	13	14
NO	6.764	2	0.0778	0.0643	0.0359	3	2	1	0.9326	0.9132	0.8705	3	3	1
PL	9.09	1	0.1535	0.1469	0.1422	12	12	12	1.1495	1.1022	1.0841	12	11	11
PT	1.886	11	0.2483	0.2211	0.1891	16	16	16	1.3366	1.24	1.1814	16	16	16
UK	0.701	14	0.1755	0.178	0.1255	14	15	8	1.1937	1.1802	1.068	14	15	10
							Mixed							
BE	3.265	7	0.1264	0.1213	0.1154	6	8	5	1.015	1.0273	1.0232	7	6	6
HU	2.722	10	0.1379	0.1126	0.1299	8	4	11	1.0144	0.9535	1.0016	6	4	5
IT	1.333	13	0.1453	0.1273	0.1281	11	10	9	1.13	1.0951	1.1113	10	10	13
ES	1.349	12	0.126	0.1186	0.1187	5	6	6	1.1317	1.1036	1.0968	11	12	12
SE	4.989	4	0.0487	0.0421	0.0511	1	1	2	0.8908	0.8957	0.9275	2	1	3

[α] The table shows the average GDP growth from 2005 to 2013, the scaled dispersion and skewness measures before the GFC (taken as 2005), just after the GFC (taken as 2010) and most recent available (2013). Data for the GDP per capita are taken from WorldBank (2016a).

4.3 Determinants of income inequality

In the following we assess which economic observables influence the income inequality measured by our v and d.[1] The heterogeneity of both v and d is shown in figure 4.5.

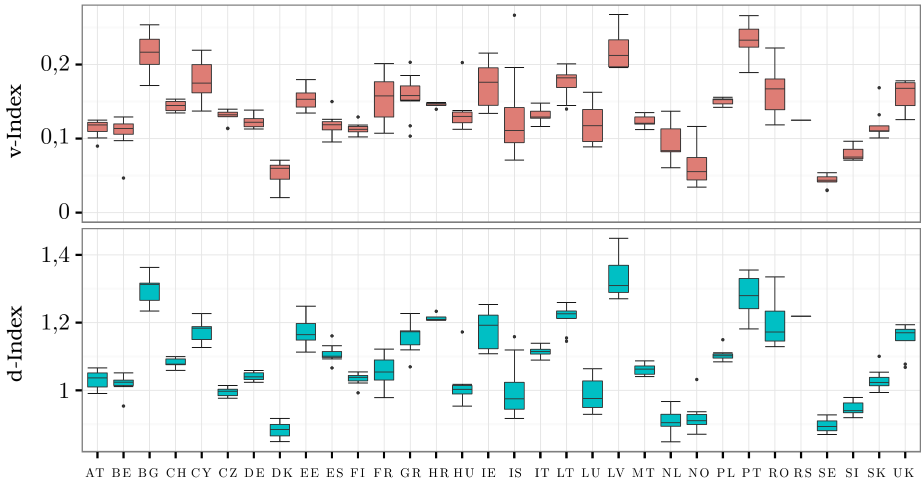

Figure 4.5: Boxplots for v and d per country. The upper box-plot shows the v-Index per country and the lower box-plot the d-Index per country from 2004 to 2013. The whiskers span over the 1.5 interquartile distance.

The upper part corresponds to the v-index and the lower part to the d-index. Shown are box-plots per country, i.e. the box spans from the lower to the upper quartile and the whiskers denote the 1.5 interquartile distance. We find a rather stable observables of inequality within a country but rather big differences between countries. This is in line with Li and Zou (1998), where data from 1947 to 1994 where used. E.g. Bulgaria, Latvia or Portugal, as countries on the upper scale on the inequality scale, show the highest inequality, when looking at the median and no observation lies in the observation of e.g. Norway or Sweden, which are located at the lower end of the inequality scale.

In order to account for this heterogeneity, we use the following fixed effects model:

$$y_{i,t} = \alpha_i + \sum_n \beta_n x^n_{i,t} + \epsilon_{i,t} \tag{4.9}$$

where $y_{i,t}$ denotes the v-index, resp. the d-index of country i at time t and the $x^n_{i,t}$ the

[1] We thank Kevin Dresmann, who finalized and calculated the following models during his Master Thesis, which I supervised.

independent variables corresponding to country i at time t. Due to different scales of our observables, we use standardizes versions of the variables, discussed below, e.g. we scale them to mean zero and standard deviation one. Other analyses (cf. Nielsen and Alderson 1995; Lee 2005) use a random effects model, where one has to assume that the constants α_i are uncorrelated with all other regressors (Wooldridge 2010). Since this seems rather implausible, we use the fixed effects model, which "... can lead to insignificant coefficients in statistical tests ..." (Lee 2005, p. 168). Our model choice is also supported by a Hausman-Test, which rejects a random effects model at least at the 5% significance level.

All our regressors are listed in table 4.4, together with their abbreviations, used in the following, the corresponding literature, where the effect of the corresponding regressors on inequality is analyzed, and the origin of the data we use. Also we group the independent variables into three groups, where the hypothesis are either a positive effect on our measure of inequality, no effect or a negative effect. This grouping is consistent with the cited literature, where the measure of inequality is almost always the Gini coefficient.

In preliminary analyses we use a reduced form of our model in (4.9) where only one regressor is used. We check if the single regressor has a significant regression coefficient. All regressions show a $R^2 < 13\%$ with 271 to 280 observations, depending on the regressor. In the following we report the cases, where we find deviations form the literature, concerning both sign of the coefficient or the significance.

- corruption:
  The positive effect, cited in the literature remains in our analysis, but we cannot reject the Null on a 10% significance level.

- education expenditure
  The negative effect, cited in the literature remains in our analysis, but we cannot reject the Null on a 10% significance level.

- foreign direct investment
  The positive effect, cited in the literature is not found in our analysis. We find negative coefficients, which are not significant on a 10% level.

- fraction of rural population

Table 4.4: Independent variables, together with the corresponding literature or hypothesis and the data-source of the used data, grouped by the hypotized correlation with the inequality measures.

Indep. Variable	Abbr.	Literature / Hypothesis	Data
		positive effect	
Corruption	co	Gupta, Davoodi, and Alonso-Terme (2002)	Transparency International (2016)
Finacial depth	fd	Li, Squire, and Zou (1998)	WB[2] (FS.AST.PRVT.GD.ZS)
Foreign direct investment	fdi	Jaumotte, Lall, and Papageorgiou (2013)	WB (BX.KLT.DINV.WD.GD.ZS)
Fraction 18 to 24 year old people having left school early (ISCED ≤ 2)	edu	Nielsen and Alderson (1995)	ES[3] (tsdsc410)
High-tec exports	hte	Jaumotte, Lall, and Papageorgiou (2013)	WB (TX.VAL.TECH.MF.ZS)
Inflation rate	ir	Aleš (2001)	ES (tec00118)
Population growth	pg	Nielsen and Alderson (1995)	ES (tps20006)
Unemployment rate	ur	Malinoski (2012)	WB (SL.UEM.TOTL.ZS)
		no effect	
GDP per capita	gdp	Adams Jr. (2003)	ES (tec00114)
KOF index of globalization	kof	Jaumotte, Lall, and Papageorgiou (2013)	Dreher (2006) and Dreher, Gaston, and Martens (2008)
		negative effect	
Average of highest education (ISCED)	edun	Nielsen and Alderson (1995)	EU-SILC (PE040)
Education exprenditure	edue	Nielsen and Alderson (1995)	WB (NY.ADJ.AEDU.GN.ZS)
Final consumption expenditure of general government	ceg	Mello (2006)	ES (tec00010)
Fraction of rural population	frp	Nielsen and Alderson (1995)	WB (SP.RUR.TOTL.ZS)
Government debt	gd	Higher debt leads to less money available for social transfers	ES (tsdde410), WB (GC.DOD.TOTL.GD.ZS), Eidgenössische Finanzverwaltung (2017)
International trade balance	itb	Jaumotte, Lall, and Papageorgiou (2013)	ES (tet00002, tec00001)
Life expectancy	life	Malinoski (2012)	WB (SP.DYN.LE00.IN)
People aged 64 or more over people aged 15-64	age	Higher average incomes at the end of working life, thus more pensioneers reduce high incomes	ES (tsdde510)
Personal freedom	pf	Li, Squire, and Zou (1998)	Gwartney, Lawson, and Hall (2015)
Tax revenue	tr	Lee (2005)	WB (GC.TAX.TOTL.GD.ZS)
Trade in percent of GDP	tpc	Jaumotte, Lall, and Papageorgiou (2013)	WB (NE.TRD.GNFS.ZS)

The negative effect, cited in the literature is not found in our analysis. We find positive coefficients, which are significant on a 1% level.

- government consumption expenditure
  The negative effect, cited in the literature is only found for the v metric and not for the d metric. Both coefficients are insignificant on a 10% level.

- personal freedom
  The negative effect, cited in the literature is not found in our analysis. We find positive coefficients, which are not significant on a 10% level.

- population growth
  The positive effect, cited in the literature remains in our analysis, but we cannot reject the Null on a 10% significance level.

- tax revenue
  The negative effect, cited in the literature is not found in our analysis. We find positive coefficients, which are insignificant on a 10% level.

- unemployment rate
  The negative effect, cited in the literature is not found in our analysis. We find positive coefficients, which are significant on a 1% level.

In the following the model in equation (4.9) is estimated and the starting point for further analysis. We analyze, if the significance from the single model before translate to the more complex model. Further we reduce the complexity of the model by eliminating insignificant variables. With this procedure we end up with 6 models, which are shown in table 4.5

Turning to model (1), we find for unemployment rate (ur), trade (tpc), financial depth (fd), education level (edun), age structure (age) and life expectancy (life) no significant coefficients for both the v-index nor d-index, although significant coefficients have been found in the reduced model before. The opposite pattern show population growth (pg), globalization (kof), foreign direct investments (fdi) and personal freedom (pf): Whereas they have not been significant in the preliminary analysis above, in model (1) we find significant coefficients at least at the 10% level. In the following we exclude the variables GDP (gdp), tax revenue (tr), government consumption expenditure (ceg),

Table 4.5: Standardizes regression coefficients for models (1) to (6), using a within-estimator. The dependent variable is denoted at the top and the corresponding t-values are given in brackets.

Var	Models					
	(1)		(2)		(3)	(4)
	v	d	v	d	v	d
age	-0.01	0.05	0.02	0.04		
	(-0.06)	(-0.36)	(0.11)	(0.35)		
ceg	-0.09	0.00				
	(-0.66)	(-0.05)				
co	-0.12	-0.09				
	(-0.73)	(-0.77)				
edu	0.19**	0.24***	0.19**	0.24***		0.16***
	(1.69)	(3.24)	(1.78)	(3.42)		(2.50)
edue	-0.21	-0.15				
	(-0.92)	(-0.96)				
edun	0.15	0.14	0.21	0.17*		
	(-0.82)	(-1.09)	(1.13)	(1.37)		
fd	0.06	-0.05	0.08	-0.04	0.21***	0.09**
	(-0.51)	(-0.60)	(0.77)	(-0.60)	(2.70)	(1.70)
fdi	-0.05**	-0.03*	-0.06**	-0.03*	-0.06**	
	(-1.67)	(-1.41)	(-1.73)	(-1.45)	(-2.08)	
frp	-1.04**	-0.66*	-1.09**	-0.58*	-0.87**	
	(-1.68)	(-1.57)	(-1.89)	(-1.49)	(-1.68)	
gd	-0.49***	-0.43***	-0.43***	-0.43***	-0.45***	-0.27***
	(-3.60)	(-4.70)	(-3.41)	(-5.01)	(-5.07)	(-4.48)
gdp	-0.45	-0.09				
	(-1.19)	(-0.34)				
hte	0.32***	0.20***	0.32***	0.19***	0.29***	0.15***
	(4.28)	(3.88)	(4.47)	(4.02)	(4.10)	(3.28)
ir	0.07**	0.07***	0.06**	0.06***	0.08***	0.05***
	(1.81)	(2.75)	(1.71)	(2.72)	(2.61)	(2.46)
itb	-0.16	-0.18**	-0.16*	-0.18**		-0.15***
	(-1.16)	(-2.00)	(-1,38)	(-2,26)		(-2.34)
kof	0.25**	0.10	0.18*	0.08		
	(1.84)	(-1.12)	(1.47)	(0.91)		
life	0.24	0.06	0.10	0.01		
	(0.77)	(0.29)	(0.33)	(0.07)		
pf	-0.15**	-0.14**	-0.16**	-0.15***	-0.15**	-0.16***
	(-1.68)	(-2.26)	(-2.19)	(-3.22)	(-2.24)	(-3.59)
pg	-0.28***	-0.20***	-0.26***	-0.19***	-0.21***	-0.14***
	(-3.51)	(-3.74)	(-3.40)	(-3.64)	(-3.15)	(-3.34)
tpc	-0.38*	-0.24*	-0.29*	-0.23**		
	(-1.64)	(-1.52)	(-1.43)	(-1.72)		
tr	0.01	-0.04				
	(-0.12)	(-0.69)				
ur	-0.08	0.00	-0.05	0.02		
	(-1.00)	(0.00)	(-0.66)	(0.30)		
year	0.07	0.59	0.03	0.06		
	(-0.62)	(0.81)	(0.33)	(0.91)		
Within-R^2	0.32	0.39	0.30	0.38	0.25	0.33
corr. Within-R^2	0.26	0.31	0.24	0.31	0.22	0.28
obs	263		268		270	270

Note: $^{***}p < 0.001$, $^{**}p < 0.005$, $^{*}p < 0.01$

corruption (co) and education expenditure (edue), which do not show a significant coefficient. This is in line with our preliminary analysis from above. The other variables show in both in model (1) and the preliminary reduced models significant coefficients with at least one of the v- oder d-index. The year variable has no significant coefficient. The estimates of the resulting models in shown in column (2) of table 4.5. The sign of the coefficients does not change compared to model (1) and the size of the effect remains roughly the same. In some cases, the significance level changes, e.g. we find a significant coefficient for education (edun) on a 10% significance level for the d-index, which has been insignificant in model (1).

In the following, we estimate models for both indices separately. Both have highly significant coefficients for government debt (gd), population growth (pg) and high-technology exports (hte) in common. Thus they are included in both models. The model (3) for the v-index additionally includes foreign direct investments (fdi), inflation rate (ir), financial depth (fd), fraction of rural population (frp) and personal freedom (pf). The coefficients do not change their sign and remain roughly constant compared to models (1) and (2), except for financial depth (fd). Whereas it used to be insignificant but positive, we now can reject the Null at a 1% level. The size increased from 0.06 to 0.21, thus by a factor of 3.5. Although having reduced the model, we find a comparable corrected Within-R^2 (21.75% compared to 25.56% in model (1)) and all coefficients are significant at least on the 5% level. The last column of table 4.5 shows the model for the d-index. Additional to the variables stated above (gd, pg, hte), inflation rate (ir), financial depth (fd), personal freedom (pf), international trade balance (itb) and early school leavers (edu) are included. Again, we find comparable results to the models (1) and (2) concerning the sign and size of the coefficients. In the case of financial depth, where the coefficient used to be negative, but insignificant, it is not positive on a significance level of 5%. All coefficients are now significant at least on the 5% level.

Since we found a rather strong correlation between the Gini and the d-index (see 1.3.1) we also proxy the Gini coefficient with our model (4). Thus, we find some inconsistencies with the corresponding literature. Although the consumption of the government (ceg) is said to have a negative effect on inequality, we cannot find a significant coefficient for our d-index, and thus for the Gini. The same holds for education expenditure (edue), education level (edun), fraction of rural population (frp), life expectancy (life), trade (tpc) and tax returns (tr). For variables with the hypothesis of a positive effect, we do not find this for foreign direct investments (fdi), corruption (co) and unemployment rate (ur).

One possible explanation could be that our first estimate of the true underlying income distribution is more accurate, due to the granularity of the EU-SILC data. This could lead to different Gini estimates, which translate to different estimated coefficients. Another possibility could be that there has been a structural break in what drives income inequality during the last decade, especially during the financial crisis. The latter itself could be an explanation for the different results. The financial crisis covers affects most of our data and thus is very prominent in our results, which is not necessarily the case in the literature, where older data were used.

Comparing the results from model (4) to the ones of model (3), we find consensus for certain variables. This is the case for e.g. inflation rate (ir) or personal freedom (pf). Here we find roughly the same coefficients. These effects explain a common movement in the time-series (see figure 4.2). The differences come from variables like the fraction of rural population, which is significant for v and a rather large effect, but not for d. Another example is government debt (gd), where the effect is roughly double the size for one of the metrics.

4.4 Concluding remarks

The asymmetry - spread phase plane diagrams of figure 4.4 and figureB.1 in the appendix reveal significant differences over the data horizon. It is easy enough to rationalize the experience of some countries, Norway being an obvious example with the strength of oil prices over the interval. Likewise, joining the EU has on the face of it been good, with fallers such as the Czech Republic or Hungary reflecting rising employment and wages as tariffs are dismantled and trade opened up.

But other comparisons are more problematical, even those that one might expect to be similar. Why is France a riser, which one might consider bad, but neighboring Germany is not, but corresponds to the mixed group? Likewise both Spain and Portugal suffered from the GFC, but why is Spain a v-d riser while the Portugal is a faller? We use a fixed-effects estimation to give some insights what causes the movements in the asymmetry - spread plane. Since our d metric correlates strong with the Gini coefficient, we compare the results with the literature covering the inequality measured by the Gini.

We show that some hypothesis do not hold and rather small models have roughly the same information content. Also we find common drivers of inequality, both in the v and d models. Structural factors, which can hardly be measures, such as impediments to the adjustment of wages or employment might have been of more importance in countries such as France, with its public sector rigidities and play an important role as well.

To summarize, the metrics and phase diagrams do not in themselves resolve issues of this kind. But in throwing up differences in such a dramatic way, they do motivate the search for answers on a more structural level. Over time there will arise fresh circumstantial challenges, whether these arise from global fallouts, localized trading blocs, or political events on a national level. If nothing else, the message of the present paper is that one way or another, these do have an observable welfare impact on the economic shares of society.

5 Asymmetry and performance metrics for financial returns

The following is based on Bowden, Posch, and Ullmann (2017a)

Performance metrics for equities or managed funds are established tools of the finance industry, objects of media reporting and investor assessment. The metrics in question are usually comparative in nature against some benchmark. For example, the Sharpe (1966) index compares the mean of the security return against the risk free rate, effectively the expected return on a portfolio long in the security return, short in the risk free rate. Likewise, Jensen's alpha looks for positive discrepancies relative to the beta adjusted market return, provided one is confident that a CAPM style equilibrium actually exists (Jensen 1968). Even short of this, however, there are some hidden assumptions in the use of standard metrics. In the case of the Sharpe index, it is tacitly assumed that the distribution of equity returns is symmetric. But this is not necessarily true, even in relatively normal times (Fama 1965; Chunhachinda et al. 1997). If returns are asymmetric over any given interval, then extraordinary exposures to loss, or else opportunities for gain, can arise. Kraus and Litzenberger (1976) proposed skewness as a second factor in the traditional Sharpe-Lintner CAPM (Sharpe 1966; Lintner 1965). Ang and Chua (1979) use this modification in order to define an excess return performance measure. For other approaches incorporating skewness, see Eling and Schuhmacher (2007) and Farinelli et al. (2008).

Any such diagnostic metric should be contextual in nature, with reference to the investor's gain or loss, in a way that the textbook third order moment for asymmetry cannot. And there is room for exploration as to just what should be assumed, even tacitly, about the underlying utility function. A case could be made that the standard concave (risk-averse) utility function is not a comprehensive representation for investor motivation. More appropriate might be a utility function that is concave on the downside and convex on the upside (Friedman and Savage 1948; Kahneman and Tversky

1979). Thus a fund manager would be averse to losses, the more so with the prospect of employment termination if losses are heavy. On the other hand, the same manager will be motivated by a more generous performance bonus for progressive gains on the upside.

The purpose of the present paper is to introduce a performance function that addresses such objectives. The methodology is an adaptation of a metric for economic disadvantage recently introduced in the context of income distribution by Bowden (2016b). The adaptation to asset returns adjusts for differing motivational perspectives; too many higher incomes are considered bad for social welfare, but high returns are good for investors! A further normalization adjustment allows a ready comparison with the corresponding Sharpe metric. The two are the same wherever the return distribution is symmetric; but they differ where this is not the case. Relative to the proposed W-metric, the Sharpe metric understates the investor gains when returns are negatively skewed, i.e. where more distributional weight accrues to higher values, with the reverse effect where the distribution is positively skewed. An empirical analysis shows that the same understatement occurs with a range of other performance metrics that have been proposed, both through time and across an extended range of equities. In this sense the proposed W-metric adds a more embracing dimension to fund performance comparisons.

5.1 Metrics and investor utility: the conceptual framework

The development that follows is general in nature. The basic framework of an investor profit or loss function based on upper and lower conditional expectations is exposited on a general level, followed by the connection with Friedman-Savage decision theory. Taking expected values results in the proposed metric for returns asymmetry, with an accompanying dual metric for spread or dispersion.

5.1.1 The $\underline{w}$in function

The proposed metric can be cast in terms of the expected value of a profit or loss function. Let x represent any outcome (e.g. income, return or any relevant measurable outcome). The associated random variable X has distribution function $F(x)$ with density $f(x)$. For expositional purposes assume $-\infty < x < \infty$ but the treatment can adapt to the half line $0 < x < \infty$ or any compact range. Since our proposed measure is based on the first conditional moment, we shall always assume the existence of the first moment of the associated random variable X. For most practical purposes, this is a weak assumption but it does rule out exceptionally long tailed distributions e.g. the Levy. But it suffices for existence of the proposed skewness parameter for empirical density functions, as well as for random variables for which higher moments do not exist (e.g. a t-distribution with $df = 2$).

For a given value $X = x$, define the progressive conditional expected values as

$$\mu_l(x) = \mathbb{E}_F\left[X|X \leq x\right] = \frac{1}{F(x)} \int_{-\infty}^{x} X \mathrm{d}F(X) \quad \text{(left conditional expectation value)} \tag{5.1}$$

$$\mu_r(x) = \mathbb{E}\left[X|X > x\right] = \frac{1}{1 - F(x)} \int_{x}^{\infty} X \mathrm{d}F(X) \quad \text{(right conditional expectation value)} \tag{5.2}$$

As $x \to \infty$, $\mu_l(x) \to \mu$ and $\mu_r(x) \to x \to \infty$, while for $x \to -\infty$, $\mu_l(x) \to x \to -\infty$ and $\mu_r(x) \to \mu$. Figure 5.1 illustrates with a logistic distribution, which has fat tails relative to the normal distribution; in this case the asymptotic convergence is very slow. Expression (5.1) as the conditional lower moment has found application in fund performance (see section 5.2.2). In contrast, the present application utilizes both upper and lower conditional moments, with a further stage of consolidation into a single valued metric. The function $v(x) = (\mu_r(x) - x) - (x - \mu_l(x))$ is referred to in the recent income distribution literature as the net economic disadvantage function. In this context, suppose my income is x. I feel disadvantaged to the extent of the average income above mine $(\mu_r(x) - x)$, but better off than those below me $(x - \mu_l(x))$.

In an investment context, where x refers to returns, things have to be reversed. On any given day suppose the return is x. Now I win to the extent that this exceeds the average return below (other days are worse), and I lose to the extent that it falls short of the average above (other days are better). This indicates that a more appropriate welfare function in the finance context is the negative of the net economic disadvantage function, i.e.

$$w(x) = -v(x) = (x - \mu_l(x)) - (\mu_r(x) - x)\,, \tag{5.3}$$

the '<u>w</u>in' function. The function $w(x)$ is concave below (convex above) a break-even value on the return axis. As $x \to -\infty$ it becomes asymptotic from above to the 45 line, and from below as $x \to \infty$. The zero point, where it crosses the x-axis, is distribution specific (see figure 5.1). An indication of its position can be inferred from the relationship

$$w(x_m) = 2\,(x_m - \mu)\,, \tag{5.4}$$

where $\mu = \mathbb{E}\,[x]$ is the expected value and x_m is the median. Thus if $\mu < x_m$, which will usually be the case with negative skewness, then has its zero to the left of the median, conversely if $\mu > x_m$. For a symmetric distribution crosses at the common mean and median, while the function itself is anti-symmetric about the x-axis. The logistic is an example as in figure 5.1.

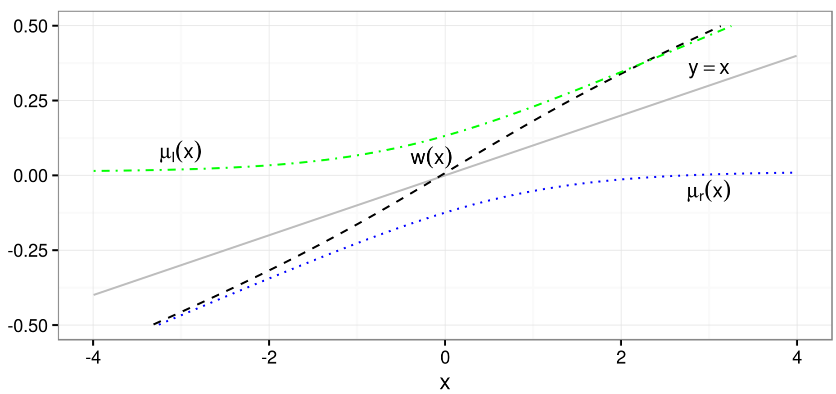

Figure 5.1: The $w(x)$ function. The function $w(x)$ illustrated with a logistic distribution. For clarity, the values have been scaled.

As earlier remarked, lower concavity versus upper convexity is a universal property (Bowden 2016a, 2016b). This motivates the description of the function $w(x)$ as a Friedman-Savage utility function. However, there are some qualifications as to scope. The classical utility function is distribution invariant, but in the present context the function $w(x)$ depends on the conditional expectation functions $\mu_l(x), \mu_r(x)$ and therefore upon the distribution $F(x)$. In this sense it could be interpreted as supplementary guide to decision making. Faced with a given returns distribution, the investor could use it to decide whether or not the implied $w(x)$ utility function is consistent with his or her own preferences as between profits and losses.

5.1.2 The expected value

The expected value of the function $w(x)$ results in a single valued metric for investor's profit or loss:

$$w = \mathbb{E}\left[w(x)\right] = \int_{-\infty}^{\infty} w(x)\mathrm{d}F(x). \tag{5.5}$$

One could imagine an investor selecting a day at random and deriving a return x on that day. Over any such random choice of day, expression (5.5) gives the expected gain or loss. The integral in (5.5) represents a double smoothing process. The first smoothing level (expressions (5.1), (5.2)) is to obtain the conditional expected value to the left and right of any chosen value x. The second is to smooth the results over all values of x.

However, the two stages can be condensed into a simple one stage expectation by means of a change of measure. The left and right entropic shifts of $F(x)$ are distributions defined by

$$F_L(y) = F(y)\left(1 + \xi_L(y)\right); \quad 1 - F_R(y) = \left(1 - F(y)\right)\left(1 + \xi_L(y)\right), \tag{5.6}$$

where the shift factors (technically Radon-Nikodym derivatives) are given by

$$\xi_L(y) = -\ln F(y); \quad \xi_R(y) = -\ln\left(1 - F(y)\right) \tag{5.7}$$

Properties of the left and right entropic shifts can be found in Bowden (2012, 2016b). In such terms,

$$\mathbb{E}\left[\mu_l(x)\right] = \int_{-\infty}^{\infty} \mu_l(x)\mathrm{d}F(x) = \int_{-\infty}^{\infty} x\mathrm{d}F_L(x) = \mu_L, \tag{5.8}$$

the simple mean of the left entropic shift. Similarly,

$$\mathbb{E}\left[\mu_r(x)\right] = \int_{-\infty}^{\infty} \mu_r(x)\mathrm{d}F(x) = \int_{-\infty}^{\infty} x\mathrm{d}F_R(x) = \mu_R, \tag{5.9}$$

the simple means of the shifted distributions defined by expressions (5.6), (5.7). Expressions (5.8), (5.9) embody the double smoothing property. The first layer is the progressive smoothing up to or beyond any given point x. The second layer smooths the result over all such points, using the density as a weighting kernel.

The expected value of the net '$\underline{w}$in' function is therefore

$$w = 2\left(\mu - \frac{1}{2}\left(\mu_L + \mu_r\right)\right) \tag{5.10}$$

An implicit asymmetry test is apparent in the form of (5.10). The entropic shifts F_L, F_R represent bodily displacements to the left and right of the natural distribution $F(x)$. If more entropic (uncertainty) mass is contained to the left, then the displacement difference is greater than to the right and the average of the two means will be less than the original. Thus a distribution $F(x)$ that is naturally negatively skewed will have $w > 0$, while positive skewness will lead to $w < 0$. For any symmetric distribution, $w = 0$. Outcomes of this kind correspond to classical metrics such as the Pearson third order moment. Unlike the standard cubic, however, there is now an explicit contextual reference, as to how investors might view things in welfare or utility terms.

5.1.3 An associated dispersion metric

A simple change of sign in expression (5.3) provides a penalty function for spread or dispersion (Bowden 2016b). This is defined as

$$d(x) = (x - \mu_l(x)) + (\mu_r(x) - x) = \mu_r(x) - \mu_l(x). \tag{5.11}$$

It follows from their respective definitions that $d(x) > w(x)$.

Taking the expected value gives a metric for dispersion:

$$d = \mathbb{E}\left[d(x)\right] = \mu_R - \mu_L > 0. \tag{5.12}$$

The underlying dispersion concept is based on the total partition entropy of the distribution. The partition entropy function $h(x)$ at any chosen point represents the uncertainty as to whether $X > x$ or $X < x$. The partition entropy function has itself given rise to a number of applications in areas such as a smoothing kernels or scaling algorithms. Integrating $h(x)$ over x gives the metric d, which can be interpreted as the total entropic spread of the distribution. There is also a covariance connection with asymmetry, as measured by the metric w. A distribution is negatively skewed if the dispersion function $d(x)$ is greater in the region where $F(x) < 0.5$, so the covariance between $d(x)$ and $F(x)$ is negative.

5.2 Performance metrics

The foregoing definitions and interpretations are employed in the present section to define and motivate the proposed performance metric as an alternative to the standard Sharpe measure. Section 5.2.2 introduced existing performance metrics, which will be used to compare our analysis to Eling and Schuhmacher (2007) in the fist place.

5.2.1 The W-metric

The basic Sharpe performance index serves as a starting point for a more general discussion. As we are focusing on investment performance, the general outcome x is replaced by the context specific return r. The standard Sharpe metric is written as

$$S(r) = \frac{\mathbb{E}\left[r\right] - r_f}{\sigma\left(r\right)}, \tag{5.13}$$

where $\mathbb{E}\left[r\right]$ is the expected return, σ its standard deviation and r_f is a risk free rate. Defining $\tilde{r} = r - r_f$ as the excess return, with distribution function $\tilde{F}$, the Sharpe ratio becomes

$$S\left(\tilde{r}\right) = \frac{\mathbb{E}_{\tilde{F}}\left[\tilde{r}\right]}{\tilde{\sigma}}, \tag{5.14}$$

where $\tilde{\sigma}$ denotes the standard deviation of excess returns.

The proposed W-metric is an alternative to the Sharpe ratio that explicitly encompasses the positive or negative benefits of asymmetry. It is defined as

$$W(r) = \frac{w(\tilde{r}) + \tilde{r}_m}{d(\tilde{r})} := \frac{\tilde{w} + \tilde{r}_m}{\tilde{d}}, \tag{5.15}$$

where $\tilde{r}_m$ is the median of the excess return distribution and w, d are defined according to expressions (5.10) and (5.12). The denominator d of the W-metric is adapted to the entropic interpretation of the asymmetry metric w in the numerator. For standardizable distributions such as the Normal, the logistic or the Gumbel, the metrics d and the standard deviation σ are proportional via the scale parameter. In such cases there is little effective difference between using either d or σ for the denominator.

In our empirical analysis we will also use the market return as a comparator. This is indicated by denoting the modified W as W^* and the Sharpe ratio S^* accordingly.

The numerator of the metric (5.15) splits into two terms: $\tilde{w}$ captures the asymmetry of the excess returns distribution while $\tilde{r}_m$ adds the median excess return. Rewriting the numerator of W results in:

$$\tilde{w} + \tilde{r}_m = (\mathbb{E}[r] - r_f) + (\tilde{w} + r_m - \mathbb{E}[r]) \tag{5.16}$$

Thus the W-metric and the Sharpe ratio's numerator differ in the term $[\tilde{w} + r_m - \mathbb{E}[r]]$. Consider the following cases:

(a) The distribution of excess returns $\tilde{r}$ is symmetric. In this case $\tilde{w} = 0$, $r_m = \mathbb{E}[r]$, and the numerators of the W-metric and Sharpe are identical.

(b) The distribution of excess returns is positively skewed. In this case $\tilde{w} < 0$, and it is likely that $r_m - \mathbb{E}[r] < 0$. Together this means that $W < S$.

(c) The distribution of excess returns is negatively skewed. In this case, $\tilde{w} > 0$, and it is likely that $r_m - \mathbb{E}[r] > 0$. Together this means that $W > S$.

Cases (b) and (c) reflect the implied Friedman-Savage investor utility basis: investors would like a positively skewed distribution; increasing marginal utility in the higher zone. But they might back away from a negatively skewed one; too much weight in the low zone, the area of more negative marginal utility.

The difference between W and S is generated by the implicit underlying utility functions, and the way that these are responsive to the distribution of returns. The

Sharpe metric S tacitly assumes a linear utility function, while the W metric is more responsive to the mixed concave-convex Friedman-Savage type utility function.

Adaptations of W can be devised for other benchmarks. Instead of using the risk free rate, the comparator could be the market return R, so $\tilde{r} = r - R$ is a compound return, long in the subject security and short in the market. Alternatively if a CAPM is thought to exist, a generalization of Jensen's alpha can be defined with the comparator return as the market, scaled by the security's beta.

5.2.2 Other performance metrics

The empirical work of section 5.3 encompasses a range of alternative performance metrics that have been proposed in the literature. We use add the following ones to compare our results to Eling and Schuhmacher (2007), Zakamouline (2009), and Ornelas, Silva Júnior, and Fernandes (2012). Whereas the former finds no difference in rankings for these performance metrics, the latter two do.

One of them, namely the Treynor ratio $\left(\mathbb{E}\left[r\right] - r_f\right)/\beta$ requires a more or less explicit CAPM reference.

Of those that do not, the Sortino and Kappa 3 ratios ("The Dutch Triangle" 1999; Kaplan and Knowles 2004) have a point of connection in that in drawing on lower and upper conditional partial moments. Given a threshold return z, the lower partial moments of order n with the corresponding threshold return z are defined (c.f. Harlow 1991) as

$$LPM_n(z) = \frac{1}{F(z)} \int_{-\infty}^{z} (z - x)^n \,\mathrm{d}F(x) \tag{5.17}$$

For the Sortino and Kappa 3 ratios:

$$Sortino(z) = \frac{\mathbb{E}\left[r\right] - z}{\sqrt[2]{LPM_2(z)}} \quad \text{resp.} \quad K_3(z) = \frac{\mathbb{E}\left[r\right] - z}{\sqrt[3]{LPM_3(z)}}. \tag{5.18}$$

The upside potential ratio (UPR) employs the higher partial moment of order one:

$$UPR(z) = \frac{\frac{1}{F(z)} \int_{z}^{\infty} (z - x) \,\mathrm{d}F(x)}{\sqrt[2]{LPM_2(z)}}. \tag{5.19}$$

All versions require the user to specify a benchmark value z, so these are not global metrics.

A further comparator is the omega ratio of Keating and Shadwick (2002). In the present notation this can be expressed as

$$\Omega(z) = \frac{\int_z^\infty (1 - F(x))\,\mathrm{d}F(x)}{\int_{-\infty}^z F(x)\mathrm{d}F(x)} := \frac{\Phi_r(z)}{\Phi_l(z)} \tag{5.20}$$

The function $\Phi_l(z)$ will be familiar from stochastic dominance theory: distribution A is second order stochastic dominant over B if $\Phi_{l,A}(z) \leq \Phi_{l,B}(z)$ for all z. Keating and Shadwick propose a less stringent version: the investor sets a threshold value of z, below which outcomes are regarded as losses and above as gains and compares the ratio (5.20) relative to that threshold.

There is no direct connection between the Omega ratio and the present W-metric, though the functions $w(x), d(x)$ utilized in the latter help to determine the shape of $\Omega(x)$ via the differential equation

$$\frac{\mathrm{d}\log\Omega(x)}{\mathrm{d}x} = -4\frac{d(x)}{d^2(x) - w^2(x)} \tag{5.21}$$

The omega ratio function is evidently a compounding of both spread and asymmetry, with the leading effect of order $d^{-1}(x)$. If the dispersion is high, a unit step to the right on the return axis does not increase the probability for losses that much, compared to a very dense distribution, where a unit step covers more probability mass.

A final group of comparators utilizes the notion of maximum drawdown (MD) (Magdon-Ismail et al. 2003). A drawdown is the maximal loss in a given time span with a designated end date and a variable start date within the given time span. The maximum drawdown denotes the maximum of the drawdowns within the given time span. Three variants utilized in the empirical comparisons of section 5.3 are the Calmar ratio (Young 1991), the Sterling ratio (Kestner 1996) and the Burke ratio (Burke 1994), with the following definitions:

$$Calmar\ ratio(r) = \frac{\mathbb{E}\left[r\right] - r_f}{-MD} \tag{5.22}$$

$$Sterling\ ratio(r) = \frac{\mathbb{E}\left[r\right] - r_f}{\mathbb{E}\left(-MD\right)} \tag{5.23}$$

$$Burke\ ratio(r) = \frac{\mathbb{E}\left[r\right] - r_f}{\sqrt{\sum -MD}} \tag{5.24}$$

5.3 Empirical application

The empirical work that follows proceeds in two phases. Section 5.3.1 looks at the relative performance of the W-metric with the conventional Sharpe ratio as the point of departure. For this more limited comparison, historical returns on just a single equity are used. Section 5.3.2 is an extended set of comparisons with the range of alternative metrics as in section 5.2.2, applied to the complete regimen of equities in the S&P500, a sample of US corporate bonds, a sample of exchange rates and a sample of hedge funds. All data are obtained by Thomson Reuters Datastream.

5.3.1 Comparison with the Sharpe ratio

To illustrate the new measure, we use the daily log returns of Ford Motors Company together with the S&P500 index as a market proxy. By doing so, we find a rather long time series of historical returns. The market yield on U.S. Treasury securities at 1-year constant maturity is employed as a proxy for the risk free rate from 1973 to 2016. Table 5.1 gives descriptive statistics. For all the different log return distributions, any evident skewness has the sign expected in the w statistic of equation (5.5)

Turning to explicit performance indicators, the Sharpe ratios and the W metrics are compared for two benchmarks: the risk free rate and the market return. The Sharpe and W measures are calculated using daily returns over each different year. The results are shown in figure 5.2. The upper part utilizes the risk free rate as benchmark as in the classic Sharpe ratio. The gray bar corresponds to W in equation (5.15) and the black bar to the Sharpe ratio. The two measures have a correlation coefficient of $\rho_{S,W} = 0.315$ at a significance level of $\alpha = 5\%$. There are several occurrences where not

Table 5.1: Descriptive statistics and the corresponding w measure for the daily returns distribution of Ford Motors Company. All values are based on log-returns as percentage values.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
			Stock			
Mean / %	-0.47	-0.05	0.01	0.03	0.14	0.56
Median / %	-0.64	0.00	0.00	-0.01	0.04	0.29
Var. / %	0.01	0.03	0.04	0.05	0.06	0.37
Skew.	-1.95	-0.21	0.04	-0.00	0.30	2.28
Ex. Kurt.	-0.70	0.44	1.04	2.22	2.53	15.88
			Market			
Mean / %	-0.23	-0.00	0.04	0.02	0.07	0.12
Median / %	-0.20	-0.00	0.06	0.04	0.10	0.19
Var. / %	0.00	0.00	0.01	0.01	0.01	0.06
Skew.	-4.94	-0.37	-0.08	-0.26	0.06	0.80
Ex. Kurt.	-0.41	0.42	1.32	2.94	1.90	52.95

only the absolute values between the Sharpe and the W-measure differ, but also the signs.

As discussed in section 5.2, one can look for two scenarios where W is potentially unequal to the Sharpe ratio S. In the sample there are 23 years with a positive skewness, of which 20 years confirm the prediction $S > W$. The empirical hit ratio of $r_m - \mathbb{E}[r] < 0$ is 52.2%. In the case of negative skewness, the $W > S$ prediction is confirmed in 16 out of 21 years. The empirical hit ratio of $r_m - \mathbb{E}[r] > 0$ with $\tilde{w} > 0$ is 75.0%. Changing the comparator from the risk free rate to the market return, we change the notation $W \to W^*$ and $S \to S^*$. The lower chart of figure 5.2 shows the results. The grey bar corresponds to W^* and the black bar to the Sharpe analogue ratio S^*. In 18 out of 44 observations the two have different signs along with an correlation coefficient of $\rho_{S^*,W^*} = 0.266$, significant at the 10% level. For the positive skewness we find $S^* > W^*$ in 16 out of 23 years with an empirical hit ratio of $r_m - \mathbb{E}[r] < 0$ is 75.0% therein. In the 21 years with a negative skewness, 16 years have $W^* > S^*$ and therein we find an empirical hit ratio of 75.0% for $r_m - \mathbb{E}[r] < 0$.

As previously discussed, the standard deviation can be used as denominator instead of the dispersion metric d, with the performance indicator as $W_S = \frac{\tilde{w}+\tilde{r}_m}{\tilde{\sigma}}$. The analysis is then repeated using W_S instead of W. The result is shown in figure 5.3. In all cases,

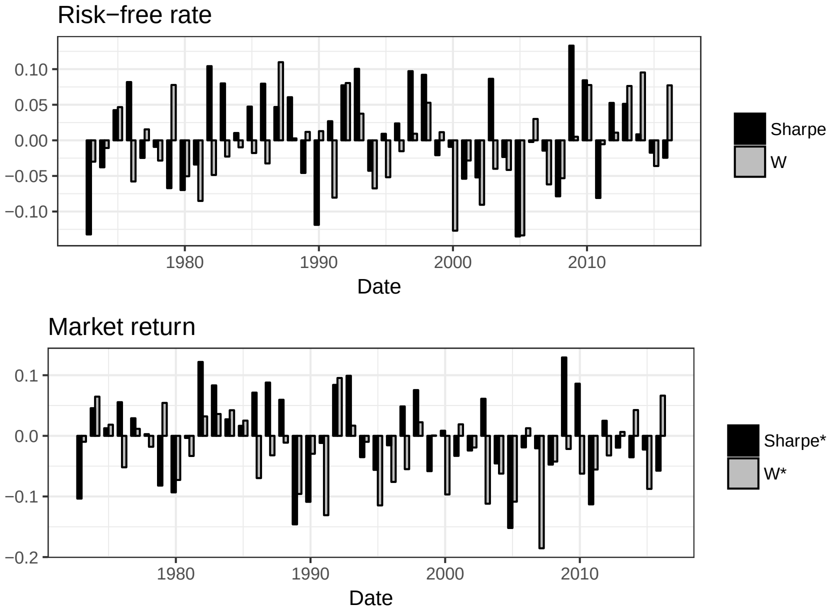

Figure 5.2: W-metric and Sharpe Ratio. Comparison of the W-metric and the Sharpe Ratio for the risk-free rate and the market return as comparator benchmarks.

the sign is preserved, i.e. $\text{sgn}(W) = \text{sgn}(W^*)$, but the absolute value of W_S is always larger. Using the standard deviation therefore leads to a higher value of the metric. For the upper chart, there are three cases where the alternative definitions lead to different conclusions as to the size of the metric. In 1980, 1998, 2001, 2005, 2008 and 2010 we find $|W| < S < |W_S|$. For the lower part of figure 5.3 the years 1989, 2005 and 2008 show a similar pattern.

5.3.2 Other literature metrics: a more extended comparison

Turning to the more extended set of measures described in section 5.2, performance comparisons utilize daily returns on four different asset classes. In the following, we do not focus on the individual metrics, like the Sharpe Ratio in the former paragraph, but rather on the rank correlation between these metrics. The general idea comes from the hedge fund industry, where the performance of a manager does not depend solely on the value of the metric, but also on the ranking with peers. The question whether the ranking differs for different performance metrics has been answered in both directions (Eling and Schuhmacher 2007; Zakamouline 2009; Ornelas, Silva Júnior, and Fernandes

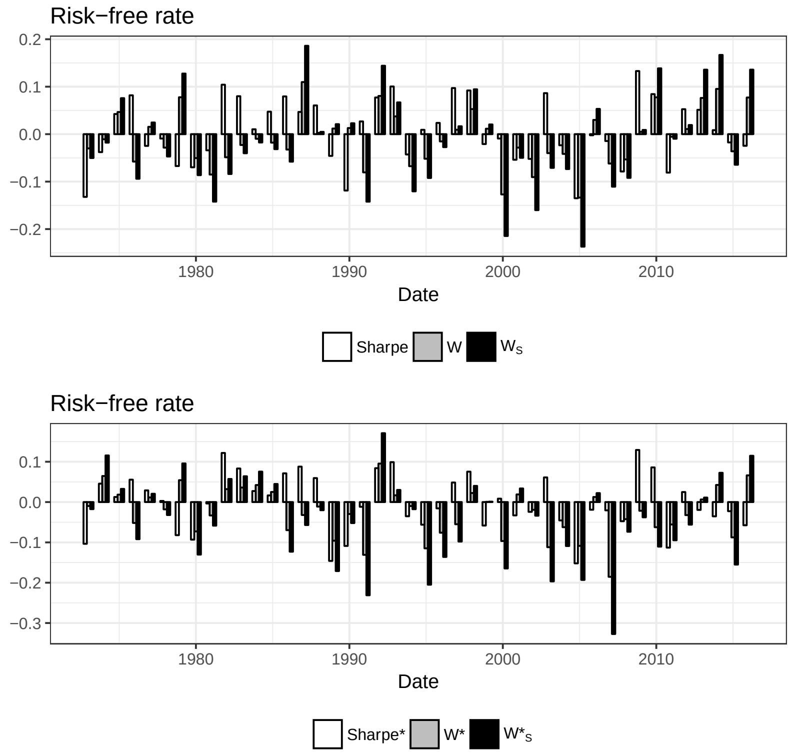

Figure 5.3: W-metric, W_S-metric and Sharpe Ratio. The W-metric, the Sharpe Ratio and the W_S-metric with SD as denominator, and (a) the risk-free rate and (b) the market return as benchmarks.

2012). In our analysis, we use

- stocks, where our universe exists of the constituents of the S&P500 as of November 11th 2015. Starting in 2000, we ignore the stocks, where missing data occur and end at the end of 2016. In these 17 years, we observe 7179 daily returns distributions.
- bonds, where our universe exists of all US corporate bonds available in Thomson Reuters Datastream. We cover the timespan from 2010 to 2015 with 330 different bonds. Again, we ignore the returns of a given bond, if we have some missing data in the time series. We end up with 1243 daily return distributions.
- exchange rates, where we use the following, all with respect to the US Dollar: Australian Dollar, Brazilian Real, British Pound, Canadian Dollar, Chinese Yuan, Danish Krone, Euro, Hong Kong Dollar, Indian Rupee, Japanese Yen, Mexican Peso and New Zealand Dollar. We use data staring in 1999 and end 2016. This results in 91 distributions for the daily returns.
- hedge funds, where we use the data given in Thomson Reuters Datastream. We observe 80 funds from 2008 to 2016. Again, we exclude a fund, if we find missing observations. We end up with 630 daily return distributions.

A sample of 10,000 returns for each class is used to give an overview of the general distribution of the returns. These are shown in figure 5.4. A kernel density estimation in blue and a normal distribution with the corresponding first two moments in red are added as comparison. We find a general deviation from a normal distribution. The distribution of the stocks seems to have the smallest deviation from the normal distribution and the bonds the biggest.

To quantify the deviation from the normal, we test each distribution with the Jarque-Bera test. The rejection rate of the null hypothesis of normal distributed returns are reported in the corresponding columns in table 5.2.

Supporting the impression from figure 5.4, we find a high rejection rate of H_0 in general, but according to the test, the exchange rates have the smallest deviation from the normal distribution. This could come from the fact, that the median excess kurtosis is smaller than the one of the stocks, or that the exchange rates show a smaller skewness in general. All descriptive statistics are reported in the corresponding blocks in table 5.3.

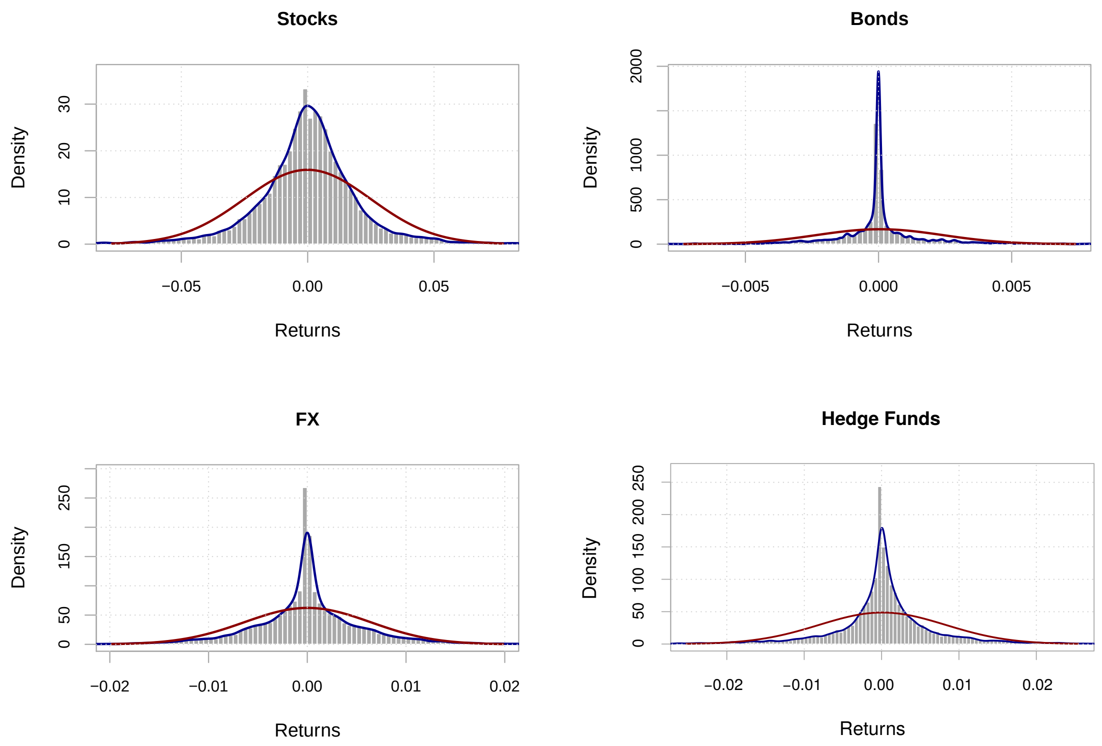

Figure 5.4: Return distribution of the different asset classes. Histogram of the log-returns for four different asset classes. Additionally a kernel density estimation in blue, as well as a normal distribution with the corresponding first two moments in red is shown.

Table 5.2: Empirical rejection rate based on the Jarque-Bera Test for the four asset classes.

α	crit. Value	H0 rejection			
		Stocks	Bonds	FX	Hedge Funds
10%	4.61	90.52%	96.86%	85.83%	93.97%
5%	5.99	88.12%	95.98%	82.59%	93.02%
1%	9.21	83.88%	94.05%	73.28%	86.67%

Table 5.3: Descriptive statistics for the return distributions for the four asset classes. For the sake of fewer decimals, the mean and standard deviation are calculated using the log returns as percentage values.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
			Stocks			
Mean	-1.42	-0.02	0.05	0.04	0.12	0.89
SD	0.61	1.37	1.85	2.18	2.59	12.80
Skewness	-14.11	-0.37	-0.03	-0.12	0.30	10.37
Excess Kurtosis	-0.56	1.09	2.19	4.99	4.57	211.30
			Bonds			
Mean	-0.06	-0.01	0.00	0.00	0.01	0.05
SD	0.00	0.07	0.16	0.19	0.30	0.72
Skewness	-15.66	-0.45	0.14	0.81	1.76	15.62
Excess Kurtosis	-0.96	2.75	11.26	45.09	74.39	244.60
			FX			
Mean	-0.12	-0.02	0.00	0.00	0.03	0.17
SD	0.00	0.33	0.55	0.54	0.74	1.95
Skewness	-15.23	-0.40	-0.08	-0.09	0.30	4.66
Excess Kurtosis	-0.41	0.71	1.89	5.51	4.72	234.80
			Hedge Funds			
Mean	-0.29	-0.01	0.00	0.01	0.03	0.18
SD	0.01	0.24	0.55	0.67	0.97	2.92
Skewness	-15.43	-0.70	-0.33	-0.49	-0.09	12.99
Excess Kurtosis	-0.23	1.38	2.15	6.92	4.24	239.70

In general the correlation of performance measures is calculated using hedge fund data. What we find from the descriptive statistics is that the other three asset classes share some features of the hedge fund distributions, such as a negative skewness (at least in some cases). So we think adding the other classes to our analysis gives additional insight into the rank correlation of the performance measures.

As earlier noted, the performance comparisons studies encompass a more extended set of measures that have been proposed in the literature. These are chosen as the Sortino ratio, Kappa 3, the UPR, the Calmar ratio, Sterling ratio, Burke ratio and the Treynor ratio to the Sharpe ratio and the Omega ratio.

For the Omega ratio, the Sortino ratio, the Kappa 3 and the UPR, we use the return distribution as the P&L distribution, e.g. we set the threshold to a value of $z = 0$. For the Treynor ratio, the CAPM beta coefficient is estimated for each year using the corresponding stock returns and the S&P500 returns as proxy for the market return. Technically this implies a forward looking beta coefficient, but the results do not change when using the lagged return distributions.

The resulting rank correlations for the four different asset classes are shown in table 5.4 to 5.7. In each table, we report both Kendall's τ in the upper triangular part and Spearman's ρ in the lower part of the correlation matrix. The diagonal kept blank. In general conformity with Eling and Schuhmacher (2007) findings, we find very high and significant correlations with most metrics for the stock returns. In general the correlations are above 80% for all asset classes for all measures except for the Treynor Ratio and the Upside Potential. The former seems to be in line with the other metrics in the case of stocks, the correlation drops to about $10\% - 20\%$ for the hedge fund and the bond returns and even turns into a negative one for the exchange rates. The latter seems to be in line except in the case of bond returns. There the correlation drops to about $40\% - 60\%$.

Summarizing, we find rather high correlation coefficients for the performance measures as Eling and Schuhmacher (2007) for hedge funds (except for the Treynor ratio), but cannot conclude that the choice between them does not make a difference. This is in line with the findings in Zakamouline (2009) and Ornelas, Silva Júnior, and Fernandes (2012), but in there significantly lower correlations have been found.

Next, we turn to the W and W^* measures as defined in section 5.2. The rank correlations are for all asset classes lower than for the other measures. For the W^*,

Table 5.4: Rank correlations for the return distribution of stocks. The upper triangular part reports the Kendall's τ and the lower part Spearman's ρ.

	Sharpe	Omega	Sortino	Kappa3	Upside	Calmar	Sterling	Burke	Treynor	W	W^*
Sharpe		0.955***	0.954***	0.934***	0.705***	0.897***	0.897***	0.930***	0.830***	0.183***	0.028
Omega	0.997***		0.966***	0.937***	0.707***	0.905***	0.905***	0.943***	0.831***	0.170***	0.021
Sortino	0.997***	0.998***		0.970***	0.725***	0.908***	0.908***	0.944***	0.827***	0.162***	0.014
Kappa3	0.994***	0.993***	0.998***		0.739***	0.898***	0.898***	0.927***	0.817***	0.157***	0.011
Upside	0.884***	0.882***	0.897***	0.908***		0.705***	0.705***	0.712***	0.657***	-0.025	-0.119***
Calmar	0.986***	0.988***	0.989***	0.986***	0.885***		1.000***	0.925***	0.827***	0.174***	0.020
Sterling	0.986***	0.988***	0.989***	0.986***	0.885***	1.000***		0.925***	0.827***	0.174***	0.020
Burke	0.993***	0.995***	0.995***	0.992***	0.888***	0.992***	0.992***		0.832***	0.172***	0.018
Treynor	0.955***	0.957***	0.955***	0.951***	0.847***	0.954***	0.954***	0.956***		0.165***	0.023
W	0.270***	0.251***	0.239***	0.233***	-0.037***	0.256***	0.256***	0.253***	0.244***		0.457***
W^*	0.042***	0.031***	0.022	0.017	-0.176***	0.029	0.029	0.027	0.034***	0.633***	

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

Table 5.5: Rank correlations for the return distribution of bonds. The upper triangular part reports the Kendall's τ and the lower part Spearman's ρ.

	Sharpe	Omega	Sortino	Kappa3	Upside	Calmar	Sterling	Burke	Treynor	W	W^*
Sharpe		0.865***	0.923***	0.903***	0.419***	0.889***	0.889***	0.924***	-0.382***	0.129***	0.039
Omega	0.969***		0.888***	0.848***	0.431***	0.863***	0.863***	0.901***	-0.378***	0.087	0.008
Sortino	0.989***	0.974***		0.959***	0.399***	0.890***	0.890***	0.952***	-0.380***	0.138***	0.036
Kappa3	0.985***	0.959***	0.997***		0.394***	0.870***	0.870***	0.919***	-0.374***	0.153***	0.042
Upside	0.576***	0.568***	0.546***	0.541***		0.409***	0.409***	0.402***	-0.199***	-0.163***	-0.154***
Calmar	0.982***	0.971***	0.982***	0.975***	0.563***		1.000***	0.900***	-0.377***	0.096***	0.033
Sterling	0.982***	0.971***	0.982***	0.975***	0.563***	1.000***		0.900***	-0.377***	0.096***	0.033
Burke	0.990***	0.980***	0.996***	0.990***	0.550***	0.984***	0.984***		-0.385***	0.124***	0.034
Treynor	-0.459***	-0.412***	-0.453***	-0.457***	-0.227***	-0.440***	-0.440***	-0.454***		-0.133***	-0.017
W	0.194***	0.123***	0.205***	0.225***	-0.241***	0.146***	0.146***	0.186***	-0.195***		0.133***
W^*	0.036	0.007	0.044	0.050	-0.212***	0.043	0.043	0.036	-0.032	0.209***	

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

Table 5.6: Rank correlations for the return distribution of exchange rates. The upper triangular part reports the Kendall's τ and the lower part Spearman's ρ.

	Sharpe	Omega	Sortino	Kappa3	Upside	Calmar	Sterling	Burke	Treynor	W	W^*
Sharpe		0.763***	0.769***	0.763***	0.653***	0.724***	0.724***	0.768***	-0.149	0.385***	0.208
Omega	0.850***		0.955***	0.927***	0.614***	0.893***	0.893***	0.954***	-0.106	0.239***	0.283***
Sortino	0.858***	0.995***		0.971***	0.616***	0.894***	0.894***	0.964***	-0.106	0.257***	0.282***
Kappa3	0.859***	0.988***	0.998***		0.612***	0.886***	0.886***	0.945***	-0.105	0.268***	0.284***
Upside	0.825***	0.782***	0.792***	0.790***		0.600***	0.600***	0.615***	-0.071	0.184	0.158
Calmar	0.847***	0.982***	0.982***	0.978***	0.775***		1.000***	0.900***	-0.117	0.252***	0.276***
Sterling	0.847***	0.982***	0.982***	0.978***	0.775***	1.000***		0.900***	-0.117	0.252***	0.276***
Burke	0.861***	0.996***	0.998***	0.995***	0.789***	0.983***	0.983***		-0.105	0.255***	0.279***
Treynor	-0.172***	-0.130	-0.133	-0.134	-0.090	-0.154	-0.154	-0.134		-0.021	-0.121
W	0.532***	0.343***	0.371***	0.389***	0.270***	0.362***	0.362***	0.371***	-0.035		0.218
W^*	0.298***	0.409***	0.406***	0.407***	0.232***	0.403***	0.403***	0.402***	-0.177***	0.317***	

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

Table 5.7: Rank correlations for the return distribution of hedge funds. The upper triangular part reports the Kendall's τ and the lower part Spearman's ρ.

	Sharpe	Omega	Sortino	Kappa3	Upside	Calmar	Sterling	Burke	Treynor	W	W^*
Sharpe		0.957***	0.976***	0.958***	0.724***	0.895***	0.895***	0.946***	0.154***	0.232***	0.064
Omega	0.993***		0.959***	0.935***	0.722***	0.898***	0.898***	0.944***	0.154***	0.216***	0.051
Sortino	0.998***	0.994***		0.974***	0.723***	0.901***	0.901***	0.953***	0.148***	0.223***	0.061
Kappa3	0.996***	0.987***	0.998***		0.727***	0.891***	0.891***	0.941***	0.143***	0.225***	0.065
Upside	0.875***	0.857***	0.872***	0.879***		0.711***	0.711***	0.723***	0.112	0.064	0.038
Calmar	0.984***	0.985***	0.986***	0.983***	0.864***		1.000***	0.924***	0.143***	0.202***	0.045
Sterling	0.984***	0.985***	0.986***	0.983***	0.864***	1.000***		0.924***	0.143***	0.202***	0.045
Burke	0.995***	0.994***	0.997***	0.993***	0.868***	0.991***	0.991***		0.148***	0.216***	0.058
Treynor	0.139***	0.133***	0.129***	0.122***	0.119***	0.131***	0.131***	0.128***		0.094	0.135
W	0.335***	0.307***	0.323***	0.325***	0.098	0.292***	0.292***	0.310***	0.131***		0.074
W^*	0.095	0.075	0.090	0.096	0.052	0.067	0.067	0.087	0.198***	0.110***	

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

the correlation is only significantly different from zero for the exchange rates. For the other classes, we find a non significant correlation most of the time. Thus it evaluates performance in a way, the other metrics cannot, resp. do not have a strong focus on. The W measure correlates positively most of the time with most of the metrics but not as strong as the ones described above. Interestingly we see both a weak positive correlation (stocks and hedge funds) and a weak negative correlation with the Treynor Ratio for both of the W/W^* measures, together with a weak positive correlation with the rest of the performance measures. The proposed metrics based on the 'win' function therefore differs significantly in the ranking form the existing measures. Since the Sharpe ratio is known for a skewness bias, the rankings of the other measures may likewise be influenced by the skewness of the distribution. In the case of the exchange rates, where the skewness is the lowest for our asset classes, we see a decrease in correlation with the Sharpe Ratio. Once this is explicitly incorporated via the W-metrics, things do change. If investors or fund managers prefer negative skewness, as argued in section 5.1, then the W^* metric provides a better set of rankings across different returns, in the sense that skewness is taken in account explicitly.

Further research in this direction may be an analysis weather the informations, revealed by the new measures, can be transformed into a profitable trading strategy. If this is the case, these informations may be seen as 'new' information and one can expect a premium when trading on them. What changes the observed correlation changes is another interesting topic. Although the distribution of Hedge Fund returns and exchange rate returns does not seem very different, we find very different rank correlations. So the question rises, what drives the correlations? This research question does not only come from this paper, but might be a natural question following the other papers concerned with the rank correlation of performance measures.

5.4 Concluding remarks

Investors naturally seek high returns and low volatility, and there is a case for negative skewness in addition. In order to evaluate return distributions, we propose the ratio of a contextually motivated skewness metric to dispersion. The latter can be either as the dual spread metric or else just the standard deviation. A sensitivity analysis shows that the bias arising from skewness in the Sharpe ratio does not crucially rely on the definition of the denominator. Thus in the Ford Motors illustration, the pattern changes in only in 11% of the cases.

Interestingly the analysis shows very different results of the correlations for the different asset classes. Keeping in mind the descriptive statistics of the return distributions, which do not show a completely different picture for the four asset classes, comparable results were expected across the asset classes. In some cases, we can support the results in Eling and Schuhmacher (2007), where a very high correlation is found. But we also find cases, where the correlation is significantly lower (from about 90% to about 50% or even negative) for some performance metrics. This is in line with Zakamouline (2009) and Ornelas, Silva Júnior, and Fernandes (2012). When taking the W/W^* metric into account, we find correlations in the range of $-10\% - 40\%$ to the existing metrics. This indicates that the use of a Friedman-Savage like function results in a very different picture, arguably resulting in a bigger information set about the underlying distribution. Also within the W/W^* we do not find a very high correlation. This indicates that although the use of the market return as benchmark does not lead to very different values in absolute terms, the ranking is significantly indicated.

Particular motivation for the augmented W-metric exists in episodes where skewness can arise, as in times of market stress. However even in more normal times, it is useful to be able to separate out the entangled contribution of skewness as a potentially important dimension to portfolio analysis. This is especially so where a given stock or ETF portfolio displays habitual tendencies in one direction or the other. The advantage of thinking in Friedman-Savage terms is that it does focus on the welfare consequences of investment decisions in the presence of potential skewness. Once we allow for investor preferences of this kind, the rankings of alternative investments can change quite dramatically.

6 Conclusion

The first part of this thesis is about an accurate estimation of the covariance/correlation matrix and detecting structural breaks within these dependency structures. Chapter 2 addresses the problem of estimating the covariance/correlation matrix with limited observations. Estimators with and without the normality assumption of returns are used and the errors of covariance estimation and correlation estimation compared. It is analyzed, if estimation improvements transfer to economic improvements measured by the Sharpe ratio and annualized volatilities of minimum-variance portfolios. Significant out-performance of some shrinking estimators in the economic sense are found, which seem to depend weakly on the normality assumption. Using a shrinking estimator with a scaled identity matrix as shrinking target, the Sharpe ratio increases by a factor of about two.

Chapter 3 tests for a constant correlation structure without any model assumption. These model free tests for constant parameters often fail to detect structural changes in high dimensions. In practice this corresponds to a portfolio with many assets and a reasonable long time series. The dimensionality of the problem is reduced by looking at a compressed panel of time series obtained by cluster analysis and the principal components of the data. With this procedure tests for constant correlation matrix can be extended from a sub portfolio to whole index, which we exemplify using a major stock index.

The second part deals with the general problem of aggregating distributions. Using conditional first moments, one can ask the question: am I better off than the others in the population? Chapter 4 deals with this question in the context of income distributions and proposes metrics for skewness and spread, based on this internal view. Using them, the trajectories of European countries from 2005 to 2013 are tracked in a phase plane. This movement enables a grouping into three groups of inequality risers, fallers and a mixed group. In a regression analysis determinants of the Gini coefficient are examined to check if these effects translate to the two metrics.

Chapter 5 turns to one source of income, investment income and the question arising from the perspective of a fund manager: How does a performance metric look like and would fund managers be ranked differently when using different performance metrics? The originated performance criterion w is more consistent with the implicit Friedman-Savage utility ordering. It weights the lower versus upper conditional expected returns, while a dual spread or dispersion metric d also exists. A point of departure is the conventional Sharpe performance ratio, with the empirical comparisons extending to a range of existing performance criteria. In contrast to existing metrics, the proposed performance metric W results in different and more embracing rankings.

A Extended tables for distances to the true correlation matrix, Sharpe ratios and annualized volatilities

Table A.1: Distances to the true correlation matrix. Results are grouped by the three data generating distributions and the number of observations horizontally. Vertically grouping is based on the estimator used and the underlying covariance type.

		250	500	750	250	500	750	250	500	750
		normal			t			gamma		
					Empirical					
	AR	25.339	17.922	14.636	25.339	17.922	14.637	25.339	17.922	14.636
	eigen	25.340	17.921	14.638	25.338	17.920	14.637	25.340	17.923	14.637
	emp	25.338	17.920	14.637	25.339	17.921	14.637	25.338	17.922	14.636
					Fisher					
	AR	1.545	5.308	7.139	1.543	5.325	7.131	1.543	5.327	7.127
μ	eigen	1.548	5.317	7.144	1.538	5.301	7.124	1.543	5.304	7.144
	emp	1.542	5.319	7.123	1.543	5.304	7.125	1.543	5.294	7.131
	AR	25.317	17.914	14.632	25.317	17.914	14.633	25.317	17.914	14.632
I	eigen	25.317	17.913	14.634	25.316	17.912	14.633	25.318	17.915	14.633
	emp	25.316	17.913	14.633	25.316	17.913	14.633	25.316	17.914	14.632
	AR	1.545	1.556	1.543	1.543	1.551	1.552	1.543	1.552	1.546
D	eigen	1.548	1.558	1.549	1.538	1.544	1.551	1.543	1.555	1.563
	emp	1.542	1.544	1.552	1.543	1.555	1.543	1.543	1.549	1.540
					Touloumis					
	AR	9.479	9.488	9.080	9.473	9.493	9.078	9.479	9.494	9.075
μ	eigen	9.476	9.490	9.084	9.480	9.483	9.074	9.485	9.486	9.084
	emp	9.469	9.491	9.075	9.475	9.486	9.074	9.470	9.483	9.077
	AR	25.317	17.914	14.632	25.317	17.914	14.632	25.317	17.914	14.632
I	eigen	25.317	17.913	14.634	25.318	17.915	14.633	25.318	17.915	14.633
	emp	25.316	17.913	14.633	25.316	17.914	14.632	25.316	17.914	14.632
	AR	1.569	1.576	1.562	1.565	1.571	1.571	1.566	1.571	1.566
D	eigen	1.571	1.577	1.569	1.561	1.563	1.571	1.567	1.574	1.582
	emp	1.565	1.564	1.571	1.566	1.574	1.564	1.565	1.569	1.559

Note: For clarity, t-statistics are omitted. All differences are significant on the 0.1% level.

Table A.2: Sharpe ratios of the minimum-variance portfolios, generated by the different covariance estimators. Results are grouped by the three data generating distributions and the number of observations horizontally. Vertically grouping is based on the estimator used and the underlying covariance type.

		250	500	750	250	500	750	250	500	750
		normal			t			gamma		
					Empirical					
	AR	-	0.288	0.473	-	0.289	0.472	-	0.289	0.471
	eigen	-	0.290	0.470	-	0.288	0.471	-	0.289	0.466
	emp	-	0.290	0.471	-	0.289	0.472	-	0.289	0.473
					Fisher					
	AR	0.578	0.575	0.618	0.579	0.575	0.618	0.579	0.576	0.617
μ	eigen	0.579	0.576	0.616	0.580	0.577	0.616	0.579	0.575	0.612
	emp	0.578	0.579	0.617	0.579	0.575	0.618	0.579	0.577	0.619
	AR	0.389	0.291	0.473	0.390	0.292	0.473	0.391	0.292	0.471
I	eigen	0.390	0.293	0.471	0.390	0.291	0.471	0.389	0.292	0.467
	emp	0.389	0.293	0.471	0.389	0.291	0.472	0.389	0.292	0.473
	AR	0.578	0.551	0.579	0.579	0.552	0.579	0.579	0.552	0.579
D	eigen	0.579	0.552	0.578	0.580	0.553	0.578	0.579	0.551	0.574
	emp	0.578	0.555	0.579	0.579	0.552	0.579	0.579	0.553	0.581
					Touloumis					
	AR	0.593	0.567	0.611	0.594	0.567	0.611	0.594	0.568	0.610
μ	eigen	0.594	0.569	0.609	0.594	0.568	0.609	0.594	0.567	0.604
	emp	0.593	0.570	0.610	0.594	0.567	0.611	0.594	0.569	0.612
	AR	0.389	0.291	0.473	0.390	0.292	0.473	0.391	0.292	0.471
I	eigen	0.390	0.293	0.471	0.390	0.291	0.471	0.389	0.292	0.467
	emp	0.389	0.293	0.471	0.389	0.291	0.472	0.389	0.292	0.473
	AR	0.579	0.554	0.583	0.581	0.555	0.584	0.581	0.555	0.584
D	eigen	0.581	0.555	0.583	0.582	0.556	0.583	0.581	0.554	0.579
	emp	0.580	0.558	0.584	0.581	0.555	0.584	0.581	0.556	0.585

Note: For clarity, t-statistics are omitted. All differences are significant on the 0.1% level.

Table A.3: Annualized volatilities of the minimum-variance portfolios, generated by the different covariance estimators. Results are grouped by the three data generating distributions and the number of observations horizontally. Vertically grouping is based on the estimator used and the underlying covariance type.

		250	500	750	250	500	750	250	500	750
		normal			t			gamma		
					Empirical					
	AR	-	2.833	1.732	-	2.824	1.739	-	2.817	1.739
	eigen	-	2.820	1.741	-	2.825	1.741	-	2.822	1.753
	emp	-	2.811	1.737	-	2.824	1.734	-	2.819	1.733
					Fischer					
	AR	1.406	1.408	1.314	1.405	1.406	1.318	1.405	1.404	1.317
m	eigen	1.406	1.404	1.319	1.401	1.404	1.319	1.405	1.407	1.328
	emp	1.405	1.401	1.317	1.406	1.407	1.316	1.404	1.403	1.313
	AR	2.098	2.808	1.731	2.096	2.799	1.737	2.094	2.792	1.738
I	eigen	2.101	2.795	1.740	2.091	2.800	1.740	2.100	2.798	1.752
	emp	2.098	2.786	1.735	2.101	2.799	1.733	2.096	2.794	1.731
	AR	1.406	1.471	1.402	1.405	1.469	1.407	1.405	1.468	1.404
D	eigen	1.406	1.468	1.407	1.401	1.466	1.407	1.405	1.470	1.416
	emp	1.405	1.464	1.405	1.406	1.469	1.403	1.404	1.466	1.400
					Touloumis					
	AR	1.369	1.429	1.331	1.369	1.427	1.335	1.369	1.425	1.335
m	eigen	1.370	1.424	1.337	1.366	1.425	1.336	1.369	1.428	1.346
	emp	1.370	1.421	1.334	1.370	1.428	1.333	1.368	1.423	1.330
	AR	2.098	2.808	1.731	2.096	2.799	1.737	2.094	2.792	1.738
I	eigen	2.101	2.795	1.740	2.091	2.800	1.740	2.100	2.798	1.752
	emp	2.098	2.786	1.735	2.101	2.799	1.733	2.096	2.794	1.731
	AR	1.400	1.462	1.389	1.400	1.460	1.393	1.400	1.458	1.390
D	eigen	1.401	1.459	1.393	1.396	1.457	1.393	1.399	1.461	1.402
	emp	1.399	1.455	1.391	1.401	1.460	1.390	1.399	1.457	1.387

Note: For clarity, t-statistics are omitted. All differences are significant on the 0.1% level.

B Phase diagrams for major countries

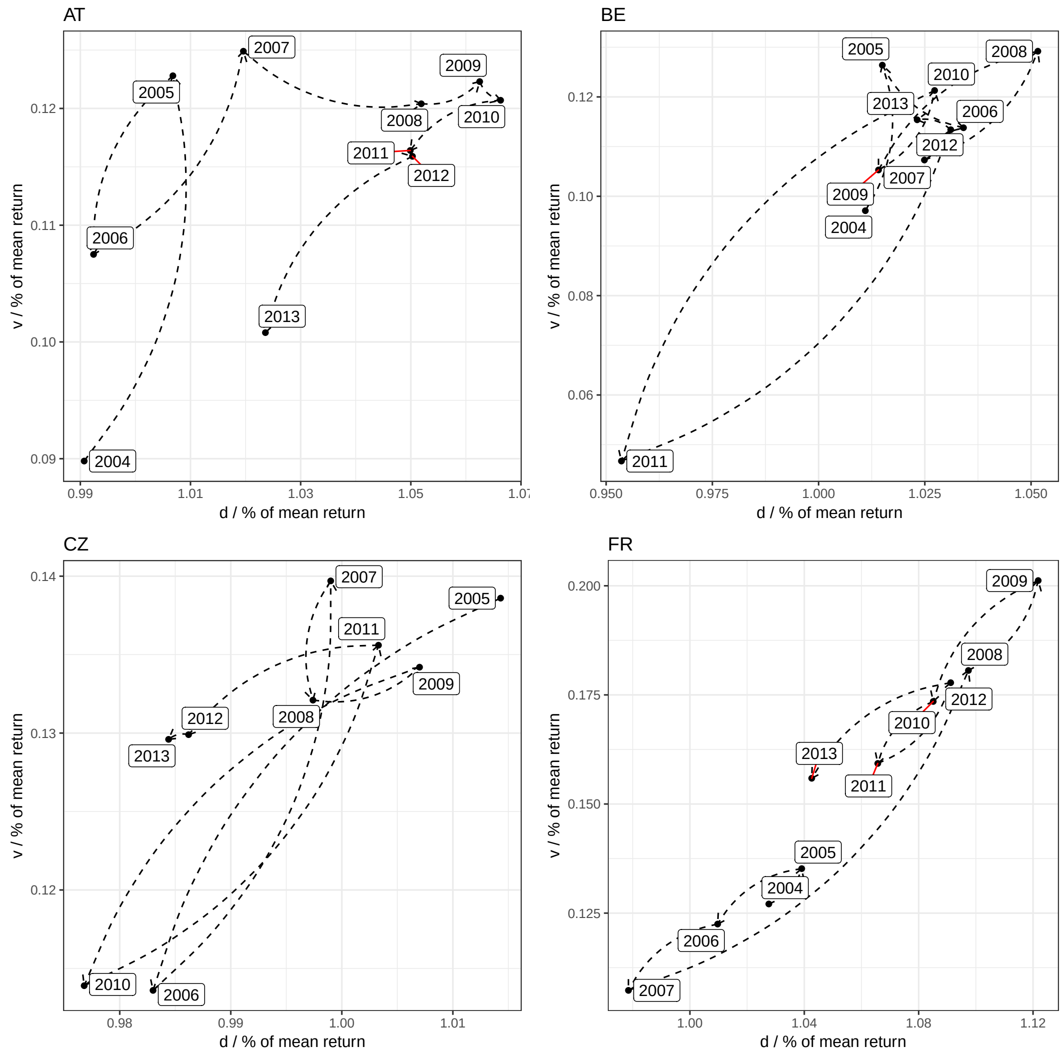

Figure B.1: Phase diagrams. Phase diagrams for the different countries for the complete time interval.

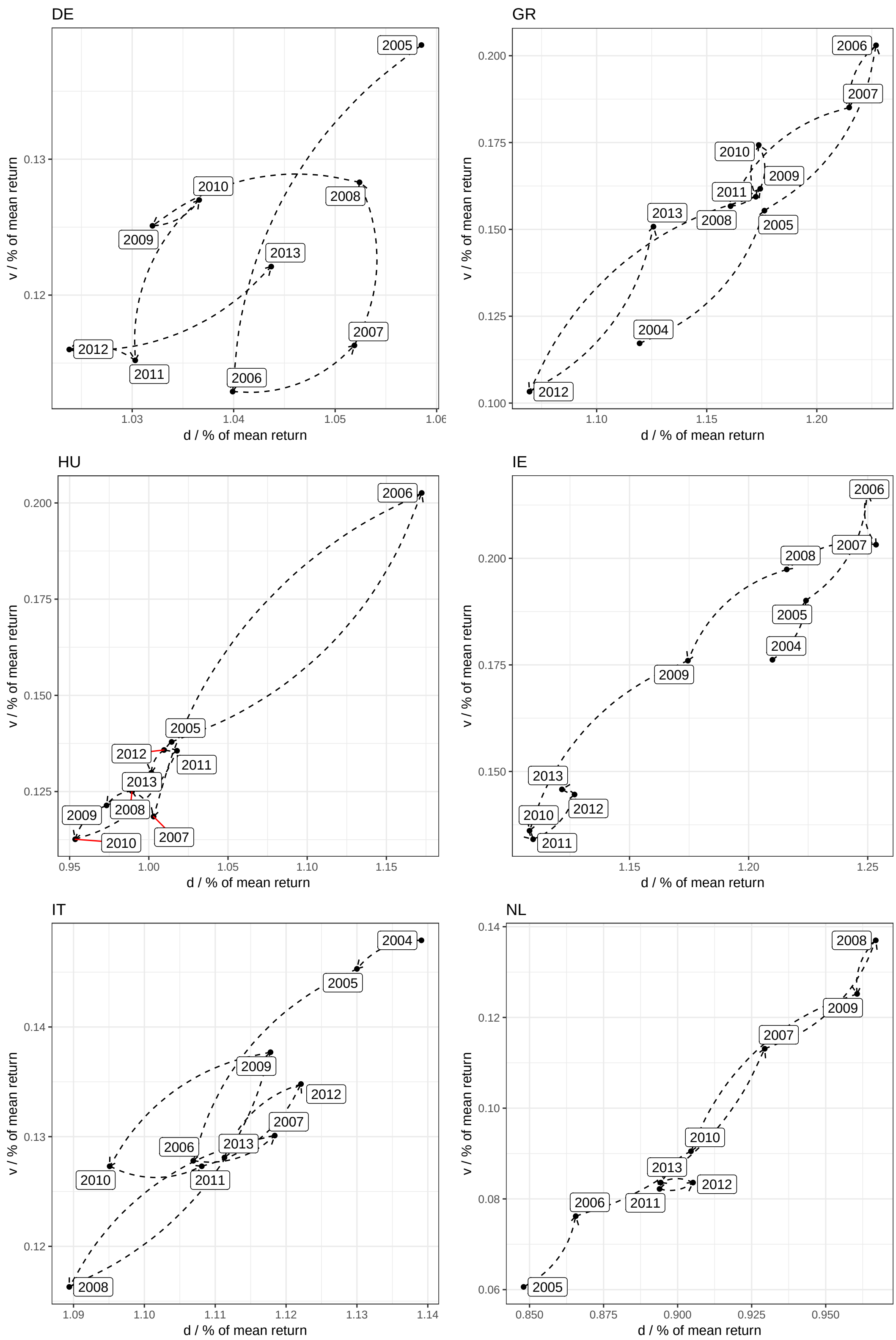

Figure B.1: continued: Phase diagrams.

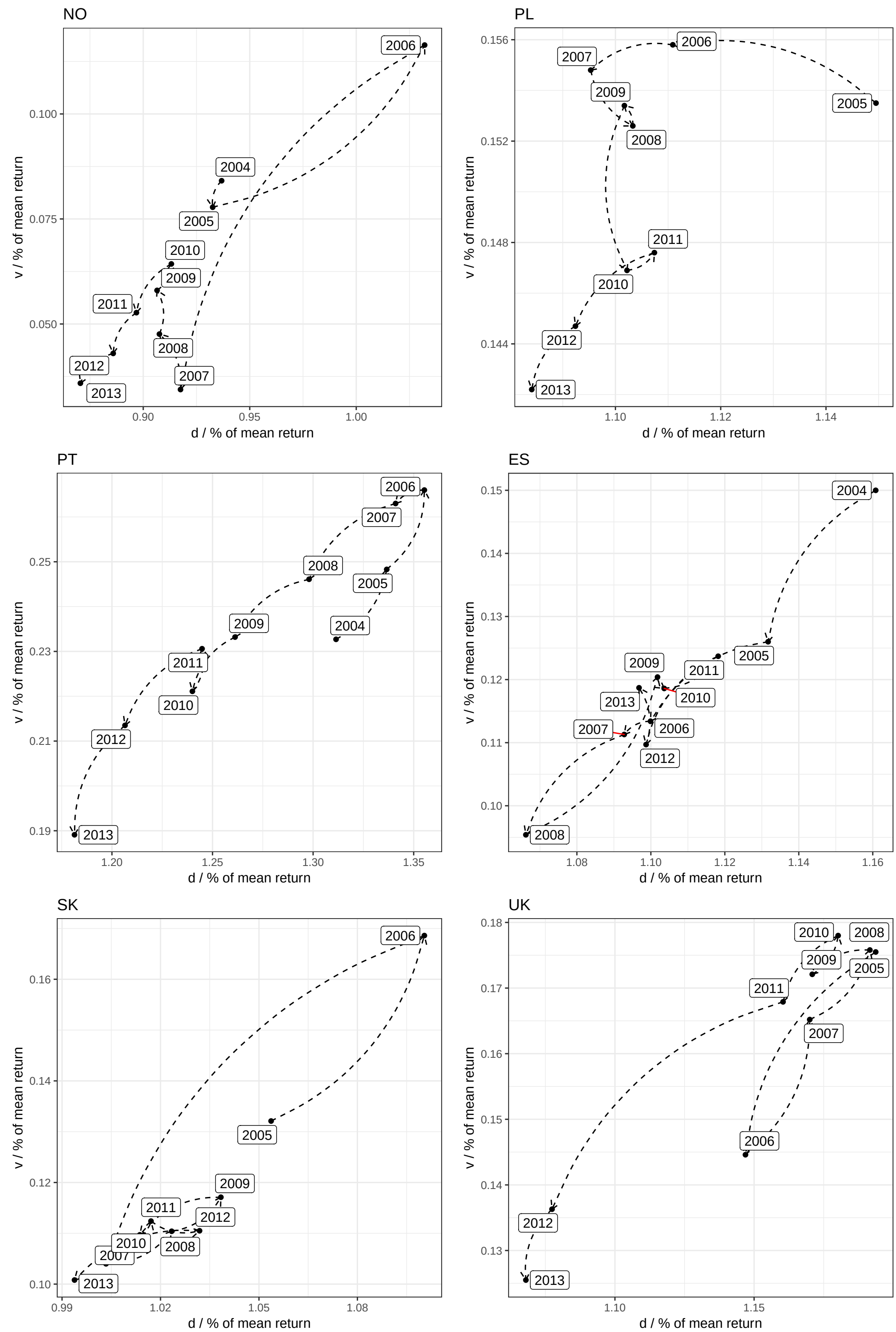

Figure B.2: continued: Phase diagrams.

Bibliography

Adams Jr., Richard H. 2003. "Economic growth, inequality, and poverty: Findings from a new data set". *Policy Research Working Paper 2972.*

Aleš, Bulíř. 2001. "Income inequality: does inflation matter?" *IMF Staff papers* 48 (1): 139–159.

Alexander, Carol. 1999. *Optimal hedging using cointegration*, 357:2039–2058. The Royal Society.

Anderberg, Michael R. 2014. *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks.* Academic press.

Andreou, Elena, and Eric Ghysels. 2009. "Structural breaks in financial time series". In *Handbook of Financial Time Series*, 839–870. Springer.

Ang, James S., and Jess H. Chua. 1979. "Composite measures for thte evaluation of investment performance". *Journal of Financial and Quantitative Analysis* 14 (2): 361–384.

Atkinson, Anthony B. 1970. "On the measurement of inequality". *Journal of Economic Theory* 2:244–263.

Aue, Alexander, and Lajos Horváth. 2013. "Structural breaks in time series". *Journal of Time Series Analysis* 34 (1): 1–16.

Bergson, Abram. 1938. "A Reformulation of Certain Aspects of Welfare Economics". *The Quarterly Journal of Economics* 52 (2): 310–334.

Bonanno, Giovanni, et al. 2004. "Networks of equities in financial markets". *The European Physical Journal B-Condensed Matter and Complex Systems* 38 (2): 363–371.

Bourguignon, Francois. 1979. "Decomposable income inequality measures". *Econometrica* 47:901–920.

Bowden, Roger J. 2016a. "Dual spread and asymmetry distribution metrics based in partition entropy". *Kiwcap Research Ltd*, Kiwcap Research Ltd: 1–13.

– . 2016b. "Giving Gini direction: An asymmetry metric for economic disadvantage". *Economics Letters* 138:96–99.

– . 2012. "Information, measure shifts and distribution metrics". *Statistics* 46 (2): 249–262.

Bowden, Roger J., Peter N. Posch, and Daniel Ullmann. 2017a. "Asymmetry and performance metrics for financial returns".

– . 2017b. "Income distribution in troubled times: Disadvantage and dispersion dynamics in Europe 2005-2013". *Finance Research Letters.*

Brida, Juan G., and Wiston A. Risso. 2010. "Hierarchical structure of the German stock market". *Expert Systems with Applications* 37 (5): 3846–3852.

Bronstein, Ilja N. 2012. *Taschenbuch Der Mathematik.* Deutsch Harri GmbH; Auflage: 4. ISBN: 3817120044.

Bücher, Axel and Kojadinovic, Ivan and Rohmer, Tom and Segers, Johan. 2014. "Detecting changes in cross-sectional dependence in multivariate time series". *Journal of Multivariate Analysis* 132:111–128.

Burke, Gibbons. 1994. "A sharper Sharpe ratio". *Futures* 23 (3): 56.

Chen, Yilun, Ami Wiesel, and Alfred O. Hero. 2009. "Shrinkage estimation of high dimensional covariance matrices". In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2937–2940. IEEE.

Chow, Gregory C. 1960. "Tests of equality between sets of coefficients in two linear regressions". *Econometrica*: 591–605.

Chunhachinda, Pomchai, et al. 1997. "Portfolio selection and skewness: Evidence from intemational stock markets". *Journal of Banking and Finance* 21:143–167.

Cochrane, John H. 2000. *Asset Pricing.*

Conceição, Pedro, and Pedro Ferreira. 2000. "The young person's guide to the Theil index: Suggesting intuitive interpretations and exploring analytical applications". *UTIP Working Paper Number 14*: 1–54.

Cowell, Frank A. 2009. “Measuring Inequality”.

– . 1980. “On the Structure of Additive Inequality Measures”. *Review of Economic Studies* 47 (3): 521–531.

Cryer, Jonathan D., and Kung-Sik Chan. 2008. *Time series analysis: with applications in R*. Springer Science & Business Media.

Dalton, Hugh. 1920. “The measurement of the inequality of incomes”. *The Economic Journal* 30 (119): 348–361.

Davies, James, and Michael Hoy. 1995. “Making Inequality Comparisons When Lorenz Curves Intersect”. *The American Economic Review*: 980–986.

Donaldson, David, and John A. Weymark. 1980. “A single-parameter generalization of the Gini indices of inequality”. *Journal of Economic Theory* 22 (1): 67–86.

Dreher, Axel. 2006. “Does globalization affect growth? Evidence from a new index of globalization”. *Applied Economics* 38 (10): 1091–1110.

Dreher, Axel, Noel Gaston, and Pim Martens. 2008. *Measuring globalisation: Gauging its consequences*. Springer Science & Business Media.

Efron, Bradley, and Carl N. Morris. 1977. *Stein’s paradox in statistics*. WH Freeman.

Eidgenössische Finanzverwaltung. 2017. “Staatsschuldenquote der Schweiz von 2007 bis 2017 (in Relation zum Bruttoinlandsprodukt)”. `https://de.statista.com/statistik/daten/studie/216761/umfrage/staatsverschuldung-der-schweiz-in-relation-zum-bruttoinlandsprodukt-bip`.

Eisen, Michael B., et al. 1998. “Cluster analysis and display of genome-wide expression patterns”. *Proceedings of the National Academy of Sciences* 95 (25): 14863–14868.

Eling, Martin, and Frank Schuhmacher. 2007. “Does the choice of performance measure influence the evaluation of hedge funds?” *Journal of Banking and Finance* 31 (9): 2632–2647.

Eltetö, Ö., and E. Frigyes. 1968. “New Income Inequality Measures As Efficient Tools for Causal Analysis and Planning”. *Econometrica* 36 (2): 383–396.

Eurostat. 2016a. “Eurostat: Database”. `http://ec.europa.eu/eurostat/data/database`.

– . 2016b. *Glossary: EU statistics on income and living conditions (EU-SILC).*

– . 2014. *Living conditions in Europe.* 2014th ed. Eurostat Statistical books.

Fama, Eugene F. 1965. "Portfolio Analysis in a Stable Paretian Market". *Management Science* 11 (3): 404–419.

Fan, Jianqing, Yingying Fan, and Jinchi Lv. 2008. "High dimensional covariance matrix estimation using a factor model". *Journal of Econometrics* 147:186–197.

Farinelli, Simone, et al. 2008. "Beyond Sharpe ratio: Optimal asset allocation using different performance ratios". *Journal of Banking and Finance* 32 (10): 2057–2063.

Fisher, Thomas J., and Xiaoqian Sun. 2011. "Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix". *Computational Statistics & Data Analysis* 55:1909–1918.

Fodor, Imola K. 2002. *A survey of dimension reduction techniques.*

Foster, James E. 1983. "An axiomatic characterization of the Theil measure of income inequality". *Journal of Economic Theory* 31 (1): 105–121.

Foster, James E., Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures". *Econometrica* 52:761–766.

Foster, James E., and Artyom A. Shneyerov. 1999. "A general class of additively decomposable inequality measures". *Economic Theory* 14 (1): 89–111.

Friedman, Milton, and Leonard J. Savage. 1948. "The Utility Analysis of Choices Involving Risk". *Journal of Political Economy* 56 (4): 279–304.

Galeano, Pedro, and Dominik Wied. 2014. "Dating multiple change points in the correlation matrix". *TEST*: 1–22.

Goodman, Leo, and Kenneth J. Arrow. 1953. "Social Choice and Individual Values". *American Sociological Review* 18 (1): 116.

Gower, John C., and G. J. S. Ross. 1969. "Minimum spanning trees and single linkage cluster analysis". *Applied statistics*: 54–64.

Greene, William H. 2008. *Econometric analysis.* Granite Hill Publishers.

Greselin, Francesca, and Ricardas Zitikis. 2015. "Measuring economic inequality and risk: a unifying approach based on personal gambles, societal preferences and references". *arXiv:1508.00127.*

Gupta, Sanjeev, Hamid R. Davoodi, and Rosa Alonso-Terme. 2002. "Does Corruption Affect Income Inequality and Poverty?" *Economics of Governance* 3 (1): 23–45.

Gwartney, James, Robert Lawson, and Joshua C. Hall. 2015. "Economic freedom dataset". *Economic freedom of the world: 2015 annual report. Vancouver, British Columbia, Canada: Fraser Institute.*

Hair, Joseph F., et al. 2006. *Multivariate data analysis.* Vol. 6. Pearson Prentice Hall Upper Saddle River, NJ.

Haldar, Pranab, et al. 2008. "Cluster analysis and clinical asthma phenotypes". *American Journal of Respiratory and Critical Care Medicine* 178 (3): 218–224.

Hallin, Marc, Davy Paindaveine, and Thomas Verdebout. 2014. "Efficient R-estimation of principal and common principal components". *Journal of the American Statistical Association* 109 (507): 1071–1083.

Harlow, William Van. 1991. "Asset Allocation in a Downside-Risk Framework". *Financial Analysts Journal* 47 (5): 28–40.

Himeno, Tetsuto, and Takayuki Yamada. 2014. "Estimations for some functions of covariance matrix in high dimension under non-normality and its applications". *Journal of Multivariate Analysis* 130:27–44.

Hirschman, Albert O. 1980. *National power and the structure of foreign trade.* Vol. 1. Univ of California Press.

– . 1964. "The paternity of an index". *The American Economic Review*: 761–762.

Iacovou, Maria, Olena Kaminska, and Horacio Levy. 2012. "Using EU-SILC data for cross-national analysis : strengths , problems and recommendations". *ISER Working Paper Series*: 1–21.

Jaumotte, Florence, Subir Lall, and Chris Papageorgiou. 2013. "Rising Income Inequality: Technology, or Trade and Financial Globalization?" *IMF Economic Review* 61 (2): 271–309.

Jensen, Michael C. 1968. “The Performance of Mutual Funds in the Period 1945-1964”. *Journal of Finance* 23 (2): 389.

Jeon, Jongwoo, Subhash Kochar, and Chul G. Park. 2006. “Dispersive ordering - Some applications and examples”. *Statistical Papers* 47:227–247.

Jolliffe, Ian. 2002. *Principal component analysis.* Wiley Online Library.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect theory: An analysis of decision under risk”. *Econometrica* 47 (2): 263–292.

Kaplan, Paul D., and James A. Knowles. 2004. “Kappa: A Generalized Downside Risk-Adjusted Performance Measure”. *Journal of Performance Measurement* 8 (3): 42–54.

Keating, Con, and William F. Shadwick. 2002. “A Universal Performance Measure”. *Journal of Performance Measurement* 6 (January): 59–84.

Kerm, Philippe van. 2007. “Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC”.

Kestner, Lars N. 1996. “Getting a handle on true performance”. *Futures* 25:44–47.

Kleiber, Christian. 1997. “The existence of population inequality measures”. *Economics Letters* 57 (1): 39–44.

Kondor, Yaakov. 1971. “An Old-New Measure of Income Inequality”. *Econometrica* 93 (6): 1042.

Kraemer, Walter. 1998. “Measurement of inequality”. *Handbook of Applied Economic Statistics*: 39–61.

Kraemer, Walter, and Tileman Conring. 2017. “Beyond inequality: A novel measure of skewness and its properties”. *Working Paper, TU Dortmund University*, Working Paper, TU Dortmund University.

Kraus, Alan, and Robert H. Litzenberger. 1976. “Skewness preference and the valuation of risky assets”. *Journal of Finance* 31 (4): 1085–1100.

Kwan, Clarence C. Y. 2010. “The Requirement of a Positive Definite Covariance Matrix of Security Returns for Mean-Variance Portfolio Analysis: A Pedagogic Illustration”. *Spreadsheets in Education (eJSiE)* 4 (1): 4.

Laloux, Laurent, et al. 1999. “Noise dressing of financial correlation matrices”. *Physical Review Letters* 83 (7): 1467.

Laloux, Laurent, et al. 2000. “Random matrix theory and financial correlations”. *International Journal of Theoretical and Applied Finance* 3 (03): 391–397.

Ledoit, Olivier, and Michael Wolf. 2004a. “A well-conditioned estimator for large-dimensional covariance matrices”. *Journal of Multivariate Analysis* 88 (2): 365–411.

– . 2004b. “Honey I shrunk the sample covariance matrix”. *Journal of Portfolio Management* 30 (4): 110–119.

Lee, Cheol-Sung. 2005. “Income Inequality, Democracy, and Public Sector Size”. *American Sociological Review* 70 (1): 158–181.

Li, Hongyi, Lyn Squire, and Heng-fu Zou. 1998. “Explaining International and Intertemporal Variations in Income Inequality”. *The Economic Journal* 108 (446): 26–43.

Li, Hongyi, and Heng-fu Zou. 1998. “Income Inequality is not Harmful for Growth: Theory and Evidence”. *Review of Development Economics* 2:318–334.

Lintner, John. 1965. “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets”. *Review of Economics and Statistics*: 13–37.

Longin, François, and Bruno Solnik. 1995. “Is the correlation in international equity returns constant: 1960-1990?” *Journal of International Money and Finance* 14 (1): 3–26.

Magdon-Ismail, Malik, et al. 2003. “The maximum drawdown of the Brownian motion”. In *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, Proceedings (CIFEr)*, 243:20–23. Institute of Electrical / Electronics Engineers Inc.

Malinoski, Mallory. 2012. “On Culture and Income Inequality: Regression Analysis of Hofstede’s International Cultural Dimensions and Gini Coefficient”. *Xavier Journal of Politics* 3 (1): 32–48.

Mantegna, Rosario N. 1999. “Hierarchical structure in financial markets”. *The European Physical Journal B-Condensed Matter and Complex Systems* 11 (1): 193–197.

Mantegna, Rosario N., and H. Eugene Stanley. 1999. *Introduction to econophysics: correlations and complexity in finance.* Cambridge university press.

Markowitz, Harry. 1952. “The utility of wealth”. *Journal of Political Economy*: 151–158.

Mello, Luiz de. 2006. “Income Inequality and Redistributive Government Spending”. *Public Finance Review* 34 (3): 282–305.

Murtagh, Fionn, and Pierre Legendre. 2014. “Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?” *Journal of Classification* 31 (3): 274–295.

Nielsen, François, and Arthur S. Alderson. 1995. “Income inequality, development, and dualism: Results from an unbalanced cross-national panel”. *American Sociological Review*: 674–701.

Onnela, Jukka-Pekka, et al. 2002. “Dynamic asset trees and portfolio analysis”. *The European Physical Journal B-Condensed Matter and Complex Systems* 30 (3): 285–288.

Onnela, Jukka-Pekka, et al. 2003. “Dynamics of market correlations: Taxonomy and portfolio analysis”. *Physical Review E* 68 (5): 56110.

Ornelas, José R. H., Antônio F. Silva Júnior, and José L. B. Fernandes. 2012. “Yes, the choice of performance measure does matter for ranking of us mutual funds”. *International Journal of Finance {&} Economics* 17:61–72.

Pearson, Karl. 1901. “LIII. On lines and planes of closest fit to systems of points in space”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–572.

Pigou, Arthur C. 1912. *Wealth and welfare.* Macmillan / Company, limited.

Plerou, Vasiliki, et al. 2002. “Random matrix approach to cross correlations in financial data”. *Physical Review E* 65 (6): 66126.

Posch, Peter N., and Daniel Ullmann. 2017. “Estimation of Large Correlation Matrix with Shrinking Methods”.

Posch, Peter N., Daniel Ullmann, and Dominik Wied. 2017. “Detecting Structural Changes in Large Portfolios”.

Pourahmadi, Mohsen. 2013. *High-Dimensional Covariance Estimation: With High-Dimensional Data.* John Wiley & Sons.

Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Harvard University Press.

Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden Day.

Sharpe, William F. 1966. "Mutual Fund Performance". *The Journal of Business* 39 (1): 119–138.

Shorrocks, Anthony F. 1980. "The class of additively decomposable inequality measures". *Econometrica*: 613–625.

Srivastava, Muni S. 2005. "Some tests concerning the covariance matrix in high dimensional data". *Journal of the Japan Statistical Society* 35 (2): 251–272.

Stigler, Stephen M. 1990. "The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators". *Statistical Science*: 147–155.

"The Dutch Triangle". 1999. *Journal of Portfolio Management* 26 (1): 50–57.

Theil, Henri. 1967. *Economics and information theory*. Amsterdam: North-Holland.

Tola, Vincenzo, et al. 2008. "Cluster analysis for portfolio optimization". *Journal of Economic Dynamics and Control* 32 (1): 235–258.

Touloumis, Anestis. 2015. "Nonparametric Stein-type Shrinkage Covariance Matrix Estimators in High-Dimensional Settings". *Computational Statistics and Data Analysis* 83:251–261.

Transparency International. 2016. "Corruption Perceptions Index - Overview". `https://www.transparency.org/research/cpi/overview`.

Tumminello, Michele, Fabrizio Lillo, and Rosario N. Mantegna. 2010. "Correlation, hierarchies, and networks in financial markets". *Journal of Economic Behavior amd Organization* 75 (1): 40–58.

Ward Jr., Joe H. 1963. "Hierarchical grouping to optimize an objective function". *Journal of the American Statistical Association* 58 (301): 236–244.

Wied, Dominik. 2015. "A nonparametric test for a constant correlation matrix". *Econometric Reviews*: 1–21.

Wied, Dominik, Walter Kraemer, and Herold Dehling. 2012. "Testing for a change in correlation at an unknown point in time using an extended functional delta method". *Econometric Theory* 28 (03): 570–589.

Wied, Dominik, et al. 2012. “A new fluctuation test for constant variances with applications to finance”. *Metrika* 75 (8): 1111–1127.

Wold, Svante, Kim Esbensen, and Paul Geladi. 1987. “Principal component analysis”. *Chemometrics and intelligent laboratory systems* 2 (1): 37–52.

Wooldridge, Jeffrey. 2010. *Econometric analysis of cross section and panel data.* MIT press.

WorldBank. 2016a. *GDP per capita. Data retrieved from World Development Indicators: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD.*

– . 2016b. *PPP conversion factor. Data retrieved from World Development Indicators: http://data.worldbank.org/indicator/PA.NUS.PPP.*

– . 2016c. “The World Bank: DataBank”. `http://data.worldbank.org/indicator`.

Yitzhaki, Shlomo. 1983. “On an Extension of the Gini Inequality Index”. *International Economic Review* 24 (3): 617–628.

Young, Terry W. 1991. “Calmar Ratio: A Smoother Tool”. *Future Magazine.*

Zakamouline, Valeri. 2009. “The Choice of Performance Measure Does Influence the Evaluation of Hedge Funds”, no. 1983: 1–36.