DISS. ETH NO. 24754

# STABLE PORTFOLIO DESIGN USING BAYESIAN CHANGE POINT MODELS AND GEOMETRIC SHAPE FACTORS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

TOBIAS SETZ

MSc ETH CSE, ETH Zurich

born on 18.06.1981

citizen of
Dintikon AG

accepted on the recommendation of

Prof. Dr. Matthias Troyer
Prof. Dr. Rosangela Loschi
Dr. Martin Mächler

2017

# Contents

# Abstract

This thesis deals with the design of stable portfolios. The term stability in the context of portfolio design was introduced by Wuertz [2010]. First the drivers of instabilities within financial returns are identified. Second a mathematical model is defined that can explain to which extend the drivers are activated.

A possible approach would be to explain unstable times of financial returns by outliers or dominant oscillators. For this thesis two different approaches are pursued. The first approach assumes that unstable times are characterized through a change within the underlying dynamics of the returns. The second approach assumes that the shape of the investment universe is changing for such times. The first approach is generally applicable for the univariate case and the second approach for the multivariate case.

A mathematical model for the first approach is the Bayesian change point (BCP) algorithm as introduced by Barry and Hartigan [1993]. The first part of this thesis implements the BCP model using a new approach which does not only react on changes within the trend but also the variance (see also Loschi et al. [1999]). In comparison to already existing implementations the variance is modelled using a generalized inverse Gaussian (GIG) distribution.

A mathematical model for the second approach are the geometric shape factors (GSF). Those are explained in detail for the first time within the second part of this thesis. The investment universe is described as an image of the feasible set. The GSF define an ellipse which has the same area, orientation, eccentricity and centre as the image of the feasible set. The GSF can be described as a function of the image moments of the feasible set. Additionally it is shown how the results of the BCP method can be used to calculate the image of the feasible set. To achieve this the orthogonalized Gnanadesikan–Kettenring (OGK) estimator [Maronna and Zamar, 2002] is used to calculate the covariance matrix. This leads to positive semidefinite and approximately affine equivariant estimations which are suitable for quadratic optimization.

In the last part of this thesis the concepts of the BCP method and the GSF are used to design portfolios which are controlled by the results of these analyses. The

results of the BCP method are used to design a dynamically hedged signal portfolio. The orientation as a geometric shape factor is used to define the distance of the target portfolio [Markowitz, 1952] to the minimum variance portfolio. In the last step these two portfolios are combined.

Besides the introduction of mathematical concepts (first and second part of this thesis) this thesis has its focus on the application of these concepts to construct novel dynamic portfolios (third part of the thesis). Particular attention is given to ensure that parameters are either set globally for any kind of returns input or if this is not possible to define routines which are independently determining these parameters based on the input without any external intervention. This to avoid that results of the past are mainly driven by parameters which are not valid anymore in the future.

# Zusammenfassung

Diese Arbeit beschäftigt sich mit dem Entwurf von stabilen Portfolios. Der Begriff Stabilität im Portfolio-Kontext wurden eingeführt von Wuertz [2010]. Als erstes macht man sich Gedanken über die Treiber von Instabilitäten in Finanzrenditen. Als zweites wird ein mathematisches Modell gesucht, dass beschreiben kann, wie stark diese Treiber aktiv sind.

Ein möglicher Ansatz wäre, dass sich instabile Phasen in Finanzrenditen durch Ausreisser oder dominante Oszillatoren erklären lassen. In dieser Arbeit werden zwei andere Ansätze verfolgt. Der erste Ansatz geht davon aus, dass sich in instabilen Phasen die Dynamik des zugrundeliegenden Prozesses der Renditen fundamental verändert. Der zweite Ansatz geht davon aus, dass sich die Form des Anlageuniversums in solchen Phasen verändert. Der erste Ansatz ist grundsätzlich für den eindimensionalen Fall und der zweite Ansatz für den multivariaten Fall.

Ein mathematisches Modell für den ersten Ansatz ist der Bayesian change point (BCP) Algorithmus von Barry and Hartigan [1993]. Im ersten Teil dieser Arbeit wird das BCP Modell mit einem neuen Ansatz implementiert, der nicht nur auf Veränderungen im Trend der Renditen reagiert, sondern auch in der Varianz (siehe dazu auch Loschi et al. [1999]). Neu dabei ist, dass die Varianz als generalisierte inverse Gauss (GIG) Verteilung beschrieben wird.

Ein mathematisches Modell für den zweiten Ansatz sind die geometrischen Form-Faktoren (GSF) [Wuertz, 2010]. Diese werden im zweiten Teil der Arbeit erstmals ausführlich erklärt. Dabei wird das Anlageuniversum durch ein Bild der Menge aller möglichen Portfolios beschrieben. Die GSF definieren eine Ellipse welche die gleiche Fläche, Orientierung, Exzentrizität sowie das gleiche Zentrum wie das Bild aller möglichen Portfolios hat. Die GSF können als Funktion der Bild-Momente der Menge aller möglichen Portfolios beschrieben werden. Ausserdem wird gezeigt, wie sich die Resultate der BCP Methode benutzen lassen, um das Bild aller möglichen Portfolios zu berechnen. Dazu wird der orthogonalisierte Gnanadesikan–Kettenring (OGK) Schätzer [Maronna and Zamar, 2002] verwendet um die Kovarianz-Matrix zu berechnen. Dies führt zu Schätzungen

welche positiv semidefinit und schätzungsweise affin-äquivariant sind und sich daher mit quadratischen Optimierern verwenden lassen.

Im letzten Teil dieser Arbeit werden die Konzepte der BCP Methode und der GSF benutzt um Portfolios zu definieren, die sich Anhand der Resultate dieser Analysen steuern lassen. Die Resultate der BCP Methode werden benutzt um ein dynamisch abgesichertes Signal-Portfolio zu konstruieren. Die Orientierung als geometrischer Form-Faktor wird benutzt um den Abstand des optimierten Zielportfolios [Markowitz, 1952] zum Portfolio mit der kleinsten Varianz zu definieren. Danach werden diese zwei Portfolios miteinander kombiniert.

Neben der Einführung von mathematischen Konzepten (erster und zweiter Teil der Arbeit) legt diese Arbeit auch einen Fokus auf die Umsetzung dieser Konzepte zur Konstruktion neuartiger dynamischer Portfolios (dritter Teil der Arbeit). Dabei wird auch darauf geachtet, dass sich Parameter entweder global für beliebige Renditen als Input definieren lassen und wenn dies nicht möglich ist Routinen zu definieren, welche diese Parameter selbständig basierend auf dem Input festlegen ohne dass von aussen in das Modell eingegriffen wird. Dies um zu verhindern, dass Resultate in der Vergangenheit mehrheitlich durch Parameter getrieben sind, welche in der Zukunft keine Gültigkeit mehr haben.

# 1 Introduction

Part I of this thesis deals with the Bayesian change point (BCP) model as introduced by Barry and Hartigan [1992, 1993] and later extended by Loschi et al. [1999, 2003] (see Chapter 2 and Chapter 3). Given a set of data the model involves repeated Bayesian analyses of contiguous subsets of that data (clusters). These analyses are used to calculate a probability of change at any point within the dataset. Barry and Hartigan [1992, 1993] assume that the data within a given cluster is normally (N) distributed and that the mean of that normal distribution is also normally (N) distributed depending on the variance. The variance has an improper distribution which is constant for the whole dataset. The change points are therefore only based on changes within the mean of the data. Loschi et al. [1999, 2003] extended the model by assuming that the variance is distributed according to an inverse gamma (IG) distribution inside every cluster (the BCP N-NIG model). The change points are therefore based on changes within the mean and/or the variance of the data. The contribution of this thesis is to modify the model of Loschi et al. [2003] by replacing the IG distribution with a generalized inverse Gaussian (GIG) distribution (the BCP N-NGIG model) (see Chapter 5). Additionally it is shown how to implement the BCP N-NGIG model such that it is numerically stable for large datasets. The IG distribution is a special case of the GIG distribution. Therefore the BCP N-NGIG model represents a generalisation of the BCP N-NIG model (see Chapter 6). The N-NGIG model (without using it in the BCP context) was studied by Thabane and Safiul Haq [1999]. They derived the posterior distribution and the prediction distribution of future responses and showed that these distributions belong to the family of generalized modified Bessel distributions. For the BCP model the marginal likelihood of the data and the posterior marginal (or conditional) distributions of the mean and the variance are needed. This thesis contributes to the topic by deriving closed form expressions of these distributions and showing that they belong to the family of generalized hyperbolic (GH) and GIG distributions (see Chapter 4). A special emphasis was also given on the question of how to choose the hyperparameters in an automated way. For the prior of the mean it is proposed to use the sample mean. Since inside the clusters the estimation of the mean

can move around this global guess. For the prior of the variance it is proposed to assume that the variance is distributed according to a gamma distribution which is a special case of the GIG distribution. The hyperparameters can then be set by using the sample variance and using an additional parameter to control the dispersion of the variance (see Chapter 6). It was shown that the BCP N-NGIG model as well as the BCP model of Barry and Hartigan [1993] can both detect changes within the mean of the underlying dynamic reliably if the variance is kept constant. For the case where the variance changes as well it was shown that the BCP N-NGIG model can still reliably detect the changes while the model of Barry and Hartigan can get in trouble. Additionally both models were compared on a real financial returns series where the BCP N-NGIG model shows smother results for the estimation of the mean.

While Part I of this thesis presented risk measures for the univariate case, Part II of this thesis presents a new kind of multivariate risk measures. They are based on the portfolio optimization approach as introduced by Markowitz [1952]. In finance a portfolio is a collection of investable assets (the investment universe) where the investment weights describe which fraction of the available capital is invested into which of these underlying assets (see Chapter 8). Every possible combination of the investment weights represents a different portfolio. Visualizing the performance and variance of all these portfolios leads to the feasible set (see Chapter 9). The feasible set is basically a function of the performance and the covariance matrix of the underlying assets. There are various approaches to calculate these figures. In this thesis the sample estimator and the robust orthogonalized Gnanadesikan–Kettenring (OGK) estimator as introduced by Maronna and Zamar [2002] are considered. The OGK estimator presents a procedure to calculate the covariances from the univariate formulation of the variance. This thesis contributes to the topic by defining a new estimator which is based on the univariate BCP model of Part I (see Chapter 9). Thus respecting possible changes within the underlying dynamic of the investment universe. The results of the BCP model are used to estimate the performances and the variances of the underlying assets. To calculate the covariances the same procedure as for the OGK estimator is used. Since this approach is computationally very expensive a second approach is presented where the covariances are calculated by mixing the BCP variances with the sample correlations. Note that for this approach the correlations are not reacting on possible changes within the underlying dynamic of the investment universe. However, it was shown that the results of these two estimators lead to very similar results (see Chapter 11). The shape of the feasible set offers valuable insights into the current state of the investment universe. If the feasible set is understood

as an image one can calculate the image moments of the feasible set (see Chapter 10). Using the image moments it is possible to approximate the feasible set through an ellipse that has the same area, centre, orientation and eccentricity as the feasible set. These figures were called the geometric shape factors (GSF). The idea was first outlined by Wuertz [2010]. This thesis contributes to the topic by presenting for the first time the detailed mathematical framework to calculate these multivariate risk measures. Another interesting family of factors are the Hu moments [Hu, 1962] which are invariant under translation, scale and rotation. The GSF and the Hu moments were visualized for different investment universes on a sliding window of 36 month for the sample, robust and BCP estimators (see Chapter 11). Monthly data was used. The results generally show that the GSF are varying greatly over time. Even when using a reasonably long window of 36 month. The GSF can be associated with a geometric meaning (see Chapter 8). For example the orientation describes the direction of the risk premium. If the direction is positive taking more risk would generally lead to more performance. And the eccentricity describes the correlation between risk and performance. The higher the correlation the higher the probability that taking more risk does indeed lead to more or less performance (depending on the direction). The analyses most notably show that the risk premium is not granted unconditionally over time.

The univariate and multivariate risk measures introduced in Part I and Part II of this thesis were consequently applied in Part III of this thesis to design stable portfolios. The term stability in the context of portfolio design was first presented by Wuertz [2010]. The contributions of this thesis to the topic is in explaining the concept in more detail for the first time (see Section 13.1). Namely that stability is associated to a model of change. The model of changing dynamics (Section 14.1) considers times as stable when the underlying dynamic of the investment universe is unlikely to change. The BCP model is measuring this. If the underlying dynamics change the BCP model is quick in updating the performances and the variances of the underlying assets. The model of shape shifts (see Section 15.1) considers times as stable when the shape of the investment universe does not change. The GSF are measuring this. If the shape of the investment universe changes an adjustment of allocation tactics might make sense. This thesis contributes to the topic of portfolio design by introducing three novel approaches to design portfolios. All portfolios were backtested on a sliding window of 36 month by applying the results out of sample and using monthly data. The first portfolio (the BCP-Signal-Portfolio (BSP)) calculates the investment weights based on univariate BCP analyses of the underlying assets (see Chapter 14) and does therefore respect the model

of changing dynamics. For that the concept of signal portfolios was introduced. The results of the BCP analyses are used to calculate a signal between 0 and 1 for every underlying asset. Then every asset gets a maximum investment weight. The actual investment weight is the multiplication of this maximum weight and the signal. The signals describe the hedging degree of any underlying asset. Hedging can be achieved by moving the difference of the maximum weight and the actual weight to a cash account. The natural benchmark for this portfolio is a constant 50% hedge of any underlying asset. The backtest for the European equity universe shows that the dynamic hedge based on the BCP signals results in a more appealing risk/return profile than the passive benchmark. For the second portfolio the investment weights are calculated through Markowitz portfolio optimization (see Chapter 15). The target portfolio is defined based on the orientation of the GSF and does therefore respect the model of shape shifts. To get full control over the target portfolio the Lambda-2-Portfolio (L2P) was introduced. The weights of the target portfolio are calculated by using the multivariate BCP estimator. Therefore the portfolio does also respect the model of changing dynamics. The backtest for the European bonds and precious metals universe shows that this portfolio offers a more appealing risk/return profile than the portfolio where the capital is distributed equally to all underlying assets (EWP). The dynamic L2P using the BCP estimator was tested against various variants where it was considered the best choice in means of performance, risk profile and diversification. This includes the L2P using the sample and robust OGK estimators. The L2P where the orientation is kept constant. And a multi objective approach to define the target portfolio. The third portfolio is a combination of the BSP and the dynamic L2P using the BCP estimator (see Chapter 16). Instead of moving capital to a cash account to hedge the underlying assets this capital is moved to the L2P. The backtest shows that this portfolio offers a more appealing risk/return profile than the portfolio where the capital is distributed equally to the EWP of the European equity universe and the EWP of the European bonds and precious metals universe. All three actively managed portfolios (the BSP, the L2P and the combination) offer more preferable return and risk profiles than their passive natural benchmarks. They are simple, dynamic and stable portfolios for risk-averse and long-term investors. Simple in the sense that the portfolios can be implemented by just buying or selling the underlying assets. Which means that the realization of the portfolios do not involve concepts like short selling, leverage or the purchase of complicated financial instruments. Dynamic and stable in the sense that the investment weights are adapted at any point in time based on thorough statistical analyses that assess the stability of the investment universe. This

in order to realize the performance that the market realistically offers from a long-term perspective while reducing the risks.

# Part I

# Bayesian Change Point Model for the Normal-GIG Prior of the Gaussian Distribution

# 2 Introduction

It is well konwn that financial returns show various stylized facts such as heavy tails, skewness or dependency structures [Cont, 2001]. Many methods have been developed to model these properties. Such as more flexible distributions (stable distribution [Mandelbrot, 1963; Fama, 1965], generalized hyperbolic (GH) distribution [Barndorff-Nielsen, 1977], generalized lambda distribution [Ramberg and Schmeiser, 1974]) or time series models (ARIMA [Box and Jenkins, 1970], GARCH [Engle, 1982; Bollerslev, 1986]).

This thesis focuses on the Bayesian analysis for multiple change point problems as introduced by Barry and Hartigan [1992, 1993] and later extended by Loschi et al. [1999, 2003]. In the remainder of this thesis it will be referred as the Bayesian change point (BCP) method or model. A given set of data is usually modelled with one common distribution. In contrast the BCP method assumes that a given set of data can be modelled by contiguous subsequences (clusters) where the data inside each cluster has a common distribution. As the number of clusters is a priori unknown the BCP model also provides inference about it. The case where the data is modelled with one common distribution is therefore a special case of the BCP method. In the most extreme case every observation is generated by its own distribution. Furthermore the number of change points has not to be known as opposed to many other methods. Generally speaking all possible cluster combinations (partitions) are considered and influence the result. The more probable a partition is (based only on the data) the more influence that partition will have on the result.

Chapter 3 reviews the current state of the BCP method. The BCP method consists of two important concepts. First the Bayesian analysis of a given distribution. And second the combination of the posterior distributions from that analysis with the product partition model (PPM). Chapter 4 presents the Bayesian analysis of the Gaussian distribution using a Normal-Generalized-Inverse-Gaussian (NGIG) prior. The posterior conditional and marginal distributions as well as the marginal likelihood are derived. It is especially shown that these distributions do all belong to the family of generalized hyperbolic (GH) and generalized inverse Gaussian (GIG) distributions.

These results will be used to set-up the final BCP model in Chapter 5. Since averaging over all possible partitions is computationally very expensive the partitions are sampled through a Markov Chain Monte Carlo (MCMC) approach. To sample the partitions the marginal likelihood is used. In combination with the derived posterior marginal distributions the unknown parameters of the normal distribution can be sampled. Additionally it is shown how to generate samples of the parameters if only the posterior conditional distributions are known. This can be achieved by using a Gibbs sampling scheme. The numerical implementation of the BCP method poses some challenges. More precisely it is the evaluation of the Bessel function that breaks for large samples. It is shown how to neutralize that problem by first doing the critical computations in log-space and second using an asymptotic expansion of the Bessel function if necessary.

Chapter 6 discusses and evaluates the model. The main difference of the new model to the model of Loschi et al. [2003] is analysed. The role of the hyperparameters is examined and the method is tested on artificial datasets and an actual financial returns series. The method will be used again in Part III to construct a stable equity portfolio.

# 3 The Bayesian Change Point Method

A change point can be described as the position within a dataset where the data before and after that point cannot be explained anymore with the same statistical model. Many work has been done in identifying such points. Non-Bayesian approaches include e.g. binary segmentation [Edwards and Cavalli-Sforza, 1965; Scott and Knott, 1974; Sen and Srivastava, 1975] and Markov switching models [Hamilton, 1989]. This thesis focuses on the Bayesian method of Barry and Hartigan [1992, 1993] which is designed to find multiple change points for the normal mean. The BCP method does not necessarily look for distinct change points but assigns a probability of change to any sample inside a given dataset.

Loschi et al. [1999] extended the results from Barry and Hartigan [1993] to find multiple change points for the normal mean and variance. For that they combined the PPM with the Normal-Inverse-Gamma (NIG) prior. Additionally they introduced a Gibbs sampling scheme to solve the model. Later Loschi et al. [2003] extended the model further by introducing a prior distribution for the parameter $p$ which describes the probability of having a change point at any position within a given dataset.

In this thesis the PPM will be combined with the Normal-Generalized-Inverse-Gaussian (NGIG) prior (see Chapter 5). Thus introducing a generalisation of the model of Loschi et al. [2003] (see Section 6.1). The NGIG prior was studied by Thabane and Safiul Haq [1999] where they derived the posterior distribution and the prediction distribution of future responses in the context of generalized modified Bessel distributions. For the BCP analysis the marginal likelihood and the posterior marginal distributions are needed and were therefore derived by showing that they belong to the family of generalized hyperbolic (GH) and generalized inverse Gaussian (GIG) distributions (see Chapter 4). For the PPM the same prior distribution for the parameter $p$ is used as by Loschi et al. [2003]. The partitions are sampled through an MCMC approach by using the marginal likelihood. The unknown mean and variance are sampled by using the posterior marginal distributions. Alternatively the posterior conditional distributions could be used through a Gibbs sampling scheme and were therefore derived as well (see Chapter 4).

## 3.1 The general algorithm

In this section the general setup of the BCP model as described by Barry and Hartigan [1992, 1993] will be summarized and generalized for the case where more than one parameter is unknown. The BCP method combines Bayesian inference with the PPM. The PPM assumes that a given dataset can undergo sudden changes in its dynamic at unknown positions. The dataset can be partitioned into different blocks separated by the change points (partitions). Every partition can be described by a common multivariate probability distribution. For every partition the parameters can be found by the Bayesian analysis. Since the partitions itself do also follow a probability distribution every sample inside the dataset is in principle generated by a different dynamic.

### 3.1.1 Partitions

Given a set of observations $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ the partitions of the PPM are expressed as

$$\begin{gathered} \rho = (i_0, i_1, \ldots, i_b) \\ 0 = i_0 < i_1 < i_2 < \ldots < i_b = n \\ X_{ij} = \{X_{i+1}, X_{i+2}, \ldots, X_j\} \end{gathered} \tag{3.1}$$

where $i_k$ describes a change point at $i_k + 1$ and $b$ is the number of breakpoints. The PPM is specified as

$$\begin{gathered} \vec{\boldsymbol{\theta}} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n\} \\ \boldsymbol{\theta}_k = \{\theta_1, \theta_2, \ldots, \theta_h\} \\ f(\vec{X}|\vec{\boldsymbol{\theta}}, \rho) = \prod_{k=1}^{b} f_{i_{k-1}i_k}(X_{i_{k-1}i_k}|\boldsymbol{\theta}_{i_k}) = \prod_{ij\in\rho} f_{ij}(X_{ij}|\boldsymbol{\theta}_k) \end{gathered} \tag{3.2}$$

where $f_{ij}$ is a multivariate probability distribution for the observations $X_{ij}$ inside the cluster $ij = \{i+1, i+2 \ldots, j\}$ with parameters $\boldsymbol{\theta}_k$ and $k \in ij$. Which means that all $\boldsymbol{\theta}_k$ for $k \in ij$ are identical if the partition $\rho$ is known. For example if $f_{ij}$ is a normal distributions then $\boldsymbol{\theta}_k = \{\mu_k, \sigma_k^2\}$ (see also Eq. (5.1)).

### 3.1.2 Bayesian inference

By applying Bayes' law on a given partition we can formulate the posterior distribution of the unknown parameters as

$$f_{ij}(\boldsymbol{\theta}_k|X_{ij}) = \frac{f_{ij}(X_{ij}|\boldsymbol{\theta}_k)f_{ij}(\boldsymbol{\theta}_k)}{f_{ij}(X_{ij})} \tag{3.3}$$

where $f_{ij}(X_{ij}|\boldsymbol{\theta}_k)$ is the likelihood of the data and $f_{ij}(\boldsymbol{\theta}_k)$ the prior distribution of the parameters. The marginal likelihood (or datafactor) is calculated as

$$f_{ij}(X_{ij}) = \int f_{ij}(X_{ij}|\boldsymbol{\theta}_k)f_{ij}(\boldsymbol{\theta}_k)d^h\boldsymbol{\theta}_k \tag{3.4}$$

The expectation value of an unknown parameter for a given partition $E_{ij}(\theta_{k,l}|X_{ij})$ can be calculated from the posterior marginal distribution:

$$f_{ij}(\theta_{k,l}|X_{ij}) = \int f_{ij}(\boldsymbol{\theta}_k|X_{ij})d^{h-1}\{\boldsymbol{\theta}_k \setminus \theta_{k,l}\} \tag{3.5}$$

where $1 \leq l \leq h$. $\theta_{k,l}$ describes one of the $h$ unknown parameters of the likelihood ($\boldsymbol{\theta}_k$) at position $k$. Note that we assume that we know the partition. Therefore $\theta_{k,l}$ is identical for any $k \in ij$. Using a numerical Gibbs sampling approach it is also possible to calculate the expectation values from the posterior conditional distribution:

$$f_{ij}(\theta_{k,l}|X_{ij}, \{\boldsymbol{\theta}_k \setminus \theta_{k,l}\}) = \frac{f_{ij}(X_{ij}|\boldsymbol{\theta}_k)f_{ij}(\boldsymbol{\theta}_k)}{f_{ij}(X_{ij}|\{\boldsymbol{\theta}_k \setminus \theta_{k,l}\})} \tag{3.6}$$

The prior probability distribution of the partitions is defined as a product partition distribution:

$$f(\rho) = K\prod_{k=1}^{b} c_{i_{k-1}i_k} = K\prod_{ij\in\rho} c_{ij} \tag{3.7}$$

where $c_{ij}$ are the cohesions. The posterior probability distribution of the partitions can be calculated as

$$f(\rho|\vec{X}) = \frac{f(\vec{X}|\rho)f(\rho)}{\sum_\rho f(\vec{X}|\rho)f(\rho)} = \frac{f(\vec{X}|\rho)f(\rho)}{f(\vec{X})} \tag{3.8}$$

where $f(\vec{X}|\rho) = \prod_{ij\in\rho} f_{ij}(X_{ij})$ is the product of the marginal likelihoods as defined in Eq. (3.4). The expectation values of the unknown parameters independent from the

partitions can be calculated as

$$E(\theta_{k,l}|\vec{X}) = \sum_{\rho} E_{ij}(\theta_{k,l}|X_{ij}) f(\rho|\vec{X}) \tag{3.9}$$

where $1 \leq k \leq n$ which allows to calculate the expectation values of the $h$ unknown parameters at any position $k$.

As one can assume solving this model can be rather challenging. From an analytical point of view the problem can be separated in first deriving the expectation values for a given partition ($E_{ij}(\theta_{k,l}|X_{ij})$) and second for the case that is independent of the partitions ($E(\theta_{k,l}|\vec{X})$).

For the first part various solutions are available depending on the chosen likelihood and prior distribution since this problem represents a classical Bayesian distribution analysis. The needed expressions for the normal likelihood and the NGIG prior were not available and are derived in Chapter 4.

For the second part analytical solutions are not available. In principle an exact numerical enumeration would be possible by summing over all possible partitions $\rho$. For small sample sizes this might even be feasible. But the numerical complexity is of order $O(n!)$ and grows therefore dramatically with sample size. Also the finite numerical precision of a computer will become more and more a problem as the sample size grows. To solve this problem an MCMC approach is used as proposed by Barry and Hartigan [1992] which reduces the numerical complexity to $O(n)$. For the MCMC approach the finite numerical precision becomes a problem as well. This will be addressed in Chapter 5.

### 3.1.3 Markov Chain Monte Carlo

The Markov sampling technique [Metropolis et al., 1953; Hammersley and C., 1964; Hastings, 1970] is a prominent concept in statistical physics. Instead of summing over all possible partitions we can approximate the result by generating partition samples that are distributed according to the posterior probability distribution of the partitions ($f(\rho|\vec{X})$).

The random partition samples are generated through a Markov Chain. First we describe the partitions as

$$\begin{aligned} \rho &= (U_1, U_2, \ldots, U_n) \\ U_k &\in \{0, 1\},\ U_n = 1 \end{aligned} \tag{3.10}$$

where $U_k = 1$ describes a change point at position $k + 1$. This is equivalent with the

definition in Eq. (3.1). To start the process an initial partition is generated where $U_k = 0$ for $i < n$ and $U_n = 1$. A new partition is generated by iterating through the old partition (from position 1 to $n-1$) and at each position $i$ we set $U_k = 1$ with probability $p_i$:

$$\begin{aligned} p_i &= \frac{r}{1+r} \\ r &= \frac{f(\rho|\vec{X}, U_k = 1)}{f(\rho|\vec{X}, U_k = 0)} \end{aligned} \tag{3.11}$$

where $r$ is the ratio between $f(\rho|\vec{X}, U_k = 1)$ of accepting the change point and $f(\rho|\vec{X}, U_k = 0)$ of not accepting the change point. This allows us to generate $M$ partitions that are distributed according to $f(\rho|\vec{X})$:

$$\begin{aligned} P &= \{\rho_1, \rho_2, \ \ldots, \rho_M\} \\ \rho_m &= (U_{1,m}, U_{2,m}, \ldots, U_{n,m}) \end{aligned} \tag{3.12}$$

Using these partitions the expectation values of the parameters can be calculated as a simple average of the estimated parameters where the partitions are given:

$$\langle E(\theta_{k,l}|\vec{X})\rangle = \frac{1}{M}\sum_{\rho\in P} E_{ij}(\theta_{k,l}|X_{ij}) = \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_{k,l,m} \tag{3.13}$$

$$\hat{\theta}_{k,l}(\rho) = \hat{\theta}_{k,l,m} = E_{ij}(\theta_{k,l}|X_{ij}) \tag{3.14}$$

$$m_{k,l} = \langle E(\theta_{k,l}|\vec{X})\rangle = \langle \hat{\theta}_{k,l}\rangle = \overline{\hat{\boldsymbol{\theta}}_{k,l}} \tag{3.15}$$

where $\hat{\theta}_{k,l,m}$ is the expectation value of an unknown parameter given the partition $\rho_m$ and $m_{k,l}$ is an approximation of the expectation value of an unknown parameter independent from the partitions. The vector $\hat{\boldsymbol{\theta}}_{k,l} = (\hat{\theta}_{k,l,1}, \hat{\theta}_{k,l,2}, \ldots, \hat{\theta}_{k,l,M})$ holds the $M$ parameter samples at position $k$ calculated from the partition samples $P$. The variance within the estimation of $E(\theta_{k,l}|\vec{X})$ is calculated as

$$s^2_{k,l} = \frac{n}{n-1}\left(\overline{\hat{\boldsymbol{\theta}}_{k,l}\cdot\hat{\boldsymbol{\theta}}_{k,l}} - \overline{\hat{\boldsymbol{\theta}}_{k,l}}\cdot\overline{\hat{\boldsymbol{\theta}}_{k,l}}\right) \tag{3.16}$$

An alternative for the variance is to calculate the Highest Posterior Density (HPD) interval [Box and Tiao, 1992] of the sample expectation values $\hat{\boldsymbol{\theta}}_{k,l}$. The HPD interval $h_{k,l}(\hat{\boldsymbol{\theta}}_{k,l})$ is the shortest possible interval for a given target probability. If that probability is chosen to be e.g. 95% then the posterior probability that the true estimation of $m_{k,l}$

lies within the HPD interval is 95%. The posterior probabilities $p_k$ if the point $k+1$ is a change point can be calculated from the partition samples as

$$p_k = \frac{1}{M} \sum_{m=1}^{M} U_{k,m} \tag{3.17}$$

Note that the MCMC approach brings many advantages. The numerical complexity is reduced to $O(n)$. The posterior probability distribution of the partitions $f(\rho|\vec{X})$ as defined in Eq. (3.8) needs only to be calculated partially. This since the datafactor $f(\vec{X})$ and many more factors from the expression $f(\vec{X}|\rho)$ cancel out in Eq. (3.11).

### 3.1.4 Summary

Combining Bayesian inference, the product partition model and a Markov chain Monte Carlo approach the BCP model can be solved efficiently. To set-up the BCP model one can follow these steps:

1. To calculate the expectation values of the unknown parameters $\hat{\theta}_{k,l}(\rho) = E_{ij}(\theta_{k,l}|X_{ij})$ given the partition one needs to derive the posterior marginal (or conditional) distributions $f_{ij}(\theta_{k,l}|X_{ij})$ (see Eq. (3.5)).

2. To generate the sample partitions one needs to derive the marginal likelihood $f_{ij}(X_{ij})$ (see Eq. (3.4)).

3. Also needed to generate the sample partitions is the prior probability distribution of the partitions $f(\rho)$. For that the cohesions $c_{ij}$ have to be defined (see Eq. (3.7)).

Once these expressions are derived the expectation values independent from the distributions $\hat{\theta}_{k,l} = E(\theta_{k,l}|\vec{X})$ (see Eq. (3.9)) can be calculated. If this is numerically not feasible one can approximate the results ($m_{k,l} = \langle E(\theta_{k,l}|\vec{X})\rangle$) through the MCMC approach as presented in Section 3.1.3.

Note that for steps one to three it is preferable to derive full analytical expressions. Having the posterior marginal distributions in the form of a probability distribution it is usually a simple task to calculate the expectation values. The same is true for the marginal likelihood where we need the evaluation of the distribution itself. A numerical approach for these expressions might not be feasible. Especially for large datasets there are various numerical pitfalls in solving the respective integral expressions.

## 3.2 The specific implementation of Barry and Hartigan

In this section the specific implementation of Barry and Hartigan [1993] will be summarized. We will follow the procedure as described in Section 3.1.4. The derivations of the following expressions are given in Barry and Hartigan [1993].

### 3.2.1 Prior Design

In the paper of Barry and Hartigan [1993] the BCP model is specified for one parameter which is the mean $\mu$ of the normal distribution where it is assumed that the variance $\sigma^2$ is constant over time. The prior distribution for the mean is specified to be a normal distribution as well with unknown hyperparameters $\mu_0$ and $\sigma_0^2$:

$$(X_{ij}|\mu_k, \sigma^2) \quad \overset{iid}{\sim} \quad N(\mu_k, \sigma^2) \tag{3.18}$$

$$(\mu_k|\mu_0, \sigma_0^2) \quad \overset{iid}{\sim} \quad N(\mu_0, \sigma_0^2/(j-i)) \tag{3.19}$$

where the division by $j - i$ expresses that larger deviations from $\mu_0$ are expected in short blocks than in long blocks. The likelihood $(X_{ij}|\mu_k, \sigma^2)$ is a multivariate normal distribution where is is assumed that the observations $X_{ij} = X_{i+1}, X_{i+2}, \ldots, X_j$ are independent and identically distributed.

The cohesions $c_{ij}$ are defined after Yao [1984] as

$$\begin{aligned} c_{ij} &= (1-p)^{j-i-1}p \quad \text{if } j > n \\ &= (1-p)^{j-i-1} \quad \text{if } j = n \end{aligned} \tag{3.20}$$

which leads to the prior probability distribution of the partitions (see Eq. (3.7)) as

$$f(\rho|p) = p^{b-1}(1-p)^{n-b} \tag{3.21}$$

where $0 \leq p \leq 1$ describes the probability of having a change point at any given position $k$ within the dataset $\vec{X}$ and $b$ is the number of change points within the partition $\rho$.

### 3.2.2 Posterior Design

The expectation value of the unknown parameter $\mu_k(\rho)$ given the partition can then be calculated as

$$\hat{\mu}_k(\rho) = E_{ij}(\mu_k|X_{ij}) = (1-w)\bar{X}_{ij} + w\mu_0 \tag{3.22}$$

$$\bar{X}_{ij} = \sum_{l=i+1}^{j} \frac{X_l}{j-i} \tag{3.23}$$

$$w = \frac{\sigma^2}{\sigma^2 + \sigma_0^2} \tag{3.24}$$

and the marginal likelihood is given as

$$\begin{aligned} f_{ij}(X_{ij}) = \int f_{ij}(X_{ij}|\mu) f_{ij}(\mu) d\mu = \frac{1}{(2\pi\sigma^2)^{(j-i)/2}} \left(\frac{\sigma^2}{\sigma_0^2+\sigma^2}\right)^{1/2} e^{V_{ij}} \\ V_{ij} = -\frac{\sum_{l=i+1}^{j}(X_l - \bar{X}_{ij})^2}{2\sigma^2} - \frac{(j-i)(\bar{X}_{ij}-\mu_0)^2}{2(\sigma_0^2+\sigma^2)} \end{aligned} \tag{3.25}$$

In principle all the expressions to calculate the unknown parameters $\mu_k$ using the MCMC approach as described in Section 3.1.3 are available at this point. Nonetheless Barry and Hartigan [1993] made some Bayesian refinements which will be discussed in the next section.

### 3.2.3 Refinements

The Bayesian refinements concern the variance of the data $\sigma^2$ and the hyperparametrs of the model as described in the previous section. The hyperparametrs are the mean $\mu_0$ and the variance $\sigma_0^2$ of the unknown parameters $\mu_k$, and the transition probability $p$. These parameters are described through the following prior probability distributions:

$$f(\sigma^2) = \frac{1}{\sigma^2} \cdot I(0 \le \sigma^2 \le \infty) \tag{3.26}$$

$$f(\mu_0) = 1 \cdot I(-\infty \le \mu_0 \le \infty) \tag{3.27}$$

$$f(w) = \frac{1}{w_0} \cdot I(0 \le w \le w_0) \tag{3.28}$$

$$f(p) = \frac{1}{p_0} \cdot I(0 \le p \le p_0) \tag{3.29}$$

where $I$ is the indicator function and $w = \sigma^2/(\sigma^2 + \sigma_0^2)$. This reduces the four hyperparameters to the two hyperparameters $w_0$ that describes the maximum signal-to-noise ratio and $p_0$ that describes the maximum probability of change. Note that the variance of the data and the hyperparameters are constant for the whole dataset. They do not vary depending on the position $k$ inside the dataset.

As a consequence $f(\vec{X}|\rho)$ is not just the product of the marginal likelihoods $f_{ij}(X_{ij})$ anymore. The expression $f(\vec{X}|\rho)$ can be derived by first integrating over the hyperparameter $\sigma^2$ as

$$\begin{aligned} f(\vec{X}|\rho,\mu_0,w) &= \int f(\vec{X}|\rho,\sigma^2,\mu_0,w)f(\sigma^2)d\sigma^2 \propto \int \frac{1}{\sigma^2}\prod_{ij\in\rho} f_{ij}(X_{ij})d\sigma^2 \\ &\propto \frac{w^{b/2}}{[W+Bw+wn(\mu_0-\bar{X})^2]^{n/2}} \end{aligned} \tag{3.30}$$

$$\bar{X} = \sum_{i=1}^{n}\frac{X_i}{n}, \quad B = \sum_{ij\in\rho}(j-i)(\bar{X}_{ij}-\bar{X})^2, \quad W = \sum_{ij\in\rho}\sum_{l=i+1}^{j}(X_l-\bar{X}_{ij})^2$$

and then over the remaining parameters $\mu_0$ and $w$ as

$$\begin{aligned} f(\vec{X}|\rho) &\propto \int\int f(\vec{X}|\rho,\mu_0,w)f(\mu_0)f(w)d\mu_0 dw \\ &\propto \int_0^{w_0}\frac{w^{(b-1)/2}}{(W+Bw)^{(n-1)/2}}dw \end{aligned} \tag{3.31}$$

Since the hyperparameters are now described through distributions it will be necessary to calculate the expectation values of $\mu_0$ and $w$ to calculate $\hat{\mu}_k(\rho)$:

$$\hat{\mu}_k(\rho) = (1-w^*)\bar{X}_{ij} + w^*\mu_0^* \tag{3.32}$$

where $w^* = E(w|\vec{X})$ and $\mu_0^* = E(\mu_0|\vec{X})$. The expectation values $\mu_0^*$ and $w^*$ given the partitions $\rho$ are calculated as

$$\mu_0^* = E(\mu_0|\vec{X},\rho,w) = \int \mu_0\frac{f(\vec{X}|\rho,\mu_0,w)f(\mu_0)}{f(\vec{X}|\rho,w)}d\mu_0 = \bar{X} \tag{3.33}$$

$$w^* = E(w|\vec{X},\rho) = \int w\frac{f(\vec{X}|\rho,w)f(w)}{f(\vec{X}|\rho)}dw = \frac{\int_0^{w_0}\frac{w^{(b+1)/2}}{(W+Bw)^{(n-1)/2}}dw}{\int_0^{w_0}\frac{w^{(b-1)/2}}{(W+Bw)^{(n-1)/2}}dw} \tag{3.34}$$

Additionally the expectation value of $\sigma^2$ can be calculated as

$$\begin{aligned}\sigma^{2*} = E(\sigma^2|\vec{X},\rho) &= \int\int\int \sigma^2 \frac{f(\vec{X}|\rho,\sigma^2,\mu_0,w)f(\sigma^2)f(\mu_0)f(w)}{f(\vec{X}|\rho)} d\mu_0 dw d\sigma^2 \\ &= \frac{1}{n-3}\frac{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W+Bw)^{(n-3)/2}} dw}{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W+Bw)^{(n-1)/2}} dw}\end{aligned} \tag{3.35}$$

Since $p$ is described through a probability distribution the prior probability distribution of the partitions $f(\rho)$ has to be re-evaluated:

$$f(\rho) = \int_0^{p_0} f(\rho|p)f(p)dp = \frac{1}{p_0}\int_0^{p_0} p^{b-1}(1-p)^{n-b}dp \tag{3.36}$$

where $b$ is the number of blocks in $\rho$. Combining Eq. (3.31) and Eq. (3.35) the posterior probability distribution of the partitions can be approximated as:

$$\begin{aligned}f(\rho|\vec{X}) &= \frac{f(\vec{X}|\rho)f(\rho)}{\sum_\rho f(\vec{X}|\rho)f(\rho)} \\ &\propto \left[\int_0^{w_0} \frac{w^{(b-1)/2}}{(W+Bw)^{(n-1)/2}} dw\right]\left[\frac{1}{p_0}\int_0^{p_0} p^{b-1}(1-p)^{n-b}dp\right]\end{aligned} \tag{3.37}$$

Note that this is actually sufficient to calculate the MCMC partition samples since it is only necessary to calculate ratios of the probability distribution of the partitions (see Eq. (3.11)).

### 3.2.4 Implementation

For a numerical implementation the main problem is the integral of Eq. (3.31). For example, to generate the partition samples the ratio of Eq. (3.11) has to be calculated as

$$r = \frac{f(\rho|\vec{X},U_k=1)}{f(\rho|\vec{X},U_k=0)} = \frac{\int_0^{w_0} \frac{w^{b/2}}{(W_1+B_1w)^{(n-1)/2}} dw \int_0^{p_0} p^b(1-p)^{n-b-1}dp}{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0+B_0w)^{(n-1)/2}} dw \int_0^{p_0} p^{b-1}(1-p)^{n-b}dp} \tag{3.38}$$

In Erdman and Emerson [2007] the Eq. (3.31)-like integrals are simplified as incomplete beta integrals. For example, the ratio of Eq. (3.11) above can then be re-expressed as

$$r = C \cdot \frac{\int_0^{\frac{B_1 w_0/W_1}{1+B_1 w_0/W_1}} p^{(b+2)/2}(1-p)^{(n-b-3)/2}dp \int_0^{p_0} p^b(1-p)^{n-b-1}dp}{\int_0^{\frac{B_0 w_0/W_0}{1+B_0 w_0/W_0}} p^{(b+1)/2}(1-p)^{(n-b-2)/2}dp \int_0^{p_0} p^{b-1}(1-p)^{n-b}dp} \tag{3.39}$$

$$C = \left(\frac{W_0}{W_1}\right)^{\frac{n-b-2}{2}} \left(\frac{B_0}{B_1}\right)^{\frac{b+1}{2}} \sqrt{\frac{W_1}{B_1}}$$

which is numerically stable.

Using the ratio of Eq. (3.39) it is possible to generate $M$ random partition samples according to $f(\rho|\vec{X})$ as described in Section 3.1.3. For every sample the expectation values $\hat{\mu}_k(\rho)$ can be calculated using Eq. (3.32). Using Eq. (3.15) an approximation of the expectation values $m_k$ independent from the partitions can be calculated as

$$m_k = \langle E(\mu_k|\vec{X})\rangle = \frac{1}{M}\sum_{\rho\in P} \hat{\mu_k}(\rho) = \overline{\hat{\boldsymbol{\mu}}_k} \tag{3.40}$$

where $\hat{\boldsymbol{\mu}}_k$ is the vector that holds the estimations $\hat{\mu_k}(\rho)$ of the $M$ partition samples. The variance within the estimations can be calculated following Eq. (3.16) as

$$s_k^2 = \frac{n}{n-1}\left(\overline{\hat{\boldsymbol{\mu}}_k \cdot \hat{\boldsymbol{\mu}}_k} - \overline{\hat{\boldsymbol{\mu}}_k} \cdot \overline{\hat{\boldsymbol{\mu}}_k}\right) \tag{3.41}$$

and the posterior probabilities $p_k$ if the point $k+1$ is a change point following Eq. (3.17) as

$$p_k = \frac{1}{M}\sum_{m=1}^{M} U_{k,m} \tag{3.42}$$

where $U_{k,m}$ is the k-th entry of the m-th sample partition $\rho_m$ (see also Eq. (3.12)).

### 3.2.5 Evaluation

The specific implementation of Barry and Hartigan as described in Section 3.2.4 was implemented by Erdman and Emerson [2007] and is available in the `R` [R Core Team, 2015] package **bcp**. Thorough evaluations of the specific implementation of Barry and Hartigan are documented in Barry and Hartigan [1993], and Erdman and Emerson [2007]. The intention of this section is to demonstrate for which kind of breakpoint problems the specific implementation of Barry and Hartigan might not be suited.

While the general model as defined in Section 3.1 does allow for changes in any parameter of the likelihood the specific implementation of Barry and Hartigan assumes a constant variance (homoscedasticity). Therefore it makes sense to assume that the specific implementation of Barry and Hartigan gets into troubles if the different regimes of the dataset have not the same variance.

In Fig. 3.1 a dataset where only the mean changes is examined. The means are detected reliably. In Fig. 3.2 not only the mean changes but also the variance. As expected the change points cannot be detected reliably anymore. Especially regions with a high variance lead to large posterior probabilities.

In the remainder of Part I the general model will be solved such that it can not only detect changes for the mean but also for the variance. To achieve this goal the needed posterior distributions for the normal likelihood and the NGIG prior are derived in Chapter 4. The obtained results will then be combined with the PPM in Chapter 5.

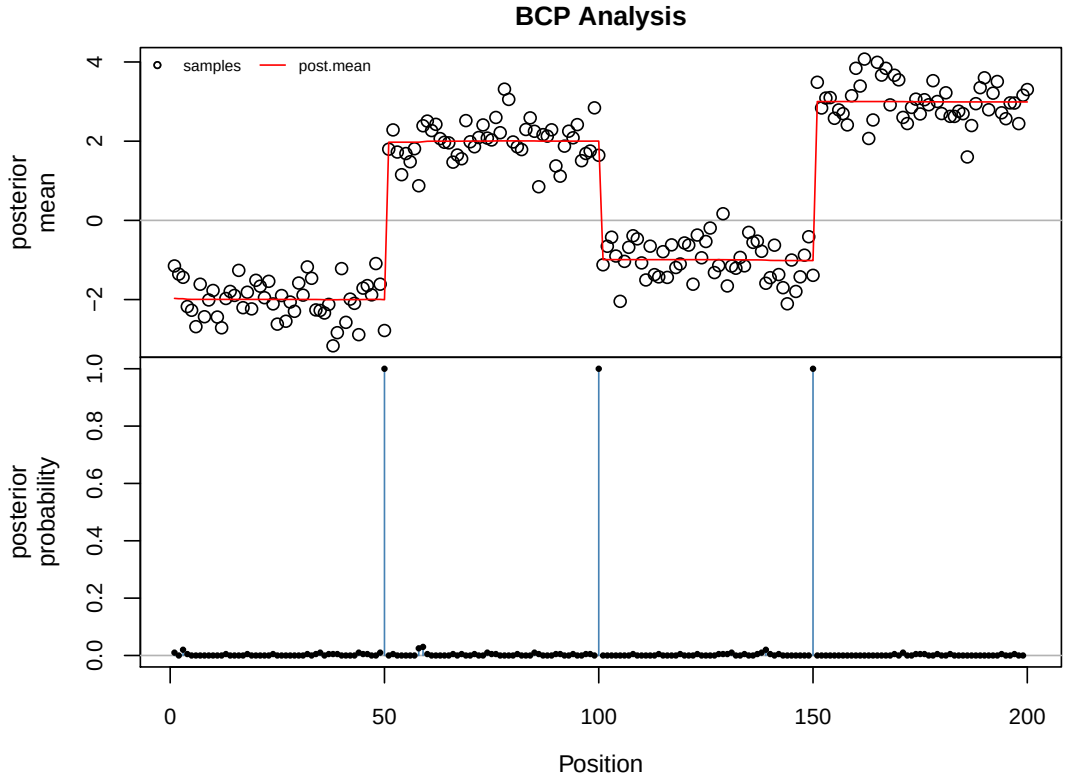


Figure 3.1: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2, the second one a mean of 2, the third one a mean of -1 and the fourth one a mean of 3. The standard deviation is constant and defined to be 0.5 for all clusters.

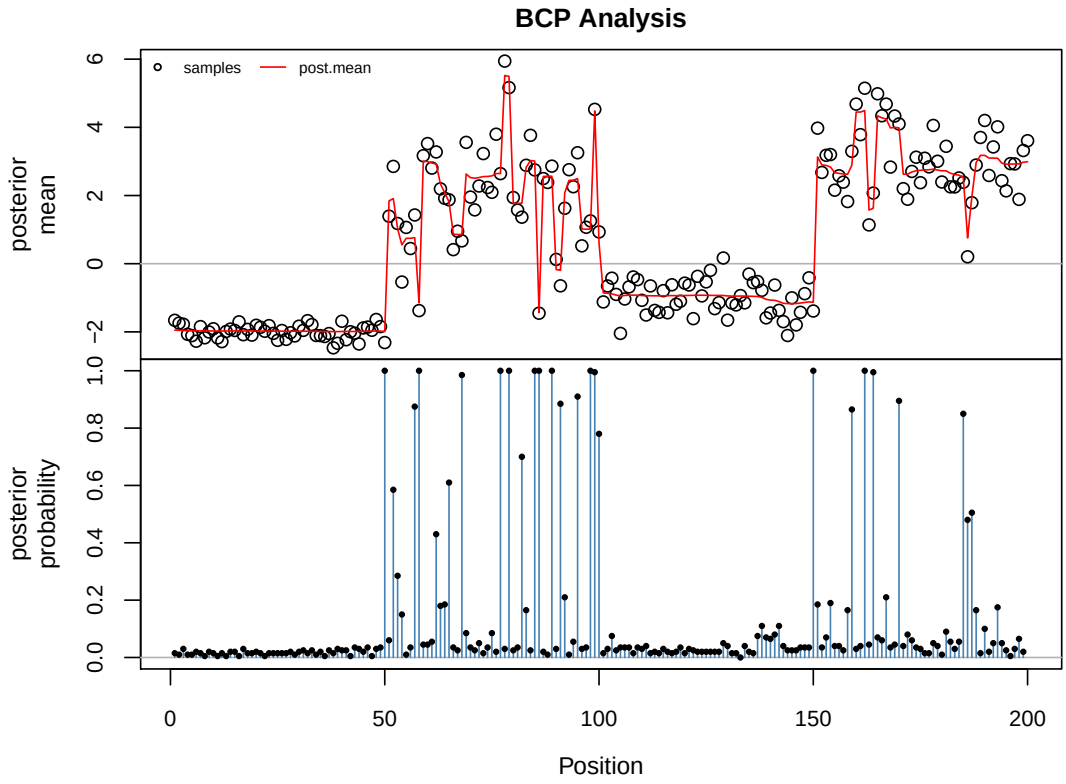


Figure 3.2: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2 and a standard deviation of 0.2, the second one a mean of 2 and a standard deviation of 1.5, the third one a mean of -1 and a standard deviation of 0.5 and the fourth one a mean of 3 and a standard deviation of 1.

# 4 Normal-GIG-Prior of the Gaussian Distribution

## 4.1 Prior Distributions

By applying Bayes' law it is possible to infer on the unknown parameters $\boldsymbol{\theta}$ of a given distribution $f(\vec{X}|\boldsymbol{\theta})$ as

$$f(\boldsymbol{\theta}|\vec{X}) = \frac{f(\vec{X}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\vec{X})} \tag{4.1}$$

where $f(\vec{X}|\boldsymbol{\theta})$ is the likelihood of the data, $f(\boldsymbol{\theta})$ the prior distribution of the parameters and $f(\vec{X})$ the marginal likelihood.

For our model the likelihood is defined to be a normal distribution (N), the prior distribution for the mean of the likelihood is defined to be a normal distribution (N) as well, and the prior distribution for the variance of the likelihood is defined to be a generalized inverse Gaussian (GIG) distribution. All random variables are assumed to be independent and identically distributed (iid). The N-NGIG model can be summarized as

$$f(\mu, \sigma^2|\vec{X}) = \frac{f(\vec{X}|\mu, \sigma^2)f(\mu, \sigma^2)}{f(\vec{X})} = \frac{f(\vec{X}|\mu, \sigma^2)f(\mu|\mu_0, v\sigma^2)f(\sigma^2|\psi, \chi, \lambda)}{f(\vec{X})} \tag{4.2}$$

$$(\vec{X}|\mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2) \tag{4.3}$$

$$(\mu|\mu_0, v\sigma^2) \overset{iid}{\sim} N(\mu_0, v\sigma^2) \tag{4.4}$$

$$(\sigma^2|\psi, \chi, \lambda) \overset{iid}{\sim} GIG(\psi, \chi, \lambda) \tag{4.5}$$

Depending on the choice of the likelihood and the prior distributions there are various analytical solutions available. If analytical solutions for the posterior marginal (or conditional) distributions and the posterior likelihood are available they can be used directly to complete the BCP analysis using the MCMC approach (see Section 3.1.3). In the following these expressions are derived for the N-NGIG model as defined in Eq. (4.2)

through Eq. (4.5).

To achieve this the posterior conditional distributions are derived in Section 4.2.2 as

$$f(\mu|\sigma^2, \vec{X}) = \frac{f(\vec{X}|\mu, \sigma^2) f(\mu, \sigma^2)}{f(\vec{X}|\sigma^2)} \tag{4.6}$$

$$f(\sigma^2|\mu, \vec{X}) = \frac{f(\vec{X}|\mu, \sigma^2) f(\mu, \sigma^2)}{f(\vec{X}|\mu)} \tag{4.7}$$

the posterior marginal distributions in Section 4.2.3 as

$$f(\mu|\vec{X}) = \int f(\mu, \sigma^2|\vec{X}) d\sigma^2 \tag{4.8}$$

$$f(\sigma^2|\vec{X}) = \int f(\mu, \sigma^2|\vec{X}) d\mu \tag{4.9}$$

and the marginal likelihood in Section 4.2.4 as

$$f(\vec{X}) = \int \int f(\vec{X}|\mu, \sigma^2) f(\mu, \sigma^2) d\mu d\sigma^2 \tag{4.10}$$

## 4.2 Posterior Distributions

### 4.2.1 Definitions

Some distributions are extensively used throughout this section. These are the iid multivariate normal distribution to model the likelihood (see also Eq. (4.3)) as

$$f(\vec{X}|\mu, \sigma^2) = \prod_{i=1}^{n} \left( \frac{1}{2\pi\sigma^2} \right)^{1/2} \exp\left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right) \tag{4.11}$$

$$= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left( -\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \right) \tag{4.12}$$

the normal distribution to model the prior for the mean (see also Eq. (4.4)) as

$$f(\mu|\mu_0, v\sigma^2) = \left( \frac{1}{2\pi v\sigma^2} \right)^{1/2} \exp\left( -\frac{(\mu - \mu_0)^2}{2v\sigma^2} \right) \tag{4.13}$$

and the GIG distribution to model the prior for the variance (see also Eq. (4.5)) as

$$f(\sigma^2|\psi,\chi,\lambda) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})}\sigma^{2(\lambda-1)}\exp\left(-\frac{\psi\sigma^2+\chi\sigma^{-2}}{2}\right) \tag{4.14}$$

where $K_\lambda$ is a modified Bessel function of the third kind.

Some of the distributions that will be derived in the next sections are distributed according to the symmetric iid generalized hyperbolic (GH) distribution. The general multivariate case is defined as

$$\begin{aligned} f(\vec{X}|\lambda,\alpha,\mu,\Delta,\delta,\boldsymbol{\beta}) = \\ \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\left(\frac{1}{\sqrt{|\Delta|}}\right)\left(\frac{\sqrt{\sigma^2+(\boldsymbol{x}-\mu)'\Delta^{-1}(\boldsymbol{x}-\mu)}}{\alpha}\right)^{\lambda-\frac{n}{2}}\left(\frac{\sqrt{\alpha^2-\boldsymbol{\beta}'\Delta\boldsymbol{\beta}}}{\delta}\right)^{\lambda} \\ \times\left(\frac{K_{\lambda-\frac{n}{2}}(\alpha\sqrt{\sigma^2+(\boldsymbol{x}-\mu)'\Delta^{-1}(\boldsymbol{x}-\mu)})}{K_\lambda(\delta\sqrt{\alpha^2-\boldsymbol{\beta}'\Delta\boldsymbol{\beta}})}\right)\exp\left(\boldsymbol{\beta}'(\boldsymbol{x}-\mu)\right) \end{aligned} \tag{4.15}$$

where $K_\lambda$ is again a modified Bessel function of the third kind. Fore the symmetric ($\beta = 0$) and iid case ($\Delta = I_n$ where $I_n$ is the identity matrix) the expression simplifies to

$$\begin{aligned} f(\vec{X}|\lambda,\alpha,\mu,\delta) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\alpha}{\tilde{c}}\right)^{\frac{n}{2}}\left(\frac{\tilde{c}}{\delta}\right)^{\lambda}\left(\frac{K_{\lambda-\frac{n}{2}}(\alpha\tilde{c})}{K_\lambda(\alpha\delta)}\right) \\ \tilde{c}(\delta,\boldsymbol{x},\mu) &= \sqrt{\delta^2+\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned} \tag{4.16}$$

for the univariate case the expression reduces further to

$$\begin{aligned} f(x|\lambda,\alpha,\mu,\delta) &= \left(\frac{1}{2\pi}\right)^{\frac{1}{2}}\left(\frac{\alpha}{\tilde{c}}\right)^{\frac{1}{2}}\left(\frac{\tilde{c}}{\delta}\right)^{\lambda}\left(\frac{K_{\lambda-\frac{1}{2}}(\alpha\tilde{c})}{K_\lambda(\alpha\delta)}\right) \\ \tilde{c}(\delta,x,\mu) &= \sqrt{\delta^2+(x-\mu)^2} \end{aligned} \tag{4.17}$$

### 4.2.2 Posterior Conditional Distributions

For the final BCP model in Chapter 5 where the results of this chapter are applied the marginal distributions are used. Note that it would also be possible to apply the MCMC procedure as described in Section 3.1.3 by using the conditional distributions (Gibbs

sampling). The conditional distribution of the variance as defined in Eq. (4.7) will be an important ingredient to calculate the marginal distributions and is therefore derived within this section. The conditional distribution of the mean as defined in Eq. (4.6) is a well known expression. Only for the sake of completeness it will be derived.

If we assume that the variance is known then the posterior conditional distribution of the mean can be calculated as

$$
\begin{aligned}
f(\mu|\sigma^2,\vec{X}) &\propto f(\vec{X}|\mu,\sigma^2)f(\mu) \\
&\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right)\exp\left(-\frac{(\mu-\mu_0)^2}{2v\sigma^2}\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}((x_i-\overline{x})-(\mu-\overline{x}))^2\right)\exp\left(-\frac{1}{2v\sigma^2}(\mu-\mu_0)^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\mu-\overline{x})^2\right)\exp\left(-\frac{1}{2v\sigma^2}(\mu^2-2\mu\mu_0+\mu_0^2)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}n(\mu-\overline{x})^2-\frac{1}{2v\sigma^2}(\mu^2-2\mu\mu_0)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(n\mu^2-2n\mu\overline{x}+n\overline{x}^2+\frac{\mu^2}{v}-\frac{2\mu\mu_0}{v}\right)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left((n+\frac{1}{v})\mu^2-2\mu(n\overline{x}+\frac{\mu_0}{v})\right)\right) \\
&\propto \exp\left(-\frac{nv+1}{2v\sigma^2}\left(\mu^2-2\mu\frac{nv\overline{x}+\mu_0}{nv+1}\right)\right) \qquad (4.18)
\end{aligned}
$$

which is the kernel of a normal distribution. Therefore we conclude that the posterior conditional distribution of the mean is normally distributed as

$$(\mu|\sigma^2,\vec{X}) \sim N\left(\tilde{\mu}=\frac{nv\overline{x}+\mu_0}{nv+1},\quad \tilde{\sigma}^2=\frac{v\sigma^2}{nv+1}\right) \qquad (4.19)$$

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad (4.20)$$

where $\overline{x}$ is the mean of the observations $\vec{X}$.

If we assume that the mean is known then the posterior conditional distribution of the

variance can be calculated as

$$\begin{aligned}
f(\sigma^2|\mu,\vec{X}) &\propto f(\vec{X}|\mu,\sigma^2)f(\mu)f(\sigma^2) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n+1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\exp\left(-\frac{1}{2\sigma^2}\frac{(\mu-\mu_0)^2}{v}\right) \\
&\times \left(\sigma^2\right)^{\lambda-1}\exp\left(-\frac{\psi\sigma^2+\chi\sigma^{-2}}{2}\right) \\
&\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1}\exp\left[-\frac{1}{2}\left(\psi\sigma^2+\frac{1}{\sigma^2}\left(\chi+\frac{(\mu-\mu_0)^2}{v}+\sum_{i=1}^{n}(x_i-\mu)^2\right)\right)\right]
\end{aligned}$$

which is the kernel of a GIG distribution. Therefore we conclude that the posterior conditional distribution of the variance is distributed according to a GIG distribution as

$$\begin{aligned}
(\sigma^2|\mu,\vec{X}) \sim GIG\,(\;\tilde{\psi}=\psi,\quad \tilde{\lambda}=\lambda-\frac{n+1}{2}, \\
\tilde{\chi}=\chi+\frac{(\mu-\mu_0)^2}{v}+\sum_{i=1}^{n}(x_i-\mu)^2\Bigg)
\end{aligned} \tag{4.21}$$

Note that there were no factors dropped that contain $\mu$. Therefore we can make use of this result to calculate the posterior marginal distributions.

### 4.2.3 Posterior Marginal Distributions

To derive the posterior marginal distribution of the mean we can make use of our findings in Eq. (4.21). If we assume that the variance is unknown then the posterior marginal distribution of the mean can be formulated as

$$\begin{aligned}
f(\mu|\vec{X}) &\propto \int f(\vec{X}|\mu,\sigma^2)f(\mu)f(\sigma^2)d\sigma^2 \\
&\propto \int \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1}\exp\left[-\frac{1}{2}\left(\psi\sigma^2+\frac{1}{\sigma^2}\left(\chi+\sum_{i=1}^{n}(x_i-\mu)^2+\frac{(\mu-\mu_0)^2}{v}\right)\right)\right]d\sigma^2 \\
&\propto \int \left(\sigma^2\right)^{\tilde{p}-1}\exp\left[-\frac{1}{2}\left(\tilde{a}\sigma^2+\tilde{b}\sigma^{-2}\right)\right]d\sigma^2
\end{aligned}$$

where

$$\tilde{p}=\lambda-\frac{n+1}{2},\quad \tilde{a}=\psi,\quad \tilde{b}=\chi+\sum_{i=1}^{n}(x_i-\mu)^2+\frac{(\mu-\mu_0)^2}{v} \tag{4.22}$$

The expression within the integral is the kernel of a GIG distribution; therefore:

$$\begin{aligned} f(\mu|\vec{X}) &\propto 2K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right)\left(\frac{\tilde{b}}{\tilde{a}}\right)^{\tilde{p}/2} \\ &\propto K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right)\tilde{b}^{\tilde{p}/2} \end{aligned} \tag{4.23}$$

The expression for $\tilde{b}$ can be reformulated as

$$\begin{aligned} \tilde{b} &= \chi + \sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2) + \frac{1}{v}(\mu^2 - 2\mu_0\mu + \mu_0^2) \\ &= \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} + \mu^2(n + \frac{1}{v}) - 2\mu\frac{v\sum_{i=1}^{n} x_i + \mu_0}{v} \\ &= \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} + \frac{nv+1}{v}\left(\mu^2 - 2\mu\frac{nv\overline{x} + \mu_0}{nv+1}\right) \\ &= \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)} + \frac{nv+1}{v}\left(\mu - \frac{nv\overline{x} + \mu_0}{nv+1}\right)^2 \\ &= \frac{nv+1}{v}\left[\frac{v}{nv+1}\left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)}\right) + \left(\mu - \frac{nv\overline{x} + \mu_0}{nv+1}\right)^2\right] \end{aligned}$$

Using this result Eq. (4.23) can be reformulated as

$$\begin{aligned} f(\mu|\vec{X}) &\propto K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right)\tilde{b}^{\tilde{p}/2} \\ &\propto K_{\lambda-\frac{n}{2}-\frac{1}{2}}\left[\psi^{1/2}\left(\frac{nv+1}{v}\right)^{1/2}\right. \\ &\qquad \left.\left(\frac{v}{nv+1}\left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)}\right) + \left(\mu - \frac{nv\overline{x} + \mu_0}{nv+1}\right)^2\right)^{1/2}\right] \\ &\quad \times \left[\left(\frac{v}{nv+1}\left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)}\right) + \left(\mu - \frac{nv\overline{x} + \mu_0}{nv+1}\right)^2\right)^{1/2}\right]^{\lambda-\frac{n}{2}-\frac{1}{2}} \end{aligned}$$

which is the kernel of a univariate symmetric iid GH distribution as defined in Eq. (4.17). Therefore we conclude that the posterior marginal distribution of the mean is distributed

according to a GH distribution as

$$
\begin{aligned}
(\mu|\vec{X}) \sim GH\,(\ \tilde{\lambda} = \lambda - \frac{n}{2}, \quad \tilde{\alpha} = \psi^{1/2}\left(\frac{nv+1}{v}\right)^{1/2}, \\
\tilde{\mu} = \frac{nv\overline{x}+\mu_0}{nv+1}, \quad \tilde{\Delta} = 1, \quad \tilde{\beta} = 0, \\
\tilde{\delta} = \sqrt{\frac{v}{nv+1}\left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\right)}
\end{aligned}
\tag{4.24}
$$

If we assume that the mean is unknown then the posterior marginal distribution of the variance can be formulated as

$$
\begin{aligned}
f(\sigma^2|\vec{X}) &\propto \int f(\vec{X}|\mu,\sigma^2) f(\mu) f(\sigma^2) d\mu \\
&\propto \int \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + \frac{1}{\sigma^2}\left(\chi + \sum_{i=1}^{n}(x_i-\mu)^2 + \frac{(\mu-\mu_0)^2}{v}\right)\right)\right] d\mu \\
&\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \\
&\times \int \exp\left[-\frac{\psi\sigma^2}{2} - \frac{1}{2\sigma^2}\left(\chi + \sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2) + \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{v}\right)\right] d\mu \\
&\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \exp\left[-\frac{1}{2}(\psi\sigma^2 + \chi\sigma^{-2}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 - \frac{\mu_0^2}{2v\sigma^2}\right] \\
&\times \int \exp\left[-\frac{n}{2\sigma^2}(\mu^2 - 2\mu\overline{x}) - \frac{1}{2v\sigma^2}(\mu^2 - 2\mu\mu_0)\right] d\mu \\
&\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + (\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v})\sigma^{-2}\right)\right] \\
&\times \int \exp\left[-\frac{1}{2\sigma^2}\left((n+\frac{1}{v})\mu^2 - 2\mu(n\overline{x} + \frac{\mu_0}{v})\right)\right] d\mu \\
&\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + (\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v})\sigma^{-2}\right)\right] \\
&\times \int \exp\left[-\frac{nv+1}{2v\sigma^2}\left(\mu^2 - 2\mu\left(\frac{nv\overline{x}+\mu_0}{nv+1}\right)\right)\right] d\mu
\end{aligned}
$$

where the expression within the integral is the kernel of a normal distribution. Note that

the expression within the integral was already derived in Eq. (4.18) where the factors containing $\sigma^2$ outside the integral were discarded. Solving the integral expression leads to:

$$\begin{aligned} f(\sigma^2|\vec{X}) &\propto \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + (\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v})\sigma^{-2}\right)\right] \\ &\times \left(2\pi\frac{v\sigma^2}{nv+1}\right)^{\frac{1}{2}} \exp\left[\frac{nv+1}{2v\sigma^2}\left(\frac{nv\overline{x}+\mu_0}{nv+1}\right)^2\right] \qquad (4.25) \\ &\propto \left(\sigma^2\right)^{\lambda-\frac{n}{2}-1} \\ &\times \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + \left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\right)\sigma^{-2}\right)\right] \qquad (4.26) \end{aligned}$$

which is the kernel of a GIG distribution. Therefore we conclude that the posterior marginal distribution of the variance is distributed according to a GIG distribution as

$$\begin{aligned} (\sigma^2|\vec{X}) \sim GIG\,(\; \tilde{\psi} = \psi, \quad \tilde{\lambda} = \lambda - \frac{n}{2}, \\ \tilde{\chi} = \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\Bigg) \end{aligned} \qquad (4.27)$$

### 4.2.4 Marginal Likelihood

The last ingredient that is needed for the BCP analysis is the marginal likelihood. It can be formulated as

$$f(\vec{X}) = \int\int f(\vec{X}|\mu,\sigma^2) f(\mu) f(\sigma^2) d\mu d\sigma^2$$

using Eq. (4.25) and Eq. (4.26) we get:

$$\begin{aligned} f(\vec{X}) &= \left(\frac{1}{2\pi}\right)^{\frac{n+1}{2}} \left(\frac{1}{v}\right)^{\frac{1}{2}} \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} \left(2\pi\frac{v}{nv+1}\right)^{\frac{1}{2}} \\ &\times \int \left(\sigma^2\right)^{\lambda-\frac{n}{2}-1} \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + \left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\right)\sigma^{-2}\right)\right] d\sigma^2 \end{aligned}$$

where the expression within the integral is the kernel of a GIG distribution. Solving the integral expression leads to:

$$
\begin{aligned}
f(\vec{X}) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} 2K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right) \left(\frac{\tilde{b}}{\tilde{a}}\right)^{\tilde{p}/2} \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \left(\frac{\psi}{\chi}\right)^{\lambda/2} \left(\frac{\tilde{b}}{\tilde{a}}\right)^{\tilde{p}/2} \frac{K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right)}{K_\lambda(\sqrt{\psi\chi})}
\end{aligned} \tag{4.28}
$$

where

$$
\tilde{p} = \lambda - \frac{n}{2}, \quad \tilde{a} = \psi, \quad \tilde{b} = \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)} \tag{4.29}
$$

re-substituting $\tilde{p}$ and $\tilde{a}$ leads to:

$$
\begin{aligned}
f(\vec{X}) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \left(\frac{\psi}{\chi}\right)^{\lambda/2} \left(\frac{\tilde{b}}{\psi}\right)^{\frac{\lambda}{2}-\frac{n}{4}} \frac{K_{\lambda-\frac{n}{2}}\left(\sqrt{\psi\tilde{b}}\right)}{K_\lambda(\sqrt{\psi\chi})} \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \frac{\psi^{n/4}}{\chi^{\lambda/2}} \tilde{b}^{\frac{1}{2}(\lambda-\frac{n}{2})} \frac{K_{\lambda-\frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)}{K_\lambda(\psi^{1/2}\chi^{1/2})} \qquad (4.30) \\
&= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \left(\frac{\psi^{1/2}}{\tilde{b}^{1/2}}\right)^{n/2} \left(\frac{\tilde{b}^{1/2}}{\chi^{1/2}}\right)^{\lambda} \frac{K_{\lambda-\frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)}{K_\lambda(\psi^{1/2}\chi^{1/2})} \qquad (4.31)
\end{aligned}
$$

The expression for $\psi\tilde{b}$ can be reformulated as

$$
\begin{aligned}
\psi\tilde{b} &= \psi\left[\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\right] \\
&= \psi\left[\chi + \frac{nv^2\sum x_i^2 + v\sum x_i^2 + nv\mu_0^2 + \mu_0^2 - v^2\left(\sum x_i\right)^2 - 2v\mu_0\sum x_i - \mu_0^2}{v(nv+1)}\right] \\
&= \frac{\psi}{nv+1}\left[\chi(nv+1) + nv\sum x_i^2 - v\left(\sum x_i\right)^2 + \sum(x_i^2 - 2\mu_0 x_i + \mu_0^2)\right] \\
&= \frac{\psi}{nv+1}\left[\chi(nv+1) + n^2v(\overline{x^2} - \overline{x}^2) + \sum(x_i - \mu_0)^2\right] \\
&= \frac{\psi}{(nv+1)^{1/n}}\left[\chi(nv+1)^{1/n} + (nv+1)^{1/n-1}\left(\sum(x_i-\mu_0)^2 + n^2v(\overline{x^2} - \overline{x}^2)\right)\right]
\end{aligned}
$$

$$= \frac{\psi}{(nv+1)^{1/n}} \left[\chi(nv+1)^{1/n} + c\left(\sum(x_i-\mu_0)^2 + s\right)\right]$$

where

$$c = (nv+1)^{1/n-1}, \quad s = n^2 v(\overline{x^2} - \overline{x}^2) \tag{4.32}$$

Solving the following equation for $\mu$:

$$\begin{aligned}
\sum(x_i-\mu)^2 &= c\left(\sum(x_i-\mu_0)^2 + s\right) \\
n\overline{x^2} - 2n\mu\overline{x} + n\mu^2 &= cn\overline{x^2} - 2cn\mu_0\overline{x} + nc\mu_0^2 + cs \\
n\mu^2 - 2n\mu\overline{x} &= c(s + n\mu_0^2 - 2n\mu_0\overline{x}) + (c-1)n\overline{x^2} \\
\mu^2 - 2\mu\overline{x} &= c(\frac{s}{n} + \mu_0^2 - 2\mu_0\overline{x}) + (c-1)\overline{x^2} \\
(\mu-\overline{x})^2 &= \overline{x}^2 + c(\frac{s}{n} + \mu_0^2 - 2\mu_0\overline{x}) + (c-1)\overline{x^2} \\
\mu &= \overline{x} + \left[\overline{x}^2 + c(\frac{s}{n} + \mu_0^2 - 2\mu_0\overline{x}) + (c-1)\overline{x^2}\right]^{1/2}
\end{aligned}$$

and matching the expression $K_{\lambda-\frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)$ to the GH distribution as defined in Eq. (4.16) we can conclude that the posterior likelihood is distributed according to a GH distribution as

$$\begin{aligned}
(\vec{X}) \sim GH\,(\ \tilde{\lambda} = \lambda, \quad \tilde{\alpha} &= \left(\frac{\psi}{(nv+1)^{1/n}}\right)^{1/2}, \\
\tilde{\delta} &= \left(\chi(nv+1)^{1/n}\right)^{1/2}, \quad \tilde{\Delta} = I_n, \quad \tilde{\beta} = 0, \\
\tilde{\mu} &= \overline{x} + \left[\overline{x}^2 + c(\frac{s}{n} + \mu_0^2 - 2\mu_0\overline{x}) + (c-1)\overline{x^2}\right]^{1/2}\Big)
\end{aligned} \tag{4.33}$$

One can easily verify that inserting the expressions of Eq. (4.33) into Eq. (4.16) will lead to Eq. (4.31):

$$\begin{aligned}
\tilde{\alpha}^{\frac{n}{2}} &= \left(\psi^{1/2}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{4}} \\
\frac{1}{\tilde{\delta}}^{\tilde{\lambda}} &= \left(\frac{1}{\chi^{1/2}}\right)^{\lambda} \left(\frac{1}{nv+1}\right)^{\frac{\lambda}{2n}}
\end{aligned}$$

$$\begin{aligned}
\tilde{c}^{\lambda-\frac{n}{2}} &= \left[\chi(nv+1)^{1/n} + c\left(\sum(x_i-\mu_0)^2 + s\right)\right]^{\frac{1}{2}\left(\lambda-\frac{n}{2}\right)} \\
&= \left[(nv+1)^{1/n}\right]^{\frac{\lambda}{2}-\frac{n}{4}} \left[\tilde{b}^{1/2}\right]^{\left(\lambda-\frac{n}{2}\right)} \\
&= \left[\tilde{b}^{1/2}\right]^{\left(\lambda-\frac{n}{2}\right)} \left(\frac{1}{nv+1}\right)^{\frac{1}{4}-\frac{\lambda}{2n}}
\end{aligned}$$

Note that using Eq. (4.21) and Eq. (4.28) leads to the posterior distribution as

$$\begin{aligned}
f(\mu,\sigma^2|\vec{X}) &= \frac{f(\vec{X}|\mu,\sigma^2)f(\mu)f(\sigma^2)}{f(\vec{X})} \\
&= \left(\frac{nv+1}{2\pi v}\right)^{\frac{1}{2}} \frac{(\tilde{a}/\tilde{b})^{\tilde{p}/2}}{2K_{\tilde{p}}\left(\sqrt{\tilde{a}\tilde{b}}\right)} \left(\sigma^2\right)^{\lambda-\frac{n+1}{2}-1} \\
&\times \exp\left[-\frac{1}{2}\left(\psi\sigma^2 + \frac{1}{\sigma^2}\left(\chi + \frac{(\mu-\mu_0)^2}{v} + \sum_{i=1}(x_i-\mu)^2\right)\right)\right] \qquad (4.34)
\end{aligned}$$

where $\tilde{a}$, $\tilde{b}$ and $\tilde{p}$ are defined in Eq. (4.29).

## 4.3 Summary

It was shown that the posterior conditional distributions, the posterior marginal distributions and the posterior likelihood all belong to the family of generalized hyperbolic (GH) and generalized inverse Gaussian (GIG) distributions. The GH and GIG distributions are very well documented in literature. The moments and expectation values of the posterior distributions are therefore readily available. Also there are various software implementations such as the **GeneralizedHyperbolic** [Scott, 2014] or **ghyp** [Luethi and Breymann, 2013] packages for `R` that makes it e.g. also possible to generate random samples from all these distributions or to fit the parameters. Using the GH and GIG formulations for the posterior distributions makes it very convenient to work with the Bayesian model as derived within this chapter.

# 5 Bayesian Change Point Model

## 5.1 Prior Design

The model is set up for the unknown mean and variance of the normal distribution. The prior distribution of the mean is specified to be a normal distribution with unknown hyperparameters $\mu_0$ and $v$ while the prior distribution of the variance is specified to be a GIG distribution with unknown hyperparameters $\psi$, $\chi$ and $\lambda$:

$$(X_{ij}|\mu_k, \sigma_k^2) \overset{iid}{\sim} N(\mu_k, \sigma_k^2) \tag{5.1}$$

$$(\mu_k|\mu_0, v\sigma_k^2) \overset{iid}{\sim} N(\mu_0, v\sigma_k^2) \tag{5.2}$$

$$(\sigma_k^2|\psi, \chi, \lambda) \overset{iid}{\sim} GIG(\psi, \chi, \lambda) \tag{5.3}$$

Note that throughout this chapter the notation as introduced in Section 3.1 is used. It follows the procedure as described in Section 3.1.4 and already applied to describe the specific implementation of Barry and Hartigan in Section 3.2.

The same cohesions as for the specific implementation of Barry and Hartigan (see Eq. (3.20)) are used:

$$\begin{aligned} c_{ij} &= (1-p)^{j-i-1}p \quad \text{if } j > n \\ &= (1-p)^{j-i-1} \quad \text{if } j = n \end{aligned} \tag{5.4}$$

which leads to the prior probability distribution of the partitions as

$$f(\rho|p) = p^{b-1}(1-p)^{n-b} \tag{5.5}$$

where $0 \leq p \leq 1$ describes the probability of having a change point at any given position $k$ within the dataset $\vec{X}$ and $b$ is the number of change points within the partition $\rho$.

## 5.2 Posterior Design

We can now make use of our findings in Chapter 4. The expectation values of the unknown mean and variance can be derived from the posterior marginal distributions. All these distributions belong to the family of GH and GIG distributions. The properties of these distributions are well known. Besides the expectation values of the mean and the variance it would also be possible to examine other moments of these parameters. Like e.g. the variance of the mean or the variance of the variance.

The posterior marginal distribution of the mean $(\mu_k|X_{ij})$ was derived in Eq. (4.24). The expectation value $\hat{\mu}_k(\rho)$ follows from that distribution as

$$\begin{aligned}(\mu_k|X_{ij}) \sim GH\,(\;\tilde{\lambda} = \lambda - \frac{n}{2}, \quad \tilde{\alpha} = \psi^{1/2}\left(\frac{nv+1}{v}\right)^{1/2},\\ \tilde{\mu} = \frac{nv\overline{x}+\mu_0}{nv+1}, \quad \tilde{\Delta} = 1, \quad \tilde{\beta} = 0,\\ \tilde{\delta} = \sqrt{\frac{v}{nv+1}\left(\chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)}\right)}\end{aligned} \tag{5.6}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.7}$$

$$\hat{\mu}_k(\rho) = E_{ij}(\mu_k|X_{ij}) = \tilde{\mu} \tag{5.8}$$

where $ij = \{i+1, j+2, \ldots, j\}$ and $n = j - i$. Assuming we can't derive the posterior marginal distribution for some reason it is also possible to estimate the expectation value of the mean by using the posterior conditional distribution $(\mu_k|\sigma_k^2, X_{ij})$ of the mean (Gibbs sampling) which was derived in Eq. (4.19). The expectation value $\hat{\mu}_k^c(\rho)$ follows from that distribution as

$$(\mu_k|\sigma_k^2, X_{ij}) \sim N\left(\tilde{\mu} = \frac{nv\overline{x}+\mu_0}{nv+1}, \quad \tilde{\sigma}^2 = \frac{v\sigma_k^2}{nv+1}\right) \tag{5.9}$$

$$\hat{\mu}_k^c(\rho) = E_{ij}^c(\mu_k|X_{ij}) = \tilde{\mu} \tag{5.10}$$

Note that the marginal and conditional expectation values are identical. For higher moments this is not true anymore.

The posterior marginal distribution of the variance $(\sigma_k^2|X_{ij})$ was derived in Eq. (4.27).

The expectation value $\hat{\sigma}_k^2(\rho)$ follows from that distribution as

$$\begin{aligned}(\sigma_k^2|X_{ij}) \sim GIG\,(\;\tilde{\psi} = \psi, \quad \tilde{\lambda} = \lambda - \frac{n}{2},\\ \tilde{\chi} = \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)}\Bigg)\end{aligned} \tag{5.11}$$

$$\hat{\sigma}_k^2(\rho) = E_{ij}(\sigma_k^2|X_{ij}) = \left(\frac{\tilde{\chi}}{\tilde{\psi}}\right)^{\frac{1}{2}} \frac{K_{\tilde{\lambda}+1}\left((\tilde{\psi}\tilde{\chi})^{\frac{1}{2}}\right)}{K_{\tilde{\lambda}}\left((\tilde{\psi}\tilde{\chi})^{\frac{1}{2}}\right)} \tag{5.12}$$

The posterior conditional distribution $(\sigma_k^2|\mu_k, X_{ij})$ of the variance was derived in Eq. (4.21). The expectation value $(\hat{\sigma}_k^2)^c(\rho)$ follows from that distribution as

$$\begin{aligned}(\sigma_k^2|\mu_k, X_{ij}) \sim GIG\,(\;\tilde{\psi} = \psi, \quad \tilde{\lambda} = \lambda - \frac{n+1}{2},\\ \tilde{\chi} = \chi + \frac{(\mu_k - \mu_0)^2}{v} + \sum_{i=1}^{n}(x_i - \mu_k)^2\Bigg)\end{aligned} \tag{5.13}$$

$$(\hat{\sigma}_k^2)^c(\rho) = E_{ij}^c(\sigma_k^2|X_{ij}) = \left(\frac{\tilde{\chi}}{\tilde{\psi}}\right)^{\frac{1}{2}} \frac{K_{\tilde{\lambda}+1}\left((\tilde{\psi}\tilde{\chi})^{\frac{1}{2}}\right)}{K_{\tilde{\lambda}}\left((\tilde{\psi}\tilde{\chi})^{\frac{1}{2}}\right)} \tag{5.14}$$

Note that the marginal and conditional expectation values are not identical.

The marginal likelihood was derived in Eq. (4.33) as

$$\begin{aligned}(X_{ij}) \sim GH\,(\;\tilde{\lambda} = \lambda, \quad \tilde{\alpha} = \left(\frac{\psi}{(nv+1)^{1/n}}\right)^{1/2},\\ \tilde{\delta} = \left(\chi(nv+1)^{1/n}\right)^{1/2}, \quad \tilde{\Delta} = I_n, \quad \tilde{\beta} = 0,\\ \tilde{\mu} = \overline{x} + \left[\overline{x}^2 + c(\frac{s}{n} + \mu_0^2 - 2\mu_0\overline{x}) + (c-1)\overline{x^2}\right]^{1/2}\Bigg)\end{aligned} \tag{5.15}$$

However for the implementation we will use the formulation as given in Eq. (4.30) and Eq. (4.29):

$$f_{ij}(X_{ij}) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \frac{\psi^{n/4}}{\chi^{\lambda/2}} \tilde{b}^{\frac{1}{2}(\lambda - \frac{n}{2})} \frac{K_{\lambda - \frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)}{K_{\lambda}(\psi^{1/2}\chi^{1/2})} \tag{5.16}$$

$$\tilde{b} = \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x} + \mu_0)^2}{v(nv+1)} \tag{5.17}$$

## 5.3 Refinements

In the specific implementation of Barry and Hartigan [1993] all the hyperparameters were described through probability distributions. For this implementation the hyperparameters ($\mu_0$, $v$, $\psi$, $\chi$, $\lambda$) are assumed to be known a priori. Only for $p$ the beta distribution is used as

$$p \sim Beta(\alpha, \beta) \tag{5.18}$$

$$f(p|\alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \tag{5.19}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{5.20}$$

where $\Gamma$ is the gamma function. The final model has therefore seven hyperparameters $\mu_0$, $v$, $\psi$, $\chi$, $\lambda$, $\alpha$ and $\beta$. The parameters $\mu_0$ and $v$ are used to model the prior knowledge about the mean. The parameters $\psi$, $\chi$ and $\lambda$ are used to model the prior knowledge about the variance. And the parameters $\alpha$ and $\beta$ are used to model the prior knowledge about the change point probability.

As a consequence the prior probability distribution of the partitions $f(\rho)$ has to be re-evaluated:

$$\begin{aligned} f(\rho) &= \int_0^1 f(\rho|p) f(p) dp \\ &= \int_0^1 p^{b-1}(1-p)^{n-b} p^{\alpha-1}(1-p)^{\beta-1} \frac{1}{B(\alpha, \beta)} dp \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 p^{(\alpha+b-1)-1}(1-p)^{(\beta+n-b)-1} dp \end{aligned}$$

where the expression within the integral is the kernel of a beta distribution; therefore:

$$f(\rho) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+b-1)\Gamma(\beta+n-b)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n-1)} \tag{5.21}$$

Note that the posterior distribution of the change point probability can be derived as

$$\begin{aligned} f(p|\vec{X},\rho) &\propto f(\rho|p)f(p) \\ &\propto p^{(\alpha+b-1)-1}(1-p)^{(\beta+n-b)-1} \\ (p|\vec{X},\rho) &\sim Beta(\alpha+b-1,\beta+n-b) \end{aligned} \tag{5.22}$$

which is a beta distribution. This allows to calculate the expectation value of the change point probability given the partition as

$$\hat{p}(\rho) = E(p|\vec{X},\rho) = \frac{\alpha+b-1}{\alpha+\beta+n-1} \tag{5.23}$$

Combining Eq. (5.16) and Eq. (5.21) the posterior probability distribution of the partitions can be approximated as

$$f(\rho|\vec{X}) = \frac{f(\vec{X}|\rho)f(\rho)}{\sum_\rho f(\vec{X}|\rho)f(\rho)} \tag{5.24}$$

$$f(\vec{X}|\rho) = \prod_{ij\in\rho} f_{ij}(X_{ij}) \tag{5.25}$$

where the marginal likelihood for all observations $\vec{X}$ is in product form. The expectation values of the unknown parameters independent from the partitions can be calculated as

$$\hat{\theta}_k = E(\theta_k|\vec{X}) = \sum_\rho E_{ij}(\theta_k|X_{ij})f(\rho|\vec{X}) \tag{5.26}$$

where $1 \leq k \leq n$ and $\theta$ represents the unknown parameter $\mu$ or $\sigma^2$.

## 5.4 Implementation

Solving Eq. (5.24) will not be feasible from a numerical point of view. Therefore the partitions $P = \{\rho_1, \rho_2, \ldots, \rho_M\}$ are sampled using the MCMC approach as described in Section 3.1.3. For that the ratio as defined in Eq. (3.31) has to be calculated as

$$r = \frac{f(\rho|\vec{X},U_k=1)}{f(\rho|\vec{X},U_k=0)} = \frac{f(\vec{X}|\rho,U_k=1)f(\rho|U_k=1)}{f(\vec{X}|\rho,U_k=0)f(\rho|U_k=0)} \tag{5.27}$$

Using Eq. (5.16) and Eq. (5.21) the ratio can be expressed as

$$\begin{aligned} r &= \frac{\prod_{ij\in\rho} f_{ij}(X_{ij}|U_k=1) f(\rho|U_k=1)}{\prod_{ij\in\rho} f_{ij}(X_{ij}|U_k=0) f(\rho|U_k=0)} \\ &= \frac{f_{lk}(X_{lk}) f_{kr}(X_{kr})}{f_{lr}(X_{lr})} \frac{\Gamma(\alpha+b)\Gamma(\beta+n-b-1)}{\Gamma(\alpha+b-1)\Gamma(\beta+n-b)} \\ &= \frac{f_{lk}(X_{lk}) f_{kr}(X_{kr})}{f_{lr}(X_{lr})} \frac{\alpha+b-1}{\beta+n-b-1} \end{aligned} \tag{5.28}$$

where $l$ is the first change point left of the position $k$ and $r$ is the first change point right of the position $k$. In Eq. (5.28) it is sufficient to calculate the marginal likelihood from Eq. (5.16) as

$$f_{ij}(X_{ij}) \propto \left(\frac{1}{nv+1}\right)^{\frac{1}{2}} \left(\frac{1}{\chi}\right)^{\lambda/2} \tilde{b}^{\frac{1}{2}(\lambda-\frac{n}{2})} \frac{K_{\lambda-\frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)}{K_\lambda(\psi^{1/2}\chi^{1/2})} \tag{5.29}$$

$$\tilde{b} = \chi + \sum_{i=1}^{n} x_i^2 + \frac{\mu_0^2}{v} - \frac{(nv\overline{x}+\mu_0)^2}{v(nv+1)} \tag{5.30}$$

All other factors cancel out.

For a numerical implementation the main problem are the Bessel functions $K_\nu(x)$ in Eq. (5.29) and Eq. (5.12). The expression gets numerically unstable if either $x$ or $\nu$ get large. The problem for large $x$ can be solved through exponential scaling by rather calculating $e^x K_\nu(x)$ than $K_\nu(x)$. This method is implemented for `R` in the package **base** [R Core Team, 2015]. The problem for large $\nu$ can be solved by using an asymptotic expansion in Debye polynomials for large $\nu$. The method of Abramowitz and Stegun [1964] together with exponential scaling is implemented for `R` in the package **Bessel** [Maechler, 2015].

Using these methods will still not be sufficient to evaluate Eq. (5.28) and Eq. (5.12). Additionally the marginal likelihoods and therefore the Bessel functions have to be calculated in log-space. The ratio is evaluated as

$$r = \exp\left[\log(f_{lk}(X_{lk})) + \log(f_{kr}(X_{kr})) - \log(f_{lr}(X_{lr}))\right] \frac{\alpha+b-1}{\beta+n-b-1} \tag{5.31}$$

The logarithm of the marginal likelihood can be calculated as

$$\log(f_{ij}(X_{ij})) = -\frac{1}{2}\log(nv+1) - \frac{\lambda}{2}\log(\chi) + \frac{1}{2}(\lambda-\frac{n}{2})\log(\tilde{b})$$

$$+\log\left(K_{\lambda-\frac{n}{2}}\left(\psi^{1/2}\tilde{b}^{1/2}\right)\right)-\log\left(K_{\lambda}(\psi^{1/2}\chi^{1/2})\right) \tag{5.32}$$

The logarithm of the Bessel function using the exponential scaling can be calculated as

$$\log(K_{\nu}(x))=\log(e^{x}K_{\nu}(x))-x \tag{5.33}$$

For the evaluations of Eq. (5.32) the method was implemented by calculating the exponentially scaled logarithmic values of the Bessel functions. If $\nu > 30$ the exponentially scaled logarithmic values of the Bessel functions using the asymptotic expansion was used.

Using the ratio of Eq. (5.31) it is possible to generate $M$ random partition samples according to $f(\rho|\vec{X})$ as described in Section 3.1.3. For every sample the expectation values $\hat{\mu}_k(\rho)$ (see Eq. (5.8)) and $\hat{\sigma}_k^2(\rho)$ (see Eq. (5.12)) can be calculated. Using Eq. (3.15) an approximation of the expectation values $m_k$ independent from the partitions can be calculated as

$$m_k=\langle E(\theta_k|\vec{X})\rangle=\frac{1}{M}\sum_{\rho\in P}\hat{\theta_k}(\rho)=\overline{\hat{\boldsymbol{\theta}}_k} \tag{5.34}$$

where $\theta$ represents the unknown parameter $\mu$ or $\sigma^2$ and $\hat{\boldsymbol{\theta}}_k$ is the vector that holds the estimations $\hat{\theta_k}(\rho)$ of the $M$ partition samples. The variance within the estimations can be calculated following Eq. (3.16) as

$$s_k^2=\frac{n}{n-1}\left(\overline{\hat{\boldsymbol{\theta}}_k\cdot\hat{\boldsymbol{\theta}}_k}-\overline{\hat{\boldsymbol{\theta}}_k}\cdot\overline{\hat{\boldsymbol{\theta}}_k}\right) \tag{5.35}$$

Alternatively the conditional expectation values (Eq. (5.10) and Eq. (5.14)) could be used (Gibbs sampling). The posterior probabilities $p_k$ if the point $k+1$ is a change point can be calculated following Eq. (3.17) as

$$p_k=\frac{1}{M}\sum_{m=1}^{M}U_{k,m} \tag{5.36}$$

where $U_{k,m}$ is the k-th entry of the m-th sample partition $\rho_m$ (see also Eq. (3.12)).

# 6 Discussion

## 6.1 Comparison

The main difference of the model described in Chapter 4 (the N-NGIG model) to the model used by Loschi et al. [2003] (the N-NIG model) is that the prior for the variance is a GIG distribution (see Eq. (4.5)) instead of an inverse gamma distribution. The gamma (G) and the inverse gamma (IG) distributions are defined as

$$G(x|\eta,\kappa) = \frac{\kappa^{\eta}}{\Gamma(\eta)} x^{\eta-1} \exp\left(-\kappa x\right) \tag{6.1}$$

$$IG(x|\eta,\kappa) = \frac{\kappa^{\eta}}{\Gamma(\eta)} x^{-\eta-1} \exp\left(-\frac{\kappa}{x}\right) \tag{6.2}$$

These distributions are both special cases of the GIG distribution [Jorgensen, 1982] and related to each other as

$$G(\eta,\kappa) = GIG(\psi = 2\kappa, \chi = 0, \lambda = \eta) \tag{6.3}$$

$$IG(\eta,\kappa) = GIG(\psi = 0, \chi = 2\kappa, \lambda = -\eta) \tag{6.4}$$

using the definition of the GIG distribution in Eq. (4.14). Therefore the BCP N-NGIG model represents a generalisation of the BCP N-NIG model.

Note that choosing a gamma prior ($\chi = 0$) does not mean that the posterior distribution is a gamma distribution (see Eq. (4.27)). For the gamma prior the marginal likelihood becomes a symmetric variance gamma distribution (see Eq. (4.33)). The inverse gamma distribution can only be modelled approximately in the limit of $\psi \to 0$. The reason is that the marginal likelihood is only defined for $\tilde{\alpha} > 0$ (see Eq. (4.33)). In that limit the posterior distribution is only an inverse gamma distribution if $\tilde{\lambda} < 0$ (see Eq. (4.27)).

In the framework of the N-NIG model the variance can only be modelled through an IG distribution. While for the N-NGIG model various distributions are possible. Besides the gamma distribution and the approximated IG distribution these are for example

the (reciprocal) inverse Gaussian distribution or the hyperbolic distribution. Of course one can also choose parameter values where the GIG distribution can not be described through another distribution. Generally the GIG distribution allows for more flexibility in modelling the prior and the posterior distribution of the variance. More precisely this model allows for more combinations of the mean, the variance, the skewness and the kurtosis of the variance.

The sample variance is defined to be the mean of the squared mean deviations ($\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}(x_i - \overline{x})^2$). One could use the squared mean deviations ($(x_i - \overline{x})^2$) itself as a model for the variance. If the data is assumed to be normally distributed then so are the mean deviations ($X \sim N(0, \sigma^2)$). This would suggest that the squared mean deviations are distributed as

$$X^2 \sim \sigma^2 \chi_1^2 \sim G(\frac{1}{2}, \frac{1}{2}\sigma^{-2}) \sim GIG(\sigma^{-2}, 0, \frac{1}{2}) \tag{6.5}$$

where $\chi_k^2$ represents the chi-squared distribution. This implies that the variance is distributed according to a gamma distribution. Which raises the question why the inverse gamma distribution is used to model the variance. The reason is that the gamma distribution is not a conjugate prior for the normal assumption of the data. Only if the gamma distribution is considered to be a special case of the GIG distribution it is possible to use the gamma distribution as the prior distribution of the variance.

## 6.2 Hyperparameters

The hyperparameters $\mu_0$ and $v$ are used to model the prior knowledge about the mean. One might say that all the knowledge we have are our observations $\vec{X}$. And that the best guess for the mean of the data would be the mean of the observations.

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6.6}$$

What this means for the Bayesian model can be seen in Eq. (5.8). Choosing for $\mu_0$ the mean of the observations means that the expectation value of the unknown mean is exactly the mean of the observations. It is disputable on whether a Bayesian analysis is of much benefit under these circumstances. This is different if breakpoints are introduced. The assumption is that the different clusters have means that are distributed around $\mu_0$. Taking for $\mu_0$ the mean of all the samples might therefore be a reasonable choice since

inside the clusters the means will be different from $\mu_0$ and if there are no change points the mean will converge into the sample mean.

Eq. (5.8) shows another interesting property; the more observations are available the more the sample mean will dominate the outcome. For few data the assumptions about $\mu_0$ will dominate. Also it shows how $v$ influences the outcome. The smaller $v$ the more $\mu_0$ will dominate the outcome. Which means that if we are insecure about $\mu_0$ we would choose a flat prior ($v$ large). These are general principles about priors. If we make them flat then the data will dominate the outcome. If we have few data then the priors will dominate the outcome. This should be kept in mind when modelling our prior knowledge of the variance through $\psi$, $\chi$ and $\lambda$ or our prior knowledge of the change point probability through $\alpha$ and $\beta$.

The variance of the GIG distribution can generally not be modified without modifying the mean. This is different for the gamma distribution since

$$E[G(\eta, \kappa)] = \frac{\eta}{\kappa}, \quad Var[G(\eta, \kappa)] = \frac{\eta}{\kappa^2} \tag{6.7}$$

If the prior distribution of the variance is modelled by using the gamma distribution (see Eq. (6.5)) the parameters of the GIG distribution can be estimated as

$$\psi = \frac{1}{\hat{\sigma}^2} v_2^{-1}, \quad \chi = 0, \quad \lambda = \frac{1}{2} v_2^{-1} \tag{6.8}$$

where $v_2$ alters the variance of the GIG distribution but not the mean. For a flat prior $v_2$ would be chosen large. Note that this is just a suggestion for an automated way of setting the hyperparameters. Since for a large amount of time series it might become infeasible to set these parameters manually. Especially if those series are analysed on a sliding window where it might be preferable to set the parameters individually for each window. But of course the parameters can also be set manually.

Fig. 6.1 visualizes the likelihood as defined in Eq. (4.3). The multivariate likelihood is always evaluated for the same data $\vec{X}$ but with varying parameters $\mu$ and $\sigma^2$. This can be seen as a representation of the maximum likelihood estimation (MLE) that would find the unknown mean and variance at the maximum of the likelihood function.

Fig. 6.2 visualizes the prior distribution as defined in Eq. (4.4) and Eq. (4.5). Choosing $\mu_0 = 0.9$ and $v = 0.3$ is quite a narrow prior around the mean which means that we are confident in that mean. But still it allows for some deviation from that assumption. Choosing $\psi = 0.01$, $\chi = 0.1$ and $\lambda = 0.5$ is quite a flat prior with a mean of 103 and a

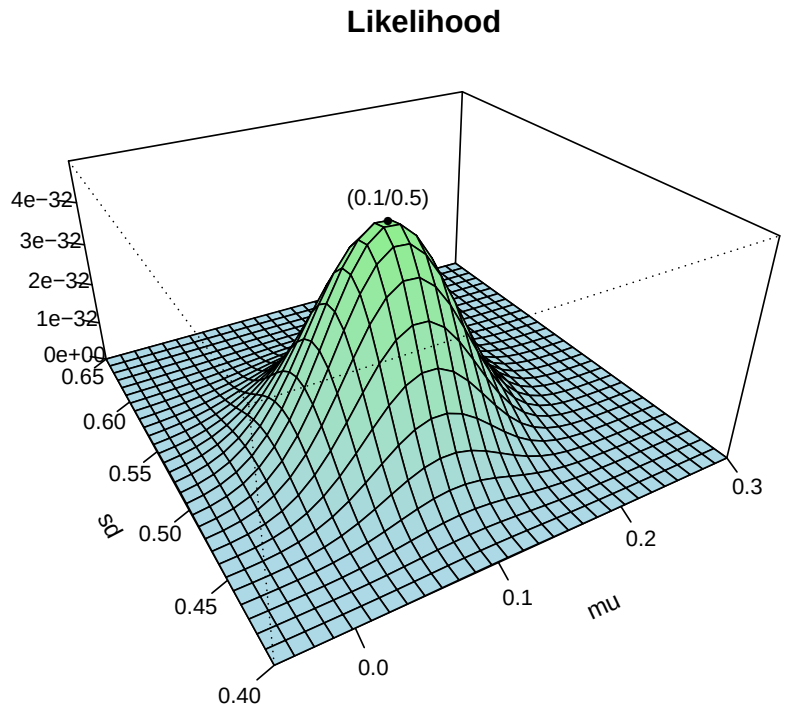


Figure 6.1: The plot shows on the z-axis the density of the multivariate normal distribution $f(\vec{X}|\mu, \sigma^2)$ for the $\mu's$ on the x-axis and the $\sigma^2$'s on the y-axis. The data $\vec{X}$ has 100 normally distributed samples with a mean of 0.1 and a standard deviation of 0.5.

variance of 20316. Which means that we have no idea what the variance could be. But note that at the same time we give variances very far away from 0.5 a chance.

Fig. 6.3 visualizes the posterior distribution as defined in Eq. (4.34). The prior for the mean influences the outcome such that the maximum of the posterior distribution is larger than 0.1. But still the data dominates the outcome. As already mentioned the prior for the variance is very flat. Which also means that variances very far away from 0.5 are taken into consideration which moves the variance away from 0.5.

If the priors would be fitted to the data; which means to define a vary narrow prior for the mean around 0.1 and a very narrow prior for the variance around 0.5. Then the posterior distribution would have its maximum almost exactly at the same location as the

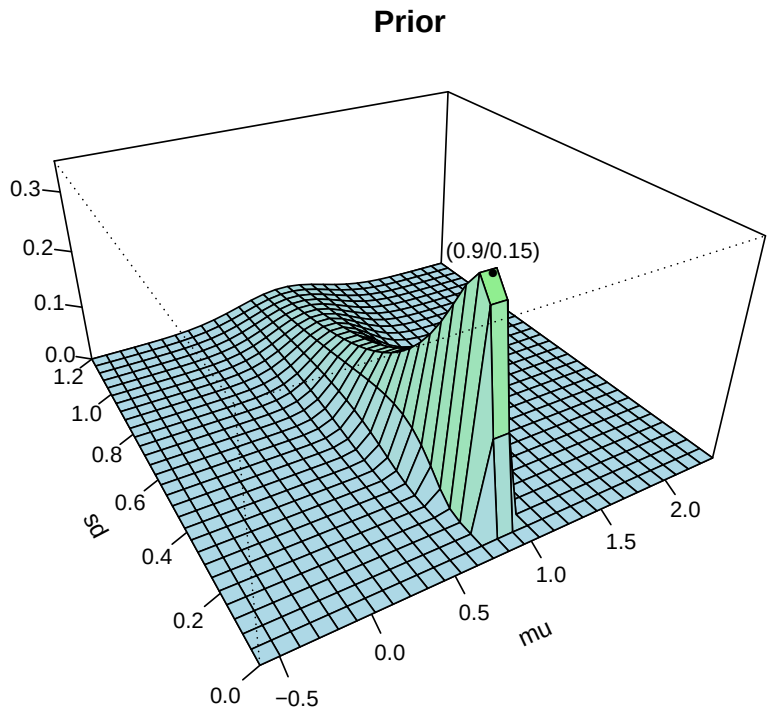


Figure 6.2: The plot shows on the z-axis the density of the prior distribution $f(\mu, \sigma^2) = f(\mu|\mu_0, v\sigma^2)f(\sigma^2|\psi, \chi, \lambda)$ for the $\mu$'s on the x-axis and the $\sigma^2$'s on the y-axis. The hyperparameters are: $\mu_0 = 0.9$, $v = 0.3$, $\psi = 0.01$, $\chi = 0.1$ and $\lambda = 0.5$.

likelihood. If we assume that there are no change points in the data then the posterior Bayesian estimations of the expectation values of the mean and the variance might not bring much benefit compared to a sample fit (although we would get more information about the mean and the variance than only the expectation values). On the other hand if there are change points it might make sense to generate the prior distributions around the estimations from the whole dataset by allowing enough deviation from these global estimations inside the clusters. This allows the expectation values of the mean and the variance inside the clusters to move around this global estimations.

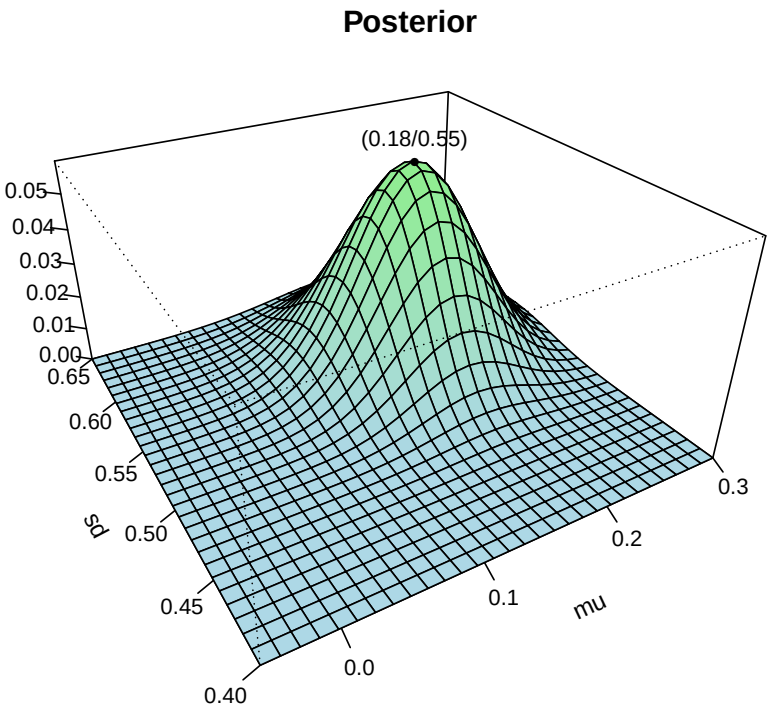


Figure 6.3: The plot shows on the z-axis the density of the posterior distribution $f(\mu, \sigma^2|\vec{X}, \mu_0, v, \psi, \chi, \lambda)$ for the $\mu$'s on the x-axis and the $\sigma^2$'s on the y-axis. The data $\vec{X}$ has 100 normal distributed samples with mean 0.1 and a standard deviation 0.5. The hyperparameters are: $\mu_0 = 0.9$, $v = 0.3$, $\psi = 0.01$, $\chi = 0.1$ and $\lambda = 0.5$.

## 6.3 Evaluation

It was already shown in Section 3.2.5 that the specific implementation of Barry and Hartigan [1992] was not designed to detect changes in the variance. In the following the same scenes as analysed in Fig. 3.1 and Fig. 3.2 are compared to the BCP N-NGIG model as defined in Chapter 5. For all analyses 300 partition samples were created (see Section 3.1.3) where the first 100 samples are discarded (burn-in samples). For the specific implementation of Barry and Hartigan the hyperparameters were always set to $p_0 = 0.2$ and $w_0 = 0.2$. For the BCP N-NGIG model the hyperparameters for the change points were set to $\alpha = 4$ and $\beta = 16$. This represents the case where the change point

probability is distributed around 0.2. The hyperparameter $\mu_0$ was set to be the mean of the data (see Eq. (6.6)) and $v$ was set to be 1. This represents the case were we believe that the means inside the clusters are distributed around $\mu_0$ and allow for some deviation through $v$. For the same reason the hyperparameters $\psi$, $\chi$ and $\lambda$ were calibrated on the data (see Eq. (6.8)) and $v_2$ set to 1. Note that based on the dataset one might have better ideas about the nature of the unknown parameters $\mu$ and $\sigma^2$. Then this can be modelled through the hyperparameters.

As already shown in Fig. 3.1 the specific implementation of Barry and Hartigan can detect changes for the case where only the mean changes but not the variance (Fig. 6.4). Fig. 6.5 shows the analysis of this data using the BCP N-NGIG model. The change points are detected reliably as well. Note that variances are very sensitive to outliers. The real change points are at $k = \{50, 150, 200\}$. Once in a while the Markov Chain will put the change point not exactly there which leads to an outlier within the estimation of the variances. For the mean this effect is less dominant. As already shown in Fig. 3.2 the specific implementation of Barry and Hartigan can get in trouble if the variance within the clusters change (Fig. 6.6). Fig. 6.7 shows the analysis of this data using the BCP N-NGIG model where the change points are detected reliably.

Fig. 6.8 shows the BCP analysis using the specific implementation of Barry and Hartigan for artificial normal returns that have the same mean and variance as the European equity index. Also the number of returns $n = 280$ is the same. Which means that all the returns are generated from the same dynamic. The specific implementation of Barry and Hartigan as well as the BCP N-NGIG model (see Fig. 6.9) can detect this reliably. Ultimately the idea is to use the BCP N-NGIG model to generate trading signals for financial time series (Part III). Fig. 6.10 shows the BCP analysis using the specific implementation of Barry and Hartigan for the returns of the European equity index and Fig. 6.11 shows the same analysis using the BCP N-NGIG model. Both analyses show that these financial returns can hardly be explained with a common dynamic as it is the case for the normal random returns. It is the times where the underlying dynamic of the returns are about to change that we would like to detect in order to adjust the estimation of the trend and the risk early. In phases of crisis (like around the year 2000) the specific implementation of Barry and Hartigan shows lots of structural breaks and therefore the estimation of the mean can jump. For the BCP N-NGIG model the estimation of the trend within these phases is smoother. Since the BCP N-NGIG model does not assume that all the clusters have to be explained with the same variance.

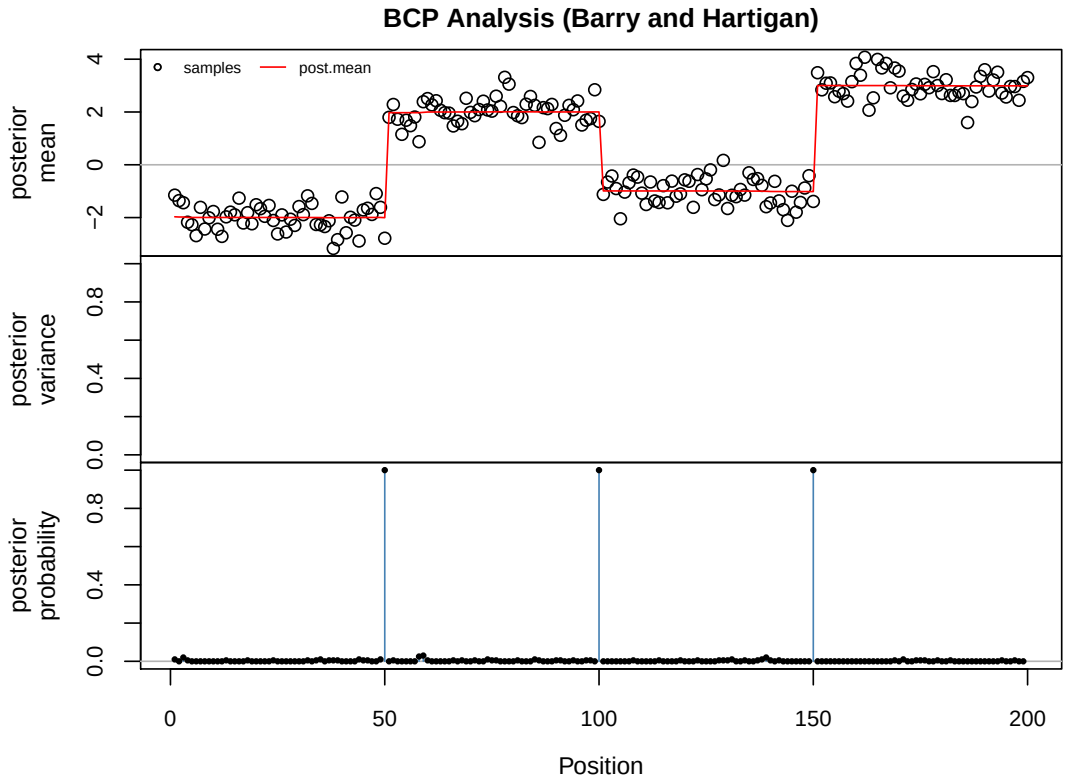


Figure 6.4: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2, the second one a mean of 2, the third one a mean of -1 and the fourth one a mean of 3. The standard deviation is constant and defined to be 0.5 for all clusters. Hyperparameters: $p_0 = 0.2$, $w_0 = 0.2$.

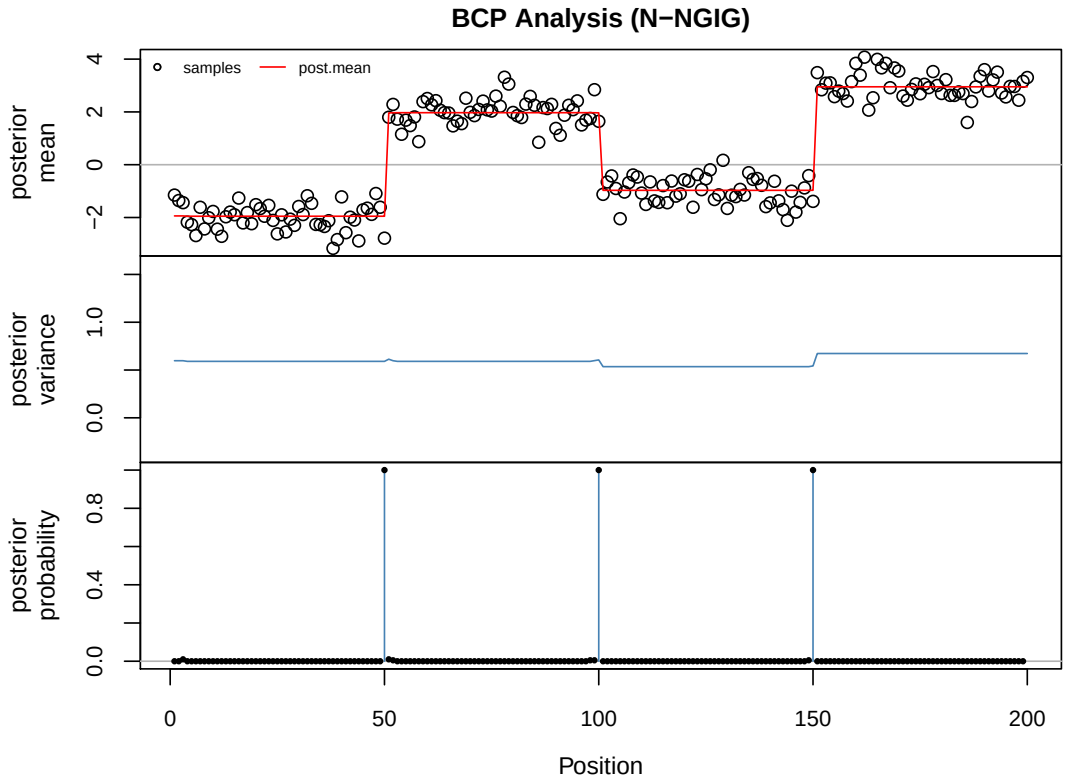


Figure 6.5: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2, the second one a mean of 2, the third one a mean of -1 and the fourth one a mean of 3. The standard deviation is constant and defined to be 0.5 for all clusters. Hyperparameters: $\mu_0 = 0.5$, $v = 1$, $\psi = 0.22$, $\chi = 0$, $\lambda = 0.5$, $\alpha = 4$, $\beta = 16$.

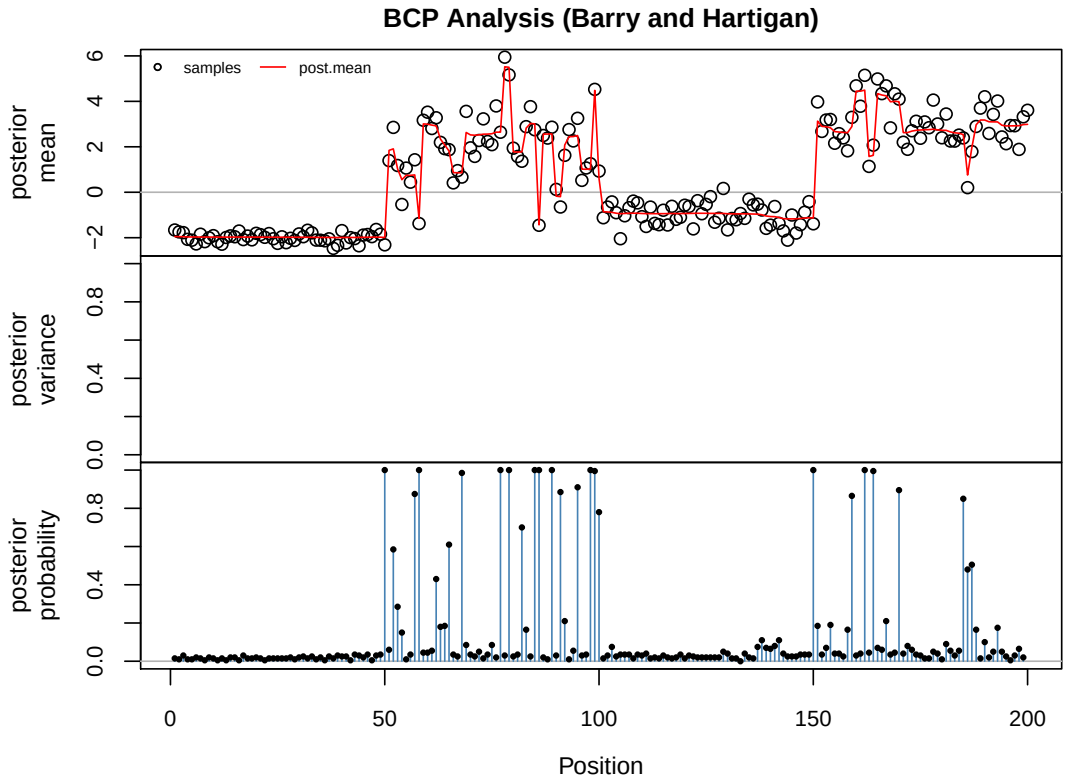


Figure 6.6: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2 and a standard deviation of 0.2, the second one a mean of 2 and a standard deviation of 1.5, the third one a mean of -1 and a standard deviation of 0.5 and the fourth one a mean of 3 and a standard deviation of 1. Hyperparameters: $p_0 = 0.2$, $w_0 = 0.2$.

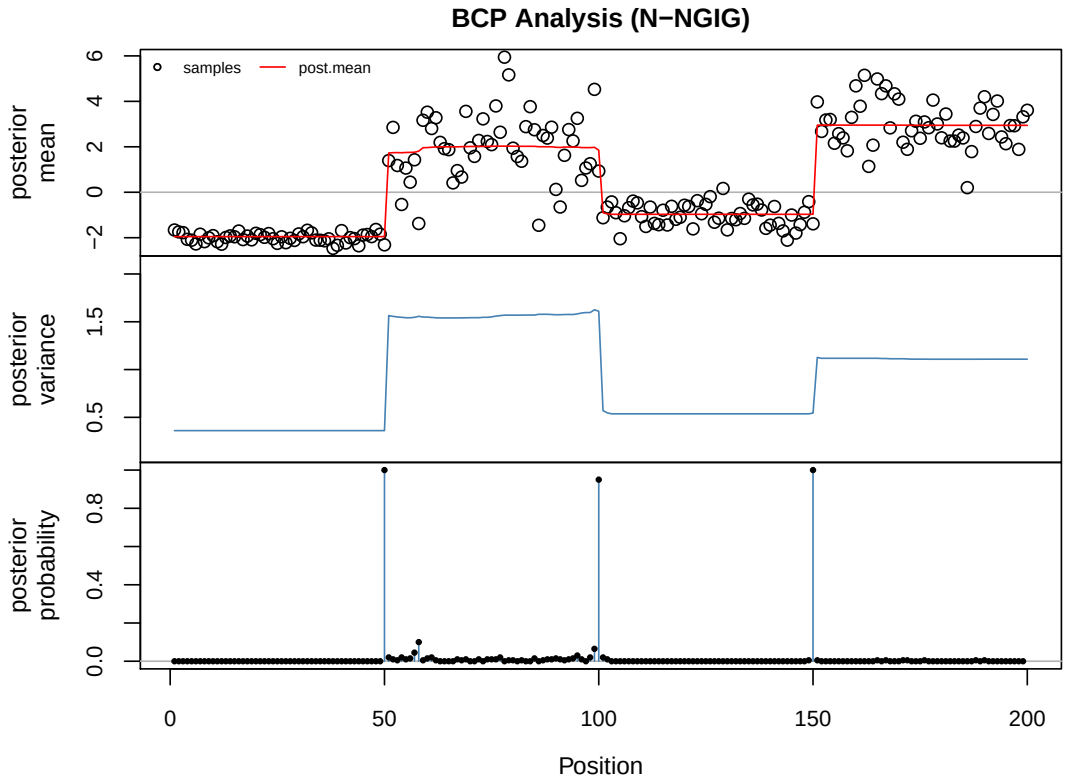


Figure 6.7: The dataset has four clusters of 50 normal random samples. The first one has a mean of -2 and a standard deviation of 0.2, the second one a mean of 2 and a standard deviation of 1.5, the third one a mean of -1 and a standard deviation of 0.5 and the fourth one a mean of 3 and a standard deviation of 1. Hyperparameters: $\mu_0 = 0.5$, $v = 1$, $\psi = 0.19$, $\chi = 0$, $\lambda = 0.5$, $\alpha = 4$, $\beta = 16$.

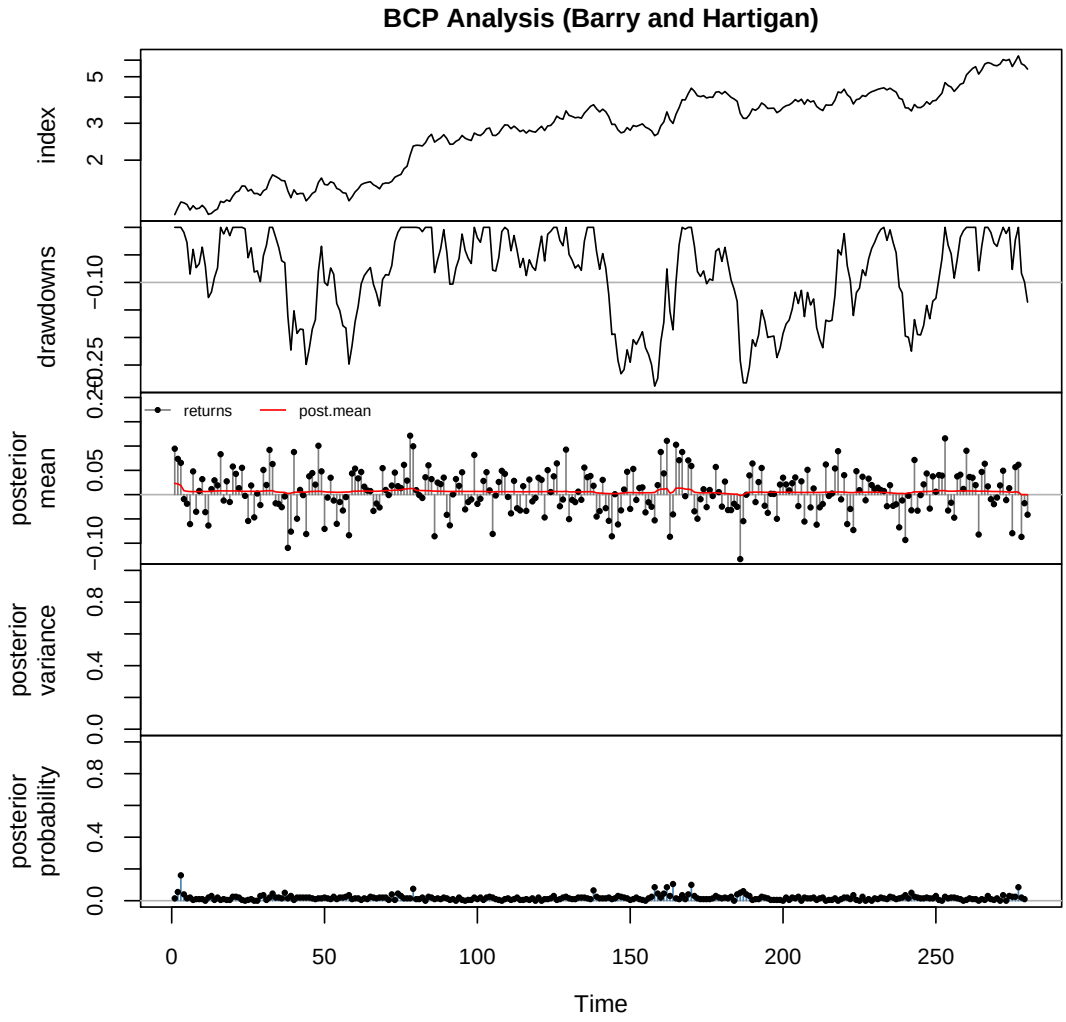


Figure 6.8: The artificial index has normally distributed returns that have the same mean (0.6%) and standard deviation (4.5%) as the European equity index (EQEU.EUR). Hyperparameters: $p_0 = 0.2$, $w_0 = 0.2$. See also Appendix A for more information about the data.

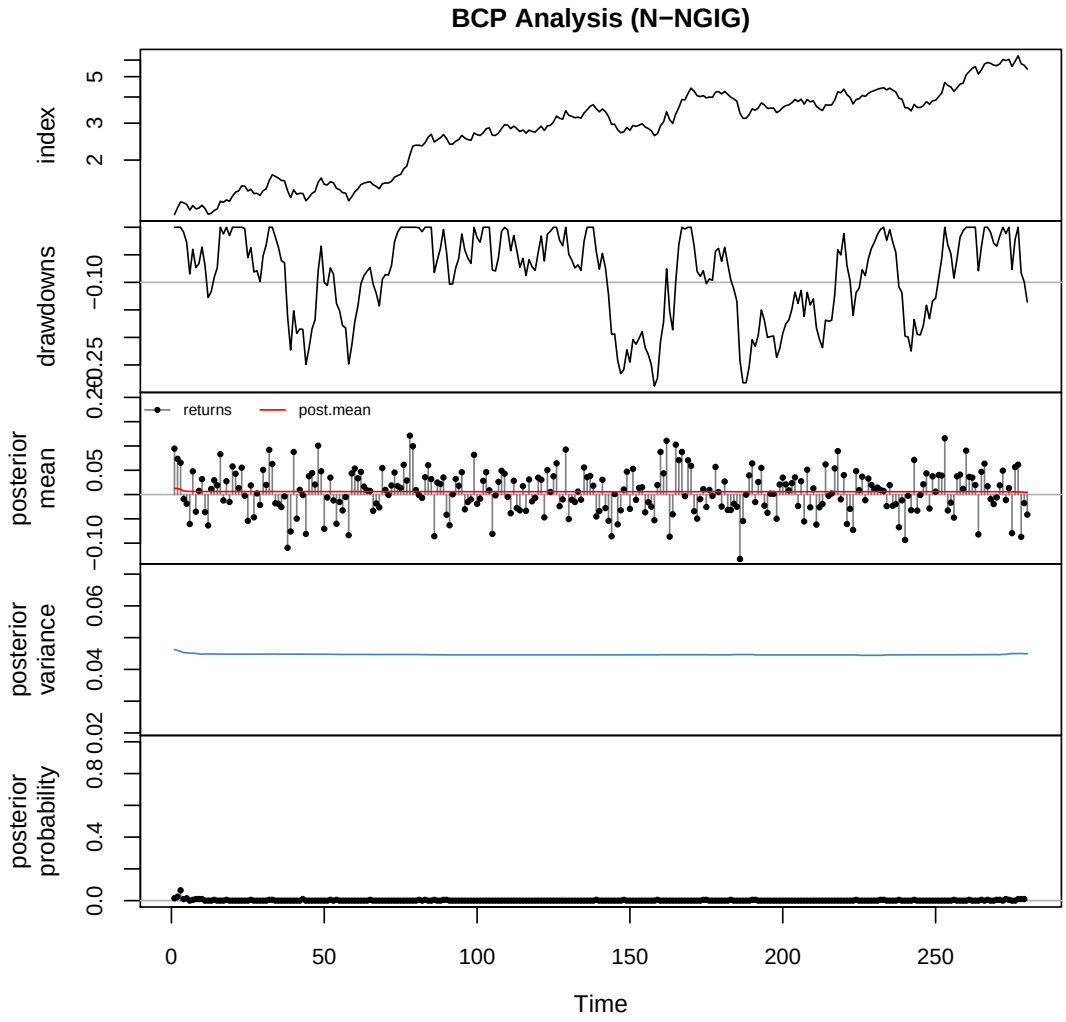


Figure 6.9: The artificial index has normally distributed returns that have the same mean (0.6%) and standard deviation (4.5%) as the European equity index (EQEU.EUR). Hyperparameters: $\mu_0 = 0.01$, $\upsilon = 1$, $\psi = 502.36$, $\chi = 0$, $\lambda = 0.5$, $\alpha = 4$, $\beta = 16$. See also Appendix A for more information about the data.

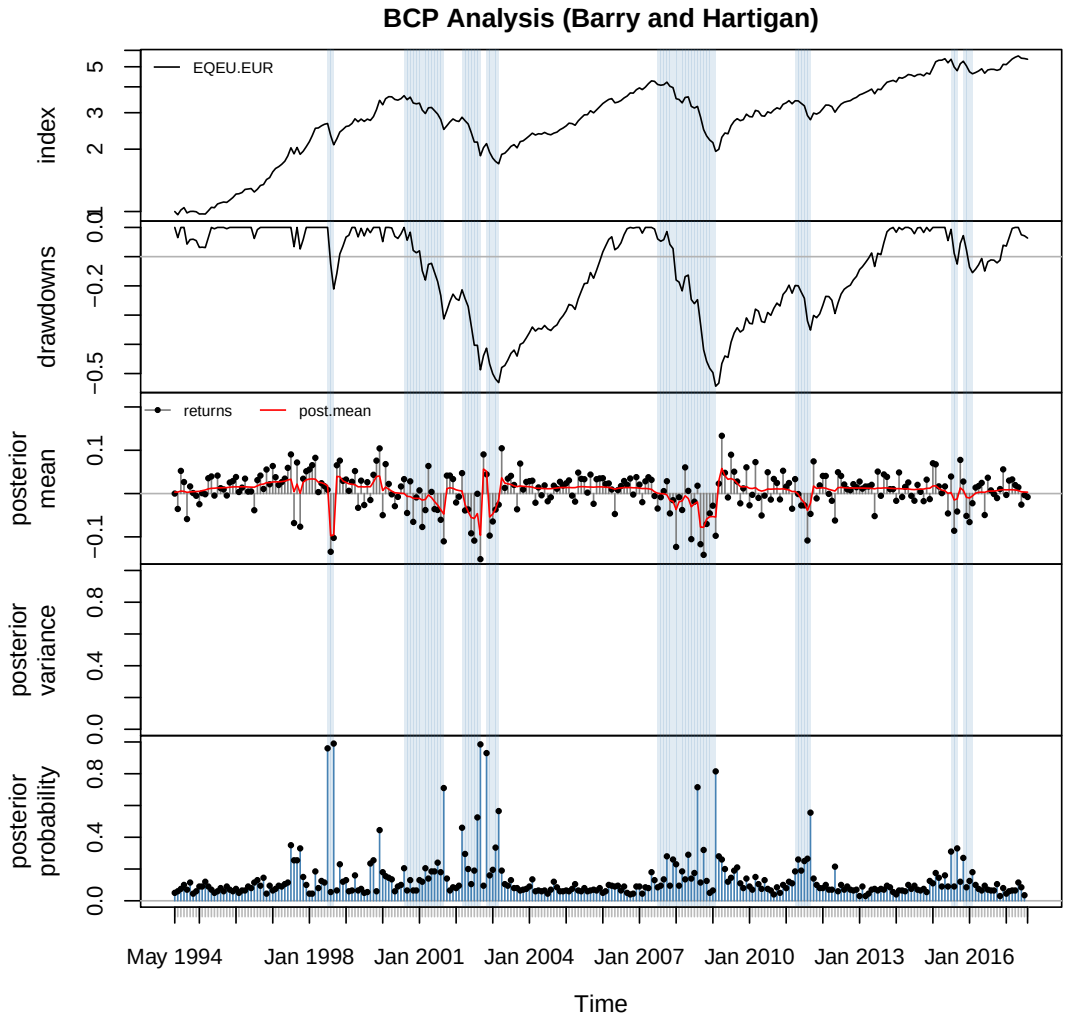


Figure 6.10: The European equity index. The shaded areas indicate times where the posterior mean is negative. Hyperparameters: $p_0 = 0.2$, $w_0 = 0.2$. See also Appendix A for more information about the data.

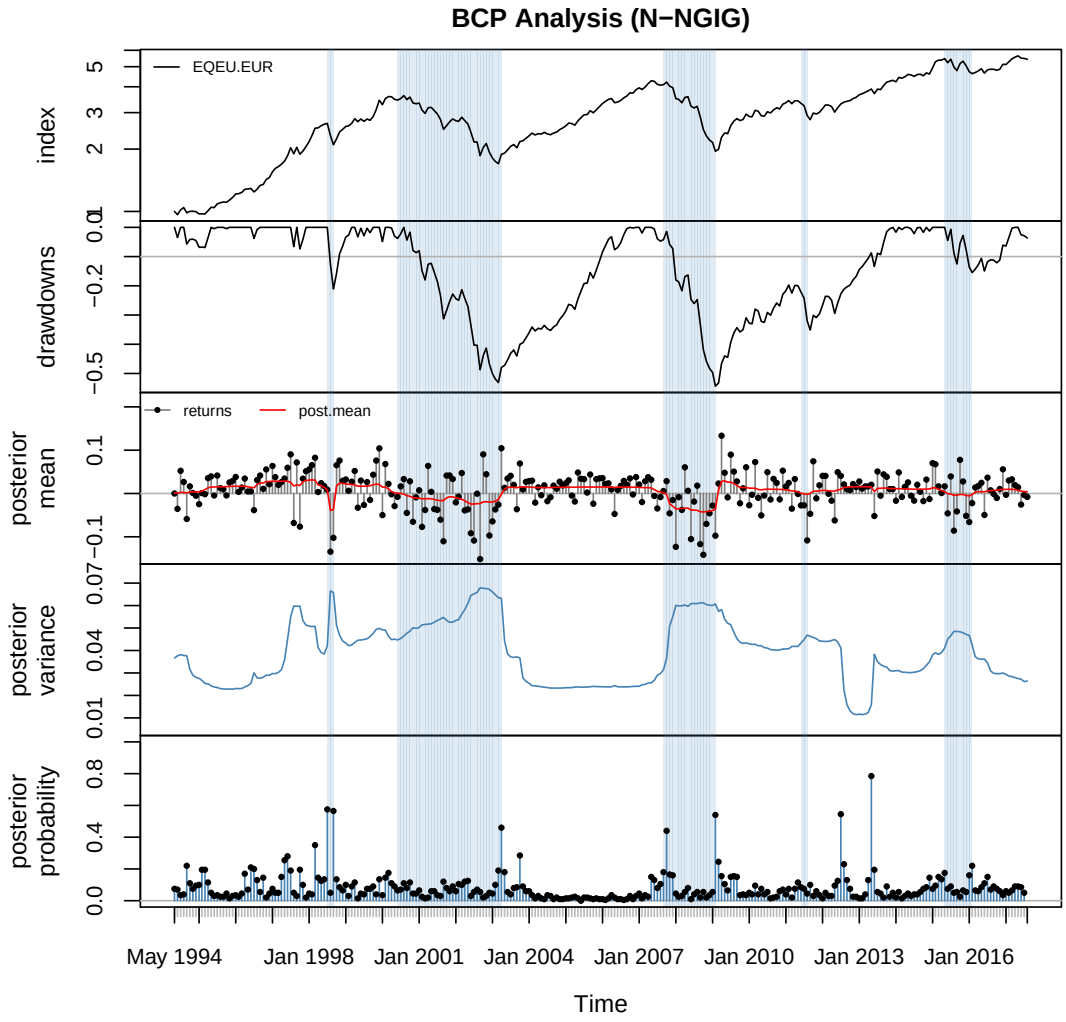


Figure 6.11: The European equity index. The shaded areas indicate times where the posterior mean is negative. Hyperparameters: $\mu_0 = 0.01$, $v = 1$, $\psi = 502.36$, $\chi = 0$, $\lambda = 0.5$, $\alpha = 4$, $\beta = 16$. See also Appendix A for more information about the data.

# 7 Summary and Outlook

The posterior marginal and conditional distributions as well as the marginal likelihood were derived and it was shown that they all belong to the family of generalized hyperbolic (GH) and generalized inverse Gaussian (GIG) distributions. Therefore the mean and the variance for the N-NGIG model have proper conjugate priors. It was shown that the BCP N-NGIG model of this thesis represents a generalisation of the BCP N-NIG model as introduced by Loschi et al. [2003]. Since the family of GH and GIG distributions is well documented in literature and since various software implementations are available it is convenient to work with these kind of distributions. The calculation of moments, fitting of data and random sampling for these distributions is readily available. The implementation of the BCP N-NGIG model can reliably detect change points for scenarios where both the mean and the variance change. Analyses of real world financial returns lead therefore to smother results for the estimation of the trend.

For the BCP model it was sufficient to derive the posterior marginal distributions and the posterior likelihood. For a complete Bayesian analysis it would be of interest to also find closed forms of the posterior distribution and the predictive distribution. While those distributions have been derived by Thabane and Safiul Haq [1999] in the context of generalized modified Bessel distributions it might be possible to define these distributions in the context of GH distributions. Also the Bayesian refinements of the new model (see Section 5.3) could be extended to more hyperparameters. A further iteration of the model could be to change the normal assumption of the data. A possible candidate could be the normal inverse Gaussian (NIG) distribution which has four parameters that can also explain the skewness and the kurtosis of the data. Although one has to consider that defining prior distributions for all four parameters would result in quite a challenging setup with many more hyperparameters. Additionally the model could be set up for the multivariate case in order to calculate the covariances. It would also be of interest to answer the question on how robust the new method is when it is faced with many different scenarios where the mean and/or the variance change for various parts of the data.

# Part II

# Morphological Shape Factors of the Feasible Set

# 8 Introduction

The first part of this thesis discussed risk measures for the univariate case. This part of the thesis deals with a new kind of multivariate risk measures. The first step of an investment process is usually to define the investment universe. After that decisions have to be made on how the available capital should be allocated to the different assets within the investment universe. Throughout this part of the thesis it is always assumed that the investor can only buy assets (no short-selling) with no leverage.

There are various approaches on how the available capital could be allocated. There are for example discretionary approaches where a well informed investment committee makes those decisions. In this thesis the focus lies on quantitative approaches. Investment decisions could for example be based on the univariate risk measures presented in Part I and shown in Part III. A very popular multivariate alternative is the portfolio optimization approach as presented by Markowitz [1952].

Markowitz portfolio optimization characterizes the investment universe through the individual performance of the assets within the investment universe and the covariance matrix of those underlying assets. Depending on how the available capital is allocated to the different assets various portfolios can be designed that have itself their very own performance and variance (risk). Reversing this approach one can design portfolios that have a desired performance and variance. Visualizing all possible performance and variance combinations results in an image of the feasible set.

The feasible set is basically a function of the performance and the covariance matrix of the underlying assets. There are various approaches on how to calculate those statistics. Every approach will lead to a different image of the feasible set. In Chapter 9 the calculation of the feasible set by using different estimators is discussed. Besides the sample and a robust estimator the BCP method as described in Chapter 5 is used to calculate the performance and the risk of the underlying assets.

By describing the feasible set as an image it is possible to calculate the moments of that image. The representation of the feasible set as an image will be discussed in Chapter 10. Characterizing images through their moments is a concept that is widely

used in image processing and computer vision. In Chapter 10 the image moments as published by Hu [1962] are presented. Those moments are independent of location, scale and/or orientation. As such they could be used to identify unique shapes of the feasible set.

In portfolio optimization it is crucial to have a good idea of the shape of the feasible set. This makes the process of finding the right portfolio more convenient. Combinations of the raw and central image moments can be associated with a geometric meaning. This makes it possible to design the geometric shape factors (GSF) as discussed in Chapter 10. More precisely the GSF define an ellipse that has the same area, centre, orientation and eccentricity as the feasible set. The orientation and eccentricity can be found through a Principal Component Analysis (PCA). Either by using the covariance matrix of the image of the feasible set or the covariance matrix of the hull of the feasible set.

While this allows to get an idea of the shape of the feasible set at first glance the GSF do also have a statistical meaning. The area is a measure for the optimization benefit. Which means that if the area is large then the space of possible portfolios is large as well. The other measures generally describe the risk premium. The centre ratio is a measure for the expected performance per unit of risk. A high centre ratio is preferable. The orientation describes the direction of the risk premium. If the direction is positive taking more risk would generally lead to more performance. The eccentricity describes the correlation between risk and performance. The higher the correlation the higher the probability that taking more risk does indeed lead to more or less performance (depending on the direction). Therefore the GSF offer a convenient insight into the shape of the feasible set and some central statistical properties of the investment universe.

In Chapter 11 the Hu moments and GSF are calculated for selected investment universes. The sample, robust and BCP estimations are compared. In Chapter 12 the results are summarized. Note that all analyses within this thesis were made by using monthly data.

# 9 The Feasible Set

The feasible region is the solution space of an optimization problem that satisfies the constraints of the problem. Portfolio optimization as initially presented by Markowitz [1952] describes a given investment universe through the performance and the covariance matrix of the different assets within the investment universe. In all generality the basic problem can be defined as follows:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & f(\boldsymbol{w}) \\ & \text{s.t.} \\ & g(\boldsymbol{w}) = 0 \\ & l_h \leq h(\boldsymbol{w}) \leq u_h \\ & l_w \leq \boldsymbol{w} \leq u_w \end{aligned} \tag{9.1}$$

where $\boldsymbol{w} = \{w_1, w_2, \ldots, w_n\}$ describes the investment weights of the underlying assets. If for example $w_2 = 0.1$ it means that 10% of the available capital is invested in asset number two. $g(\boldsymbol{w}) = 0$ expresses the (non-)linear equality constraints, $l[h] \leq h(\boldsymbol{w}) \leq u[h]$ the (non-)linear inequality constraints and $l[w] \leq \boldsymbol{w} \leq u[w]$ the upper and lower bounds of the weights. The objective function $f(\boldsymbol{w})$ is often chosen as one of the following equations

$$f(\boldsymbol{w}) = -\boldsymbol{\mu}'\boldsymbol{w} \tag{9.2}$$

$$f(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w} \tag{9.3}$$

where Eq. (9.2) describes the maximization of the profit and Eq. (9.3) the minimization of the variance (risk). $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_n\}$ is the vector that holds the expected performances of the underlying assets and $\Sigma$ the covariance matrix of those assets. If for example Eq. (9.3) is used to solve the problem of Eq. (9.1) the result will be the weights of the portfolio that has the lowest variance given the constraints. This portfolio is also

called the minimum variance portfolio (MVP). Another prominent portfolio is the equal weights portfolio (EWP). The global EWP is defined as

$$\boldsymbol{w} = \left\{ w_1 = \frac{1}{n}, w_2 = \frac{1}{n}, \ldots, w_n = \frac{1}{n} \right\} \tag{9.4}$$

which is actually the portfolio that has the highest diversification. Under the presence of constraints the maximum diversification portfolio can be found by defining the objective function as

$$f(\boldsymbol{w}) = \mathrm{var}(\boldsymbol{w}) \tag{9.5}$$

The performances $\boldsymbol{\mu}$ and the covariance matrix $\Sigma$ of the underlying assets can be estimated from the returns $R = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n\}$ of the assets. The returns of the assets over time $\boldsymbol{r}_k = \{r_1, r_2, \ldots, r_s\}$ are calculated from the index values (prices) of the assets over time $\boldsymbol{p}_k = \{p_1, p_2, \ldots, p_s\}$ as

$$r_{t,k} = \ln \frac{p_{t,k}}{p_{t-1,k}} = \ln(p_{t,k}) - \ln(p_{t-1,k}) \tag{9.6}$$

which is the logarithmic return for the asset $k$ at time $t$. An alternative would be to use discrete returns. In contrary to the logarithmic returns the discrete returns are not symmetric (skewed) or time additive (non-normal). This can lead to undesired artefacts when estimating parameters or solving optimization problems. Therefore all analyses within this thesis are made by using logarithmic returns.

Often the sample estimators of the normal distribution are used to calculate $\boldsymbol{\mu}$ and $\Sigma$. Those estimators do not reflect many of the stylized facts encountered when dealing with financial returns [Cont, 2001]. Alternative approaches to calculate $\boldsymbol{\mu}$ and $\Sigma$ include the use of other elliptical distributions that can model fat tails or the use of robust estimators. To model even more stylized facts more elaborate objective functions have to be defined. In contrary to the quadratic objective functions of Eq. (9.3) this will usually lead to more complicated non-linear objective functions. For such non-linear problems it is crucial to use solid optimizers in order to find the global minimum.

Using alternative approaches to calculate $\boldsymbol{\mu}$ and $\Sigma$ or even completely changing the objective function will result in different images of the feasible set. The calculation of the feasible set is explained by using the unbiased sample estimators for $\boldsymbol{\mu}$ and $\Sigma$ in Section 9.1. To illustrate the difference between sample and alternative calculations of the

feasible set two alternative approaches will be discussed. In Section 9.2 the feasible set is calculated by using a robust estimator for $\boldsymbol{\mu}$ and $\Sigma$. In Section 9.3 the BCP estimates (as defined in Chapter 5) of the underlying assets are used to estimate $\boldsymbol{\mu}$ and $\Sigma$.

## 9.1 Sample Estimation

To calculate the hull of the feasible set the quadratic optimization problem as defined in Eq. (9.7) is solved repeatedly for different target returns $\bar{\boldsymbol{r}} = \{\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_m\}$.

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \pm \boldsymbol{w}' \Sigma \boldsymbol{w} \\ & \text{s.t.} \\ & \mathbf{1}' \boldsymbol{w} = 1 \\ & \boldsymbol{\mu}' \boldsymbol{w} = \bar{r}_i \\ & 0 \leq w_i \leq 1 \end{aligned} \tag{9.7}$$

The sequence of equally spaced target returns $\bar{\boldsymbol{r}}$ goes from the minimum of the expected returns $\bar{r}_1 = \min(\boldsymbol{\mu})$ to the maximum of the expected returns $\bar{r}_m = \max(\boldsymbol{\mu})$. The more target returns ($m$) that are chosen between the minimum and the maximum the higher the resolution of the hull. For every target return the problem of Eq. (9.7) is solved for $+\boldsymbol{w}'\Sigma\boldsymbol{w}$ and $-\boldsymbol{w}'\Sigma\boldsymbol{w}$.

Note that the weights are restricted such that they have to be between 0 and 1 (long-only) and sum up to one (fully-invested). This represents a natural and straightforward investment style. Many investors do or can not deviate from these restrictions. Since this problem cannot be solved analytically anymore for more than two assets it must be solved numerically. For `R` the problem of Eq. (9.7) can be solved using the **quadprog** package [Turlach and Weingessel, 2013] which can solve optimization problems for linear or quadratic objective functions and linear constraints. For more general problems where the objective functions and the constraints are (non-)linear the **Rsolnp** package [Ghalanos and Theussl, 2015] can be used. Various industrial solvers can be accessed through the AMPL interface.

The solution will be a portfolio defined through the weights ($\boldsymbol{w}$) that minimize the objective function while respecting the constraints. The portfolio will have a return of $\boldsymbol{\mu}'\boldsymbol{w} = \bar{r}_i$ and a variance of $\boldsymbol{w}'\Sigma\boldsymbol{w}$ where the variance is the lowest or highest possible variance for that return. Every portfolio can be visualized through its expected return

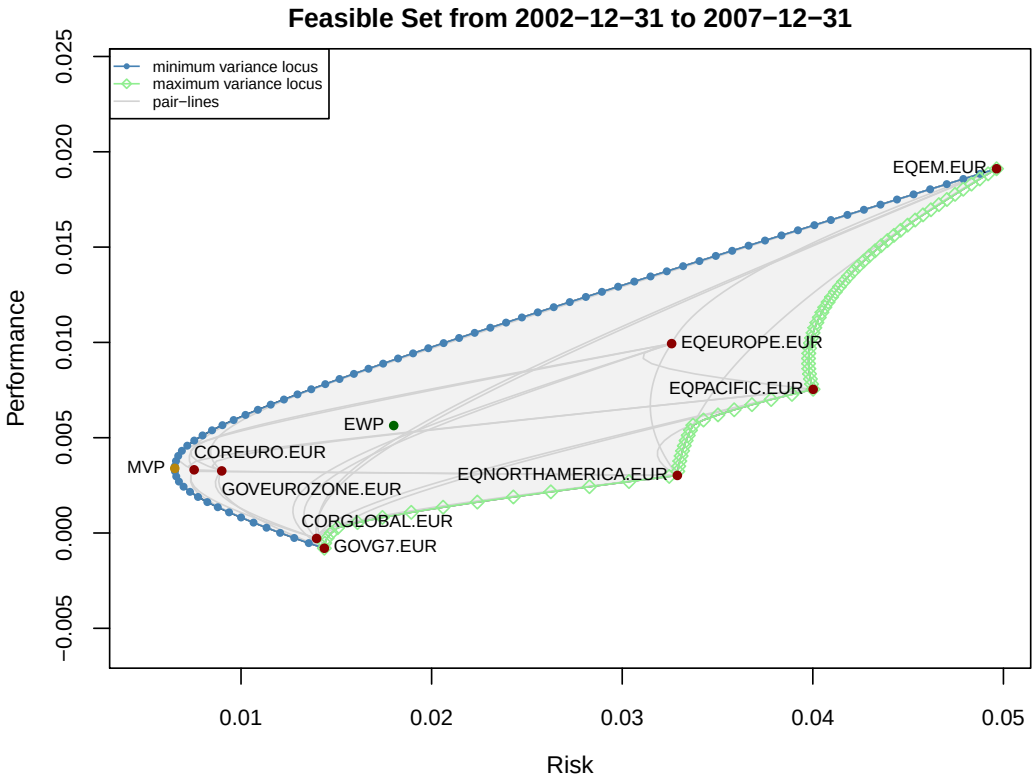


Figure 9.1: The feasible set of the global equity and bond universe using sample estimations. See also Appendix A for more information about the data.

(performance) and variance (risk) within the risk-performance-plane which leads to the hull of the feasible set (see Fig. 9.1). Randomly choosing weights between zero and one such that they sum up to one will always lead to a portfolio that lies inside the feasible set. Note that the axes of Fig. 9.1 are chosen such that they have an aspect ratio of 1. This ensures that a circle is actually shown as a circle and not as an ellipse.

In Fig. 9.1 the sample estimator was used to calculate the expected returns $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_n\}$ and the covariance matrix $\Sigma$ of the underlying assets. The sample mean and covariances can be calculated as

$$\mu_k = \hat{\mu}(\boldsymbol{r}_k) = \frac{1}{s}\sum_{t=1}^{s} r_{t,k} \tag{9.8}$$

$$\sigma_{kl} = \widehat{\text{cov}}(\boldsymbol{r}_k, \boldsymbol{r}_l) = \frac{1}{s-1}\sum_{t=1}^{s}(r_{t,k} - \mu_k)(r_{t,l} - \mu_l) \tag{9.9}$$

where $r_{t,k}$ is the logarithmic return for asst $k$ at time $t$.

The left part of the hull that goes from the asset with the lowest expected return to the asset with the highest expected return is called the minimum variance locus. Vice versa the right part of the hull is called the maximum variance locus. The part of the minimum variance locus that goes from the MVP to the asset with the highest expected return is called the efficient frontier. If only variance and expected return matter any portfolio that does not live on the efficient frontier is not optimal and should therefore not be chosen. If other statistics as for example the diversification are considered then this might be different.

The maximum variance locus can be found by minimizing $-\boldsymbol{w}'\Sigma\boldsymbol{w}$. But since that part of the hull is concave many optimizers struggle to find the minimum. On the other hand the maximum variance locus does always consist of only two assets and can therefore be calculated analytically. For that the problem as defined in Eq. (9.7) has to be solved for two assets. This can be done by solving the equation system of the constraints as

$$\begin{aligned} w_1 &= \frac{\bar{r}_i - \mu_2}{\mu_1 - \mu_2} \\ w_2 &= 1 - w_1 \\ \sigma_p &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ r_p &= \bar{r}_i \end{aligned} \tag{9.10}$$

where $r_p$ is the return and $\sigma_p$ the variance of the pairwise portfolio that has weights $\boldsymbol{w} = \{w_1, w_2\}$. After calculating the efficient frontiers of all pairwise portfolios (pair-lines) the maximum variance locus consists of the pairwise portfolios that have the highest variance given the target return $\bar{r}_i$. For all figures within this thesis that show the maximum variance locus this was the method that was used.

The underlying assets of the investment universe are a mix of four equity and four bond indices (in euro). The equity indices cover the global equity market. The government bond indices cover the bond market of the G7 and the eurozone countries while the corporate bond indices cover the global corporate bond market. The time frame is chosen such that the shape of the feasible set is somewhat expected. Which means that there is

a risk premium from the bond indices (low variance) to the equity indices (high variance). The question arises on whether this is actually a reasonable assumption at all times. The remainder of this part of the thesis will focus on that question. In Chapter 15 a portfolio that positions itself within the feasible set based on these findings will be presented. All illustrations to visualize the concepts of the morphological shape factors are made with the investment universe and time frame as in Fig. 9.1. Other universes will be discussed in Chapter 11.

## 9.2 Robust Estimation

### 9.2.1 Motivation

A problem that arises when using the sample estimators is that these estimates are sensible to outliers. Robust estimations of the mean (location) and the covariance matrix (dispersion or scale) are designed to address this problem. Ultimately this will lead to different images of the feasible set. A first approach to robust estimation would be to winsorize (equalizing outliers to a fixed quantile threshold) or trim (removing outliers) the data or to use the median as an alternative for the mean and the interquartile range (IQR) as an alternative for the standard deviation.

A measure for the degree of robustness for an estimator is the breakdown point (BP) [Donoho, 1982; Donoho and J., 1983] which can take values between 0 and 0.5. It describes the extend to which outliers will affect the estimation of a parameter. The mean for example has a BP of 0 since one new sample can make the estimation arbitrarily high. While the median has a breakdown point of 0.5 since the estimation stays robust if not more than half of the samples are outliers. For the multivariate case the situation is more complex. See also Maronna and Yohai [2016] and Pfaff [2013] for more details of the following summary.

The M estimator was introduced by Huber [1964] and Maronna [1976] where the unknown multivariate parameters $\boldsymbol{\mu}$ and $\Sigma$ are the solution of a minimization problem which is defined as

$$\min_{\boldsymbol{\mu},\Sigma} \quad \sum_t \rho(\boldsymbol{r}_t, \boldsymbol{\mu}, \Sigma) \tag{9.11}$$

where $\rho$ is an arbitrary function that can be defined such that extreme data points receive less weight and $\boldsymbol{r}_t$ is the row vector of the matrix $R$ of returns that has $s$ observations for $n$ assets. Special cases of this formulation are the maximum likelihood estimator (MLE) and the method of least squares (LS). Note that for these two cases outliers do

not receive less weight. The maximum BP for M estimators is $1/n$.

To overcome the maximum BP limitation various estimators have been introduced that minimize a robust scale of the Mahalanobis distances $d_t^2 = d(\boldsymbol{r}_t, \boldsymbol{\mu}, \Sigma)^2 = (\boldsymbol{r}_t - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{r}_t - \boldsymbol{\mu})$ in some way or another. Davies [1987] introduces a general formulation as

$$\min_{\boldsymbol{\mu},\Sigma} \quad g = g(d(\boldsymbol{r}_1, \boldsymbol{\mu}, \Sigma), \ldots, d(\boldsymbol{r}_s, \boldsymbol{\mu}, \Sigma)) \quad \text{s.t.} \quad |\Sigma| = 1 \tag{9.12}$$

where $g$ is a robust metric and $|\cdot|$ the determinant of a matrix. If $g$ is the mean of the values the solution for $\boldsymbol{\mu}$ and $\Sigma$ are the sample mean and a scalar multiply of the sample covariance matrix. If $g$ is the median the solution is the minimum volume ellipsoid (MVE) estimator which estimates the parameters using the points within the smallest ellipsoid that covers at least half of the points of $R$. If $g$ is the mean of the smaller half of the values the solution is the minimum covariance determinant (MCD) estimator which estimates the parameters using the points within the ellipsoid that has the smallest determinant of all ellipsoids that contain at least half of the points of $R$. The MVE and MCD estimators were already introduced earlier by Rousseeuw [1985] and Rousseeuw and Leroy [1987] where it was shown that they have an asymptotic BP of $\alpha$ where $\alpha$ can be chosen between 0 and 0.5.

The S estimator is defined for $g$ being an M estimate of scale [Huber, 1981; Hampel et al., 1986]. This means that $g$ satisfies $s^{-1}\sum_{t=1}^{s}\rho(d_t/g) = b = E_\phi(\rho(r/k))$ where $d_t = d(\boldsymbol{r}_t, \boldsymbol{\mu}, \Sigma)$, $\rho$ is bounded and non-decreasing, $\phi$ is the standard normal distribution, $r$ the sample variable, and $k$ the consistency factor for normal samples. Since $\rho$ is bounded the BP of S estimates can be equal to 0.5.

The MM estimates were introduced by Yohai [1987] in the context of regression analysis and adapted for the multivariate case by Lopuhaa [1992] and Tatsuoka and Tyler [2000]. First the unknown parameters are estimated $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ through a robust estimation (e.g. MCD). Let $S(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ be the M estimate of scale which can be found by solving $s^{-1}\sum_{t=1}^{s}\rho(\hat{d}(\boldsymbol{r}_t, \hat{\boldsymbol{\mu}}, \hat{\Sigma})/S(\hat{\boldsymbol{\mu}}, \hat{\Sigma})) = b$. The final parameters $(\boldsymbol{\mu}, \Sigma)$ are then estimated by minimizing $\sum_{t=1}^{s}\rho(d(\boldsymbol{r}_t, \boldsymbol{\mu}, \Sigma)/(cS(\hat{\boldsymbol{\mu}}, \hat{\Sigma})))$ such that $|\Sigma| = 1$ and where $c$ is a control parameter of efficiency that depends on $\rho$. The BP of the MM estimator is 0.5.

The $\tau$ estimates were introduced by Yohai and Zamar [1988] in the context of regression analysis and adapted for the multivariate case by Lopuhaa [1992]. The concept is very much related to the S estimator. But instead of minimizing the M estimate of scale a robust scale (the $\tau$-scale) is minimized that allows to control the efficiency of the estimation. Let $S(\boldsymbol{\mu}, \Sigma)$ be the M estimate of scale which can be found by solving

$s^{-1}\sum_{t=1}^{s}\rho_1(d(\boldsymbol{r}_t,\boldsymbol{\mu},\Sigma)/S(\boldsymbol{\mu},\Sigma))=b$. The $\tau$-scale is then defined as

$$\tau^2(\boldsymbol{\mu},\Sigma)=S^2(\boldsymbol{\mu},\Sigma)\frac{1}{n}\sum_{i=1}^{n}\rho_2\left(\frac{d(\boldsymbol{r}_t,\boldsymbol{\mu},\Sigma)}{cS(\boldsymbol{\mu},\Sigma)}\right) \tag{9.13}$$

such that $|\Sigma|=1$ and where $c$ is a control parameter of efficiency that depends on $\rho_2$. The BP of the $\tau$ estimator is 0.5.

The MVE and MCD estimators are quite prominent in portfolio optimization. But one has to be careful when using these estimators for small sample sizes and/or many variables. Since the estimators can only have a BP of 0.5 when half of the sample size is still sufficient for estimating the unknown parameters. Generally the robust estimators are affine invariant but can result in non-convex problems. The orthogonalized Gnanadesikan–Kettenring (OGK) estimator [Maronna and Zamar, 2002] avoids this problem by calculating the robust covariance matrix from robust univariate estimations. This leads to a positive semidefinite and approximately affine equivariant covariance matrix which is better suited for quadratic optimization. This also represents a framework where the BCP analysis as defined in Chapter 5 can be used to calculate the covariance matrix.

### 9.2.2 The Orthogonalized Gnanadesikan–Kettenring Estimator

The Gnanadesikan–Kettenring (GK) estimator [Gnanadesikan and Kettenring, 1972] is based on the following identity

$$\sigma_{kl}=\widehat{\mathrm{cov}}(\boldsymbol{r}_k,\boldsymbol{r}_l)=\frac{1}{4}\left(\hat{\sigma}(\boldsymbol{r}_k+\boldsymbol{r}_l)^2-\hat{\sigma}(\boldsymbol{r}_k-\boldsymbol{r}_l)^2\right) \tag{9.14}$$

where $\boldsymbol{r}_k$ is the column vector of the returns matrix $R$ that has $s$ observations for $n$ assets. This means that only the univariate estimates of $\sigma(\boldsymbol{r}_k+\boldsymbol{r}_l)$ and $\sigma(\boldsymbol{r}_k-\boldsymbol{r}_l)$ are needed to calculate the covariance between $\boldsymbol{r}_k$ and $\boldsymbol{r}_l$. And if those univariate variances are robust so will be the covariance matrix. The problem is that the resulting covariance matrix $\Sigma$ is not necessarily positive semidefinite and is not affine equivariant. Especially the non-positive semidefiniteness makes it unpreferable for quadratic programming.

Maronna and Zamar [2002] propose an algorithm to make the covariance matrix positive semidefinite and approximately affine equivariant after it has been estimated through Eq. (9.14). They called this estimator the orthogonalized Gnanadesikan–Kettenring (OGK) estimator which will be summarized in the remainder of this section.

Let $R_k$ be the row vectors and $\boldsymbol{r}_k$ the column vector of the returns matrix $R$ that has $s$ observations (rows) for $n$ assets (columns), $\sigma(\cdot)$ a robust variance estimator, $\mu(\cdot)$ a robust location estimator and $v(\cdot,\cdot)$ a robust estimate for the covariance (e.g. by using $\sigma(\cdot)$ in Eq. (9.14)). Then the OGK estimator can be calculated as

1. Calculate $D = \text{diag}(\sigma(\boldsymbol{r}_1), \ldots, \sigma(\boldsymbol{r}_n))$ and $Y = RD^{-1}$.
2. Compute the matrix $U$ where $U_{ii} = 1$ and $U_{ij} = v(\boldsymbol{r}_i, \boldsymbol{r}_j)$.
3. Compute the eigenvalues $\lambda_i$ and eigenvectors $\boldsymbol{e}_i$ of U such that $U = E\Lambda E'$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $E = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]$.
4. Calculate $A = DE$ and $Z = YE$. Note that $R = ZA'$.
5. Calculate the parameters as $\Sigma(R) = A\Gamma(Z)A'$ and $\boldsymbol{\mu}(R) = A\boldsymbol{\nu}(Z)$ where $\Gamma = \text{diag}(\sigma(\boldsymbol{z}_1)^2, \ldots, \sigma(\boldsymbol{z}_n)^2)$ and $\boldsymbol{\nu} = (\mu(\boldsymbol{z}_1), \ldots, \mu(\boldsymbol{z}_n))$.

The first step makes the estimates scale-equivariant. In the second step the covariances are estimated. After the fourth step the columns of $Z$ should be approximately uncorrelated and its covariance matrix $\Gamma$ approximately diagonal. In the last step the covariance matrix $\Gamma$ is transformed back to its original coordinate system. The same is true for the location parameters.

The method can be iterated. For that $\Sigma$ and $\boldsymbol{\mu}$ are calculated for $Z$ after step 4. Then the estimates for $R$ are calculated as $\Sigma(R) = A\Sigma(Z)A'$ and $\boldsymbol{\mu}(R) = A\boldsymbol{\mu}(Z)$. This can be repeated as often as desired. Maronna and Zamar [2002] propose a further improvement by a reweighing step. For that the Mahalanobis distances of the row vectors $R_i$ are calculated using the robust estimates of $\Sigma$ and $\boldsymbol{\mu}$. Based on these distances a hard rejection scheme is proposed to calculate a weighted version of $\Sigma$ and $\boldsymbol{\mu}$.

### 9.2.3 The Univariate $\tau$ Estimate

To calculate an OGK estimate a robust estimator for the univariate variance and mean has to be chosen. The robust covariance estimate is then given through Eq. (9.14). Maronna and Zamar [2002] propose the univariate $\tau$ estimate as defined in Eq. (9.13) for

$\sigma$ and a weighted mean for $\mu$. The weighted mean can be calculated as

$$\begin{aligned} \mu_0 = \text{med}(\boldsymbol{r}_k), \quad \sigma_0 = \text{MAD}(\boldsymbol{r}_k) = \text{med}(|\boldsymbol{r}_k - \mu_0|) \\ W_c(x) = \left(1 - \left(\frac{x}{c}\right)^2\right)^2 I(|x| \leq c) \\ w_i = W_{c_1}\left(\frac{r_i - \mu_0}{\sigma_0}\right) \\ \mu_k = \hat{\mu}(\boldsymbol{r}_k) = \frac{\sum_{i=1}^{s} w_i r_i}{\sum_{i=1}^{s} w_i} \end{aligned} \tag{9.15}$$

where $I$ is the indicator function and $r_i$ is the i-th observation of the $s$ observations of $\boldsymbol{r}_k$. The standard estimations are then calculated as

$$\begin{aligned} \rho_c = min(x^2, c^2) \\ \hat{\sigma}(\boldsymbol{r}_k)^2 = \frac{\sigma_0^2}{s} \sum_{i=1}^{s} \rho_{c_2}\left(\frac{r_i - \hat{\mu}(\boldsymbol{r}_k)}{\sigma_0}\right) \end{aligned} \tag{9.16}$$

Amongst many robust estimators such as the MVE or MCD estimators the **robustbase** package [Maechler et al., 2016] for `R` implements the OGK estimate and the univariate $\tau$ estimates. There the estimation of Eq. (9.16) is divided by a consistency factor and a finite sample correction factor to make it consistent with the normal model and unbiased for small sample sizes. The estimation of the variance is then given as

$$\sigma_k^2 = \hat{\sigma}(\boldsymbol{r}_k)^2 = \frac{\sigma_0^2}{(s-2)b} \sum_{i=1}^{s} \rho_{c_2}\left(\frac{r_i - \hat{\mu}(\boldsymbol{r}_k)}{\sigma_0}\right) \tag{9.17}$$

where $s - 2$ expresses the two degrees of freedom within the calculation of $\rho_{c_2}$ and $b$ can be calculated as

$$\begin{aligned} b = E_\Phi(\rho_c) &= \int_{-c}^{c} x^2 d_\Phi(x) + c^2 P(|x| > c) \\ &= 2\int_0^c x^2 d_\Phi(x) + 2c^2(1 - \Phi(c)) \\ &= 2\left[\Phi(x) - x\varphi(x)\right]_0^c + 2c^2(1 - \Phi(c)) \\ &= 2\Phi(c) - 2c\varphi(c) - 1 + 2c^2(1 - \Phi(c)) \\ &= 2\left((1 - c^2)\Phi(c) - c\varphi(c) + c^2\right) - 1 \end{aligned} \tag{9.18}$$

where $c = c_2\Phi^{-1}(3/4)$ to take into account the consistency correction of the MAD ($\sigma_0$).

### 9.2.4 The Feasible Set

The feasible set can now be calculated as described in Section 9.1. But instead of using the sample estimators for $\Sigma$ and $\boldsymbol{\mu}$ the robust OGK estimators are used as described in Section 9.2.2 by using the univariate $\tau$ estimates as described in Section 9.2.3. Since we will use rather small sample sizes for the calculation of the feasible set in the remainder of this thesis the additional reweighing step was not considered. The result can be seen in Fig. 9.2.

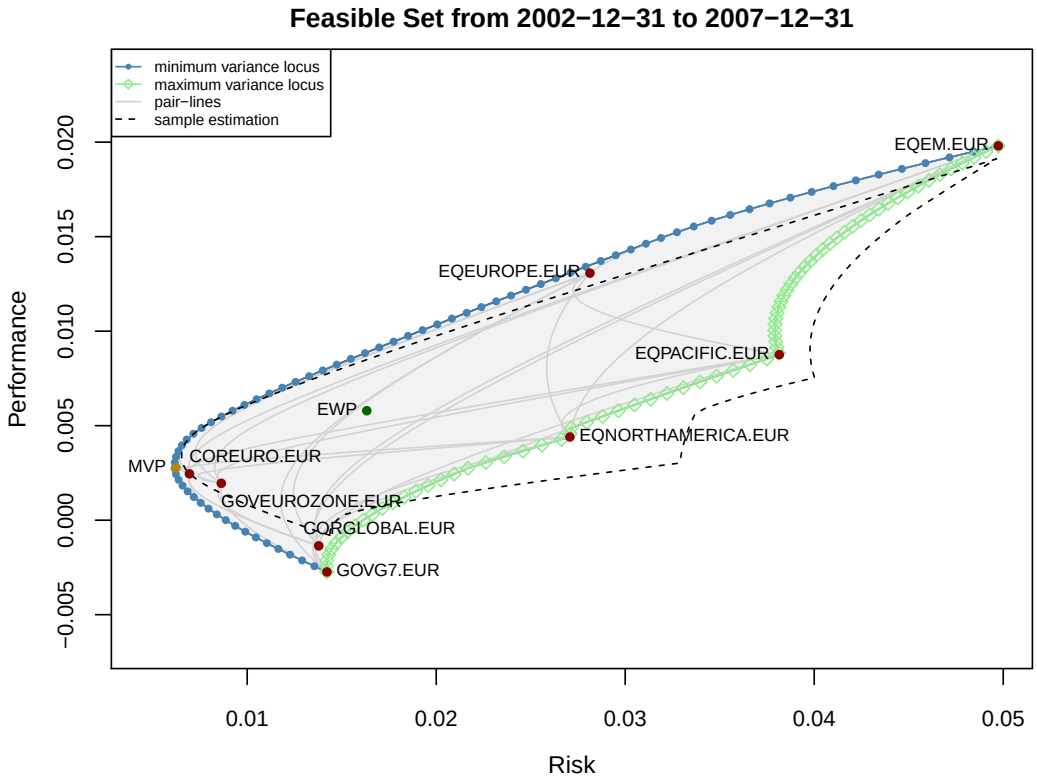


Figure 9.2: The feasible set of the global equity and bond universe using robust estimations. See also Appendix A for more information about the data.

Of course it is possible to calculate the feasible set by using any other kind of robust estimator such as MVE or MCD. Note that the interpretation of the sample and the robust feasible set differs. If one would have invested in e.g. the European equities

(EQEUROPE.EUR) then the profit would have been around 1% per month for the given time period (as can be seen in Fig. 9.1). In Fig. 9.2 the profit is calculated to be around 1.3%. But the robust estimation is not what one would have experienced (in contrary to the sample estimation). It is rather an expectation of the profit under the absence of outliers. The same is true for any other portfolio within the feasible set.

## 9.3 BCP Estimation

### 9.3.1 Motivation

Since the BCP model as described in Chapter 5 delivers univariate estimates of the location $\mu(\cdot)$ and the dispersion $\sigma(\cdot)$ parameters it is apparent to use these estimates to calculate the covariance matrices through Eq. (9.14) and to make the matrix positive semidefinite as described in Section 9.2.2.

In contrary to common estimations that in the univariate setup deliver at any point in time one value for the location and dispersion the BCP estimation delivers estimations for any point in time. A straightforward approach would be to take just the most recent estimation. When there are no change points this estimation would be similar to the sample estimator. If there are recent change points then this estimation is based on the more recent data.

This will make the feasible set change its shape quickly if a new regime is about to form. Which raises some questions about our individual understanding of financial markets. Robust estimation is driven by the assumption that the financial returns are subjected to sporadic outliers which can be ignored to a certain extend to estimate the parameters. While the BCP model might see extreme values as the formation of a new dynamic and therefore react on these outliers.

The BCP estimation is therefore driven by a different thinking that is based on changing dynamics. For the robust case robust univariate estimations are transformed into a robust multivariate estimation. For the BCP case dynamic univariate estimations are transformed into a dynamic multivariate estimation. The two approaches will be compared in Chapter 10 through the GSF.

### 9.3.2 The Exponential Moving Average

Using only the most recent estimations of a BCP analysis might result in very drastic changes of the feasible set if a new observation occurs that can not be explained with the

older observations. That is if the posterior probabilities tend towards 1. An alternative would be to take also the older estimations into account by calculating the mean of all the estimations. While that would make the changes smooth again it would lead to very similar results as using the sample estimations and therefore not be of much benefit.

A possibility to lessen the effect of drastic changes and still favouring the more recent estimations is to use an exponential moving average (EMA) [Roberts, 1959] as

$$z_t = \begin{cases} x_1 & t = 1 \\ \lambda x_t + (1-\lambda) z_{t-1} & t > 1 \end{cases} \tag{9.19}$$

where $z$ is the EMA at time $t$. If we are just interested in the most recent estimation (at time $s$) the EMA can be formulated as

$$\mathrm{EMA}_\lambda(\boldsymbol{x}) = \sum_{t=2}^{s} \lambda(1-\lambda)^{s-t} x_t + (1-\lambda)^{s-1} x_1 \tag{9.20}$$

where $0 \leq \lambda \leq 1$ controls to which extend newer observations are favoured towards older observations. The higher $\lambda$ the more newer observations are favoured.

Let $M = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_n\}$ be the matrix that holds the BCP estimations (see Eq. (3.40)) of the trends for every asset where $\boldsymbol{m}_k = \{m_1, m_2, \ldots, m_s\}$ is the vector that holds the estimations for every point in time $t$. And in the same way let $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ be the matrix that holds the BCP estimations of the standard deviations. Using an EMA the univariate BCP estimation of the trend and the standard deviation of an asset's returns $\boldsymbol{r}_k$ can be calculated as

$$\hat{\mu}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\mu}(\boldsymbol{m}_k) \tag{9.21}$$

$$\hat{\sigma}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\sigma}(\boldsymbol{s}_k) \tag{9.22}$$

Note that for $\lambda = 1$ the result will be the most recent estimation and for $\lambda = 0$ the oldest observation.

### 9.3.3 The BCP Estimator

Using Eq. (9.14), Eq. (9.21) and Eq. (9.22) the BCP estimator can be summarized as

$$\begin{aligned} \mu_k = \hat{\mu}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\mu}(\boldsymbol{m}_k), \quad \sigma_k = \hat{\sigma}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\sigma}(\boldsymbol{s}_k) \\ \sigma_{kl} = \widehat{\mathrm{cov}}(\boldsymbol{r}_k, \boldsymbol{r}_l) = \frac{1}{4}\left(\hat{\sigma}(\boldsymbol{r}_k + \boldsymbol{r}_l)^2 - \hat{\sigma}(\boldsymbol{r}_k - \boldsymbol{r}_l)^2\right) \end{aligned} \tag{9.23}$$

where in an additional step the covariance matrix can be made positive semidefinite and approximately affine equivariant by using the orthogonalization procedure as described in Section 9.2.2.

Since the estimations $\mu_k$ and $\sigma_k$ are the parameters of the (assumed) normally distributed returns $\boldsymbol{r}_k$ an alternative would be to mix the sample correlations $\hat{\rho}(\boldsymbol{r}_k, \boldsymbol{r}_k)$ with the BCP variances as

$$\begin{aligned} \mu_k = \hat{\mu}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\mu}(\boldsymbol{m}_k), \quad \sigma_k = \hat{\sigma}(\boldsymbol{r}_k) = \mathrm{EMA}_{\lambda_\sigma}(\boldsymbol{s}_k) \\ \sigma_{kl} = \widehat{\mathrm{cov}}(\boldsymbol{r}_k, \boldsymbol{r}_l) = \hat{\sigma}(\boldsymbol{r}_k)\hat{\sigma}(\boldsymbol{r}_l)\hat{\rho}(\boldsymbol{r}_k, \boldsymbol{r}_l) \end{aligned} \tag{9.24}$$

Note that for this procedure the correlations do not react to any breakpoints (in contrary to Eq. (9.23)). The advantage is that it is computationally much more efficient and that the covariance matrix is positive semidefinite.

### 9.3.4 The Feasible Set

For Eq. (9.23) there are various variants. If it is calculated without the orthogonalization step the covariance matrix will not be positive semidefinite (at least it is the case for this dataset) and the feasible set can therefore not be calculated using a quadratic optimizer. Also it would be possible to calculate the estimations for $\mu_k$ from the orthogonalization step. For the feasible set in Fig. 9.3 the orthogonalization step was performed using two iterations to make $\sigma_{kl}$ positive semidefinite and approximately affine equivariant. The additional reweighing step was not considered. The estimations for $\mu_k$ are not taken from the orthogonalization step. Additionally the figure shows the feasible set calculated using Eq. (9.24) (BCP-Sample-Correlation Mixture).

For the univariate estimators $\mu_k$ and $\sigma_k$ the parameters $\lambda_\mu = \lambda_\sigma$ were set to 0.2. This means that approximately 93% of the total weight is within the first 12 observations. The parameters to calculate the BCP estimations $M$ and $S$ were set by first calibrating the model on the data (see Eq. (6.6) and Eq. (6.8)) using $v = 2$ and $v_2 = 1.5$ to make the

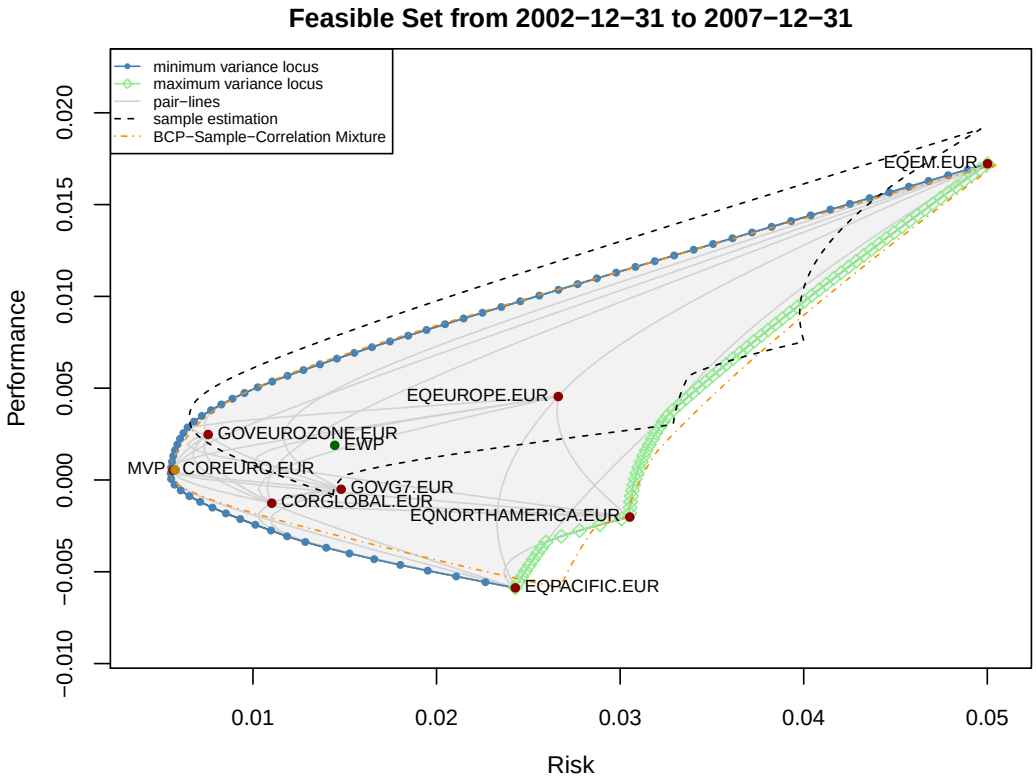


Figure 9.3: The feasible set of the global equity and bond universe using BCP estimations. See also Appendix A for more information about the data.

priors more flat. The breakpoint parameters $\alpha = 4$ and $\beta = 16$ were chosen such that the expected breakpoint probability is 20%. For performance reasons the calculation of the covariances was parallelised.

Fig. 9.1, Fig. 9.2 and Fig. 9.3 are only snapshots to demonstrate the calculations of different images of the feasible set. To get a better idea how these estimations compare the feasible sets are calculated on a sliding window (in Chapter 10) and compared through the GSF. This will also show how the feasible set changes its shape over time.

# 10 Morphological Shape Factors

Various fields of science are faced with the problem of classifying real images that show some variations to the ideal image. These includes remote sensing, astronomy, medicine or vision guided robotic systems. Image moments offer a solution to this problem since they can be constructed such that variations within the image lead to the same result. The initial mathematical framework was developed in the context of algebraic invariants [Hilbert, 1993; Gurevich, 1964; Schur, 1968]. Moment invariants for pattern recognition were introduced by Hu [1962] where various moments are defined that are invariant under scale, translation and/or rotation. A summary about more recent work can be found in Flusser [2007].

If the feasible set is understood as an image these moments can be calculated for the feasible set where some of them also have a geometrical meaning. The idea was presented by Wuertz [2010]. If this analysis is performed over time it offers valuable insights into the dynamics of the feasible set. The results will be a new set of risk measures where some of them are invariant under certain transformations.

## 10.1 The Feasible Set as an Image

To represent the feasible set as an image a grid is put on top of it. The hull of the feasible set can be described as a polygon. The grid points inside the feasible set receive a value of 1 and the grid points outside the feasible set a value of 0 as

$$\begin{aligned} I(x,y) &= 1, \qquad \text{for grid points inside the polygon} \\ I(x,y) &= 0, \qquad \text{for grid points outside the polygon} \end{aligned} \tag{10.1}$$

where $(x, y)$ are the coordinates of a grid point within the risk-performance-plane (see Fig. 10.1). The feasible set in Fig. 10.1 was calculated by using the sample estimators (see Section 9.1). The same procedure can be applied to the robust or BCP estimations.

There are various algorithms to find points inside a polygon. For `R` the **sp** package

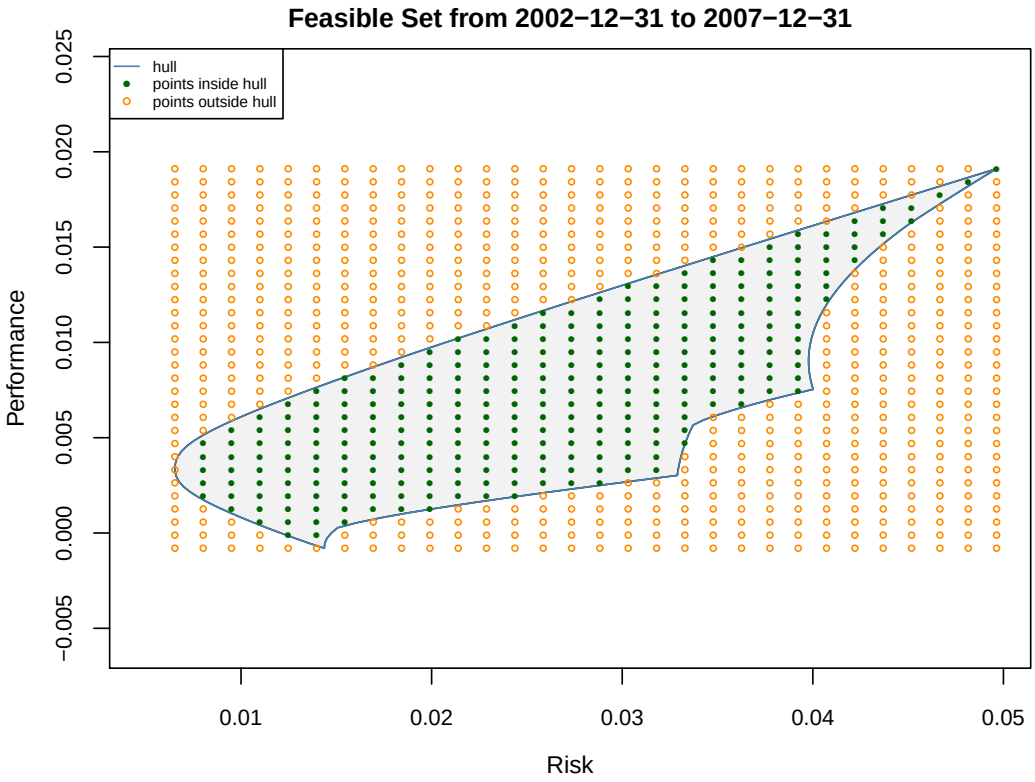


Figure 10.1: Grid points of the feasible set using 30 points in each direction.

[Pebesma and Bivand, 2005] offers an algorithm. The method is described in detail in O'Rourke [1998]. The higher the resolution of the grid (more grid points) the more exact the calculation of the moments will become.

## 10.2 Raw Moments

For the continuous case the (p+q)th order raw moments (see Hu [1962]) of a density function $f(x, y)$ are defined as

$$m_{p,q} = \int_{\infty}^{\infty} \int_{\infty}^{\infty} x^p y^q f(x, y) dx dy, \quad p, q = 0, 1, 2, \ldots \tag{10.2}$$

To calculate the raw moments of the discrete image as defined in Eq. (10.1) one can

use the discrete representation of Eq. (10.2) as

$$m_{p,q} = \sum_x \sum_y x^p y^q I(x,y) \Delta x \Delta y, \quad p,q = 0,1,2,\ldots \tag{10.3}$$

where the summation goes over all possible combinations of $x$ and $y$, and where $\Delta x$ and $\Delta y$ are constant.

Using the raw moments the area $A$, the centre $(x_c, y_c)$ and the centre ratio $c_s$ of the feasible set can be calculated as

$$A = m_{00} \tag{10.4}$$

$$(x_c, y_c) = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}) \tag{10.5}$$

$$c_s = \frac{y_c}{x_c} \tag{10.6}$$

The area is invariant under translation and rotation while the centre is invariant under scale and rotation.

## 10.3 Central Moments

For the continuous case the (p+q)th order central moments (see Hu [1962]) are defined as

$$\mu_{p,q} = \int_\infty^\infty \int_\infty^\infty (x - x_c)^p (y - y_c)^q f(x,y) dx dy, \quad p,q = 0,1,2,\ldots \tag{10.7}$$

and for the discrete case the central moments can be calculated as

$$\mu_{p,q} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q I(x,y) \Delta x \Delta y, \quad p,q = 0,1,2,\ldots \tag{10.8}$$

where the summation goes over all possible combinations of $x$ and $y$, and where $\Delta x$ and $\Delta y$ are constant.

The central moments can be formulated as a function of the raw moments. The binomial theorem for a positive integer $n$ (see e.g. Johnson et al. [2005]) is defined as

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j \tag{10.9}$$

Applying this to Eq. (10.8) leads to

$$\begin{aligned}
\mu_{p,q} &= \sum_x \sum_y (-x_c + x)^p (-y_c + y)^q I(x,y) \Delta x \Delta y \\
&= \sum_x \sum_y \left( \sum_{i=0}^{p} \binom{p}{i} (-x_c)^{p-i} x^i \sum_{j=0}^{q} \binom{q}{j} (-y_c)^{q-j} y^j \right) I(x,y) \Delta x \Delta y \\
&= \sum_{i=0}^{p} \binom{p}{i} (-x_c)^{p-i} \sum_{j=0}^{q} \binom{q}{j} (-y_c)^{q-j} \sum_x \sum_y x^i y^j I(x,y) \Delta x \Delta y \\
&= \sum_{i=0}^{p} \sum_{j=0}^{q} \binom{p}{i} \binom{q}{j} (-x_c)^{p-i} (-y_c)^{q-j} m_{i,j}
\end{aligned} \tag{10.10}$$

which means that for the calculation of the central moments only the calculation of the raw moments are needed. The central moments are translation invariant.

The orientation $\theta$ and eccentricity $e$ of the feasible set can be found through a principal component analysis (PCA) [Pearson, 1901; Hotelling, 1933; Abdi and Williams, 2010]. For that one has to solve the eigenvalue problem of the covariance matrix of the discrete image of the feasible set which is defined as

$$\Sigma = \begin{bmatrix} \mu_{2,0} & \mu_{1,1} \\ \mu_{1,1} & \mu_{0,2} \end{bmatrix} \tag{10.11}$$

The eigenvalue problem is defined as $A\lambda = \lambda \boldsymbol{v}$ where $\lambda$ is the eigenvalue and $\boldsymbol{v}$ the eigenvector of the matrix $A$. The problem can be reformulated as $(A - \lambda I)\boldsymbol{v} = 0$ where $I$ is the identity matrix. For a 2x2 matrix this leads to the following equation:

$$(A - \lambda I)\boldsymbol{v} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{10.12}$$

The equation has non-zero solutions for $\boldsymbol{v}$ if the determinant of $A - \lambda I$ equals to 0. The eigenvalues can then be calculated as

$$\begin{aligned}
|A - \lambda I| &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12} a_{21} = 0 \\
\lambda_{1,2} &= \frac{a_{11} + a_{22}}{2} \pm \frac{\sqrt{4 a_{12} a_{21} + (a_{11} - a_{22})^2}}{2}
\end{aligned} \tag{10.13}$$

$$= \frac{\mu_{2,0} + \mu_{0,2}}{2} \pm \frac{\sqrt{4\mu_{1,1}^2 + (\mu_{2,0} - \mu_{0,2})^2}}{2} \tag{10.14}$$

The eccentricity can be calculated through the eigenvalues as

$$e = \sqrt{1 - \frac{\lambda_2}{\lambda_1}} \tag{10.15}$$

The following eigenvectors satisfy Eq. (10.12):

$$\begin{aligned} \boldsymbol{v_1} &= \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} = \begin{bmatrix} a_{12} \\ \lambda - a_{11} \end{bmatrix} \\ \boldsymbol{v_2} &= \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} = \begin{bmatrix} \lambda - a_{22} \\ a_{21} \end{bmatrix} \end{aligned} \tag{10.16}$$

where $\theta$ describes the angle of the eigenvector to its nearest axis. Note that per definition the relation $(\lambda - a_{11})/a_{12} = a_{21}/(\lambda - a_{22})$ holds. Which means the orientation angle satisfies the following equation system:

$$\begin{aligned} \tan(\theta) &= \frac{\lambda - a_{11}}{a_{12}} \\ \tan(\theta) &= \frac{a_{21}}{\lambda - a_{22}} \end{aligned}$$

By eliminating $\lambda$ the orientation angle $\theta$ can be calculated as

$$\begin{aligned} a_{12}\tan(\theta) + a_{11} &= \frac{a_{21}}{\tan(\theta)} + a_{22} \\ \mu_{1,1}(1 - \tan^2(\theta)) &= (\mu_{2,0} - \mu_{0,2})\tan(\theta) \\ \theta &= \frac{1}{2}\arctan\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \end{aligned} \tag{10.17}$$

using the relation $\tan(2\theta) = 2\tan(\theta)/(1 - \tan^2(\theta))$. Note that the eigenvectors are perpendicular to each other. Calculating the angle like this leads to two singularities. If the angle approaches $\pi/4$ (towards the x-axis) it will jump to $-\pi/4$ (towards the y-axis). And if the angle approaches $\pi/4$ (towards the y-axis) it will jump to $-\pi/4$ (towards the x-axis). On the other hand; if the order of the eigenvalues changes (the smaller one becomes the bigger one); the angle will be the same.

Alternatively one could also calculate the angle from the eigenvector ($\boldsymbol{v} = [v_1, v_2]$) of

the biggest eigenvalue towards the x-axis as

$$\theta = \arctan \frac{v_2}{v_1} \tag{10.18}$$

This leads to two singularities as well. If the angle approaches $\pi/2$ (towards the x-axis) it will jump to $-\pi/2$ (towards the x-axis). If the order of the eigenvalues changes the angle will jump from $\theta$ to $\theta - (\pi/4)$.

The orientation is invariant under translation and scale while the eccentricity is invariant under translation, scale and rotation.

## 10.4 Scale and Rotation Invariant Moments

Scale and translation invariant moments were presented by Hu [1962] (see also Flusser [2007]) as

$$\nu_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\frac{p+q}{2}+1}} \tag{10.19}$$

In the same paper Hu [1962] presents seven moments that are invariant to translation, scale and rotation as

$$h_1 = \nu_{2,0} + \nu_{0,2} \tag{10.20}$$

$$h_2 = (\nu_{2,0} - \nu_{0,2})^2 + 4\nu_{1,1}^2 \tag{10.21}$$

$$h_3 = (\nu_{3,0} - 3\nu_{1,2})^2 + (3\nu_{2,1} - \nu_{0,3})^2 \tag{10.22}$$

$$h_4 = (\nu_{3,0} + \nu_{1,2})^2 + (\nu_{2,1} + \nu_{0,3})^2 \tag{10.23}$$

$$\begin{aligned} h_5 =& (\nu_{3,0} - 3\nu_{1,2})(\nu_{3,0} + \nu_{1,2})[(\nu_{3,0} + \nu_{1,2})^2 - 3(\nu_{2,1} + \nu_{0,3})^2] \\ &+ (3\nu_{2,1} - \nu_{0,3})(\nu_{2,1} + \nu_{0,3})[3(\nu_{3,0} + \nu_{1,2})^2 - (\nu_{2,1} + \nu_{0,3})^2] \end{aligned} \tag{10.24}$$

$$\begin{aligned} h_6 =& (\nu_{2,0} - \nu_{0,2})[(\nu_{3,0} + \nu_{1,2})^2 - (\nu_{2,1} + \nu_{0,3})^2] \\ &+ 4\nu_{1,1}(\nu_{3,0} + \nu_{1,2})(\nu_{2,1} + \nu_{0,3}) \end{aligned} \tag{10.25}$$

$$\begin{aligned} h_7 =& (3\nu_{2,1} - \nu_{0,3})(\nu_{3,0} + \nu_{1,2})[(\nu_{3,0} + \nu_{1,2})^2 - 3(\nu_{2,1} + \nu_{0,3})^2] \\ &- (\nu_{3,0} - 3\nu_{1,2})(\nu_{2,1} + \nu_{0,3})[3(\nu_{3,0} + \nu_{1,2})^2 - (\nu_{2,1} + \nu_{0,3})^2] \end{aligned} \tag{10.26}$$

also known as the Hu moments. The seventh moment is also skew invariant. Which

allows to distinguish between mirrored images.

## 10.5 Geometric Shape Factors

The geometric shape factors (GSF) are the parameters that define an ellipse. These are the area (Eq. (10.4)), the centre (Eq. (10.5)), the orientation (Eq. (10.18)) and the eccentricity (Eq. (10.15)). Instead of calculating theses parameters through the image moments they can also be calculated by using the polygon that describes the feasible set. These calculations are more precise and computationally more efficient since they don't rely on the representation of the feasible set as an image ($I(x, y)$) but only on the coordinates of the hull of the feasible set.

Let $P = [\boldsymbol{x}, \boldsymbol{y}]$ be the coordinates of the hull of the feasible set (the polygon). The hull is visualized in Fig. 10.2. The area and the centre of the feasible set (see e.g. Bourke [1988]) can be calculated from the coordinates of the hull as

$$A = \frac{1}{2} \sum_{i=1}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \tag{10.27}$$

$$x_c = \frac{1}{6A} \sum_{i=1}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{10.28}$$

$$y_c = \frac{1}{6A} \sum_{i=1}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{10.29}$$

The orientation and the eccentricity can be calculated in exactly the same way as described in Section 10.3. With the only difference that the covariance matrix is calculated from the coordinates of the hull as

$$\Sigma = \begin{bmatrix} \mathrm{var}(\boldsymbol{x}) & \mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) \\ \mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) & \mathrm{var}(\boldsymbol{y}) \end{bmatrix} \tag{10.30}$$

$$\theta = \arctan \frac{v_2}{v_1} \tag{10.31}$$

$$e = \sqrt{1 - \frac{\lambda_2}{\lambda_1}} \tag{10.32}$$

where $\lambda_{1,2}$ are the eigenvalues of $\Sigma$ and $\boldsymbol{v} = [v_1, v_2]$ is the eigenvector of the biggest eigenvalue. Note that the orientation and the eccentricity of the points inside the polygon

tend towards the orientation and the eccentricity of the hull of the polygon if the resolution of the grid is increased. The same is true for the area and the centre.

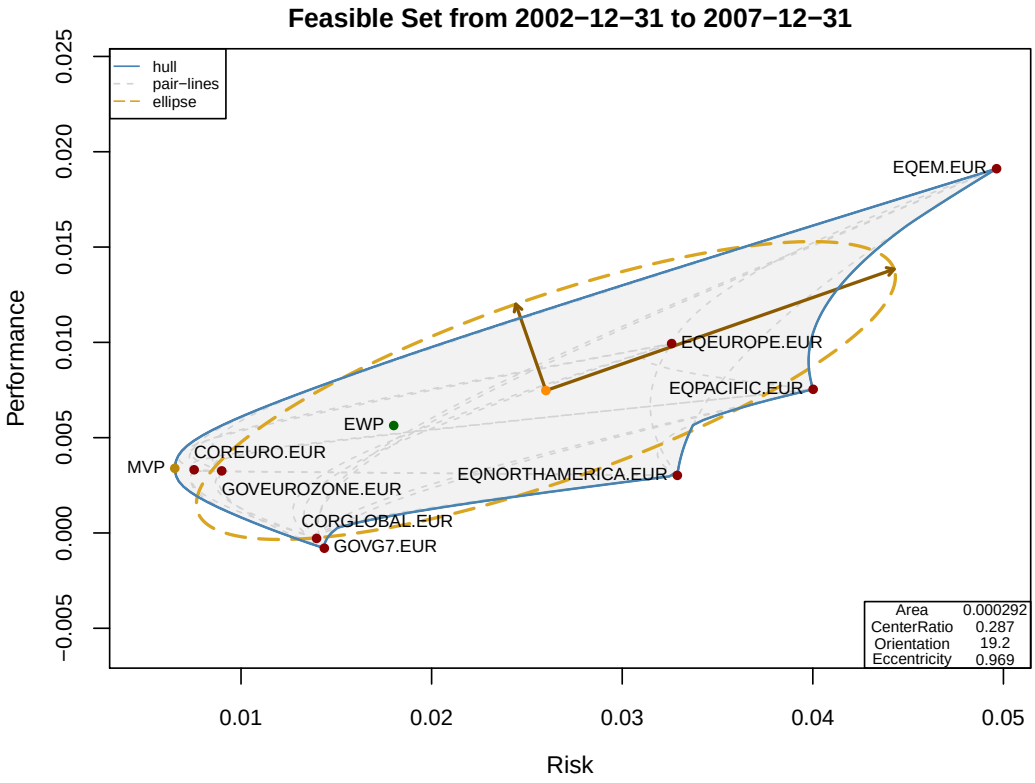


Figure 10.2: The feasible set of the global equity and bond universe using sample estimations. Plotted on the top of the feasible set is the ellipse that has the same area, centre, orientation and eccentricity as the feasible set. See also Appendix A for more information about the data.

To plot the ellipse that has area $A$, centre $(x_c, y_c)$, orientation $\theta$ and eccentricity $e$ the semi major axis $a$ and the semi minor axis $b$ are calculated as

$$ab = \frac{A}{\pi}, \quad \frac{a}{b} = \frac{1}{\sqrt{1-e^2}}$$

$$a = \frac{A}{\pi\sqrt{1-e^2}} \tag{10.33}$$

$$b = \frac{A\sqrt{1-e^2}}{\pi} \tag{10.34}$$

The coordinates of the ellipse $E = [\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}]$ can then be calculated as

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{x}}\\ \hat{\boldsymbol{y}}\end{bmatrix} &= \begin{bmatrix}x_c\\ y_c\end{bmatrix} + \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}a\cos(t)\\ b\sin(t)\end{bmatrix}\\ &= \begin{bmatrix}x_c + a\cos(t)\cos(\theta) - b\sin(t)\sin(\theta)\\ y_c + a\cos(t)\sin(\theta) + b\sin(t)\cos(\theta)\end{bmatrix}\end{aligned} \tag{10.35}$$

for $0 \leq t \leq 2\pi$. The result is shown in Fig. 10.2. In the same way the GSF can be calculated for the feasible set using robust or BCP estimators. In Chapter 11 the GSF will be calculated over time for different investment universes. By visualizing the area, centre ratio, orientation and eccentricity over time it is possible to get an idea of the shape of the feasible set at any point in time without visualizing the feasible set itself.

# 11 Applications

The images from Fig. 11.1 to Fig. 11.8 show the GSF as described in Section 10.5 and the Hu moments (Eq. (10.20) to Eq. (10.26)) for different investment universes. Different methods to calculate the feasible set are compared. This are the sample estimations (Eq. (9.8) and Eq. (9.9)), the robust estimations (see Section 9.2.2 and Section 9.2.3) and the two versions of the BCP estimations (Eq. (9.23) and Eq. (9.24)). The GSF and the Hu moments were calculated at any point in time by using the last 36 monthly returns.

## 11.1 Global Universe

The global equity and bond universe (Fig. 11.1 and Fig. 11.2) is the same as used in Chapter 9 and Chapter 10. It consists of global equities, government bonds for the G7 and eurozone countries and global corporate bonds (in euro). The analyses exemplarily illustrate the behaviour during a crisis. In Fig. 10.2 a time frame was chosen where the feasible set has a shape as it is often assumed. It shows a risk premium (positive orientation) from the bonds towards the equities. In Fig. 11.1 the time period from around mid of 2007 until around the beginning of 2009 can be associated with a crisis within the given investment universe.

It is apparent that the orientation becomes negative during that time which means that the risk premium becomes negative. Generally speaking one can observe some patterns during the unfolding of a crisis. At first single assets start to move away from the bulk of the assets; this makes the area larger. Usually these assets tend to move towards a lower performance and a higher risk (variance). For the global equity and bond universe the equities will turn around the bonds. This moves the orientation from positive to negative. Once the bulk starts to follow the first movers the centre will move towards a lower performance and a higher risk as well. The eccentricity seems to be rather high during these times which means that confidence in the (negative) risk premium is high.

The Hu moments are invariant under translation, scale and rotation. Which means that if the feasible set is just moving to a different location (translation), the performance

and the variance of the underlying assets change but not the Sharpe ratio (scale) or the feasible set rotates the Hu moments will be the same. Whenever the Hu moments are changing it means that the feasible set shows a unique shape and that the underlying investment universe is therefore in a unique state. During the time period from 2013 to 2015 the investment universe was more or less identical under translation, scale and rotation. Whenever the Hu moments change drastically it means that there are fundamental changes within the investment universe that cannot be associated with translation, scaling or rotation.

Comparing the different estimators shows that the two variants of the BCP estimation deliver almost the same results. The line that is labelled with "BCP1" shows the estimation using Eq. (9.23) and the line that is labelled with "BCP2" shows the estimation using Eq. (9.24). The robust estimation does not change as much as the BCP estimations and the sample estimations lie somewhere in between. Interestingly the robust estimation shows the largest changes for the estimation of the eccentricity between 2015 and 2016.

## 11.2 Various Universes

The European equity universe (Fig. 11.3 and Fig. 11.4) consists of the 19 European industry sectors (in euro). It shows similar behaviour during the crises of around 2002 and 2008 as the global equity and bond universe. The mechanisms are the same. During the crisis of 2008 it shows singularities within the orientation (see also Section 10.3 for the reason why). Therefore one has to be careful when designing a portfolio based upon the orientation to make sure that the desired portfolio is the same when the orientation is around $\pm 90$ degrees (see also Chapter 15).

The pure bond universe (Fig. 11.5 and Fig. 11.6) consists of the government bonds of 11 of the countries that have adopted the euro. Additionally it contains a bond index for the United States, the United Kingdom and the European Union as a whole (in euro). It shows the largest changes during the European debt crisis which started to unfold in 2010. It was mainly dominated through the movement of the Greek and Portuguese bonds away from the bulk. The Hu moments show that up to that point the bond universe was identical under translation, scale and rotation.

The European bonds and precious metals universe (Fig. 11.7 and Fig. 11.8) consists of three indices which describe the eurozone bond market with maturities from one to ten years and an index that describes the global corporate bond market for bonds issued in euros. The precious metals are described through the euro hedged prices of gold, silver,

palladium and platinum. The results are mainly characterized through the movements of the precious metals around the bonds. For portfolios that consist of different asset classes the orientation usually doesn't show singularities. Also the eccentricity is generally high. This universe will be used in Chapter 15 to define a portfolio that based on the orientation tries to position itself optimally within the feasible set.

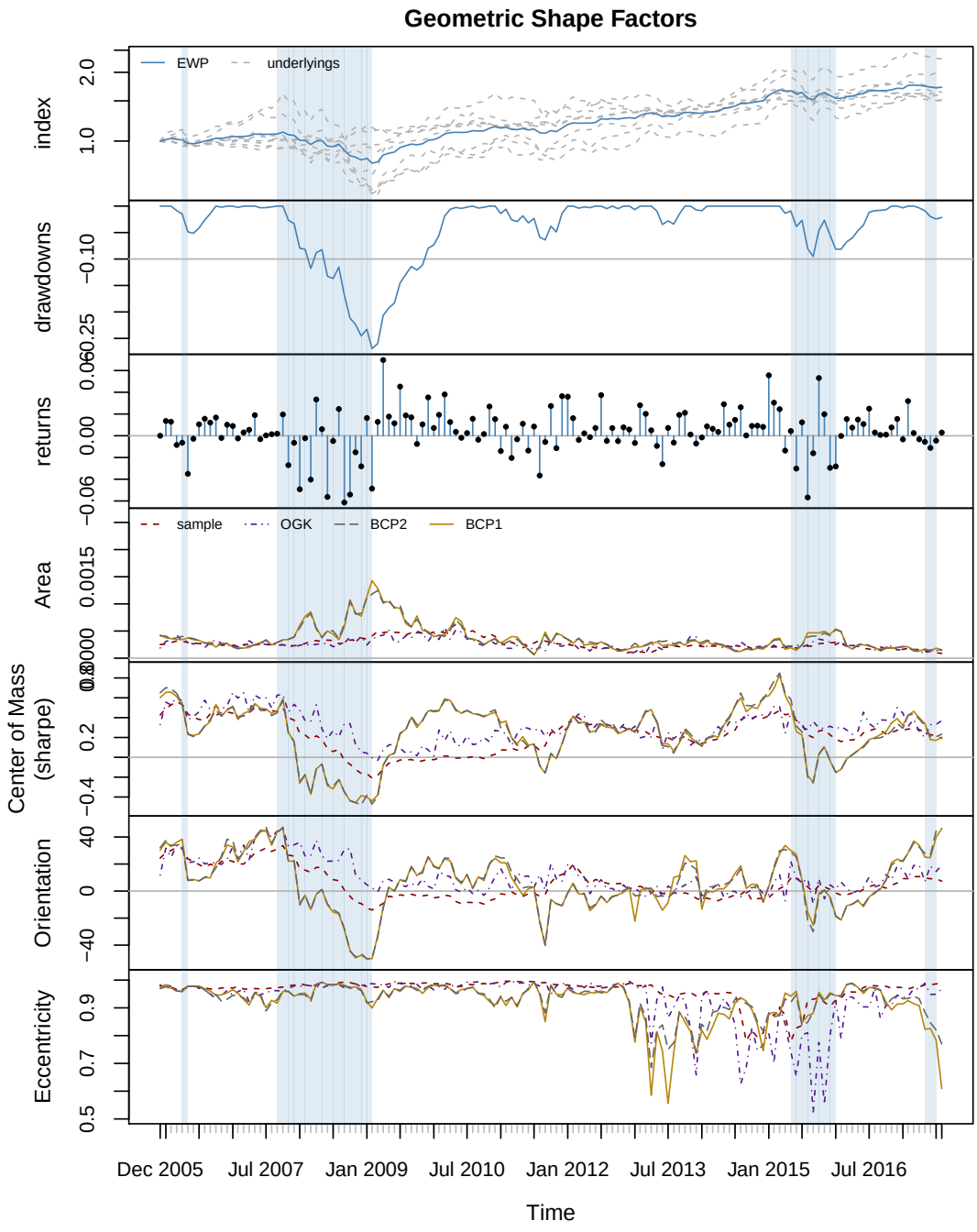


Figure 11.1: The GSF for the global equity and bond universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

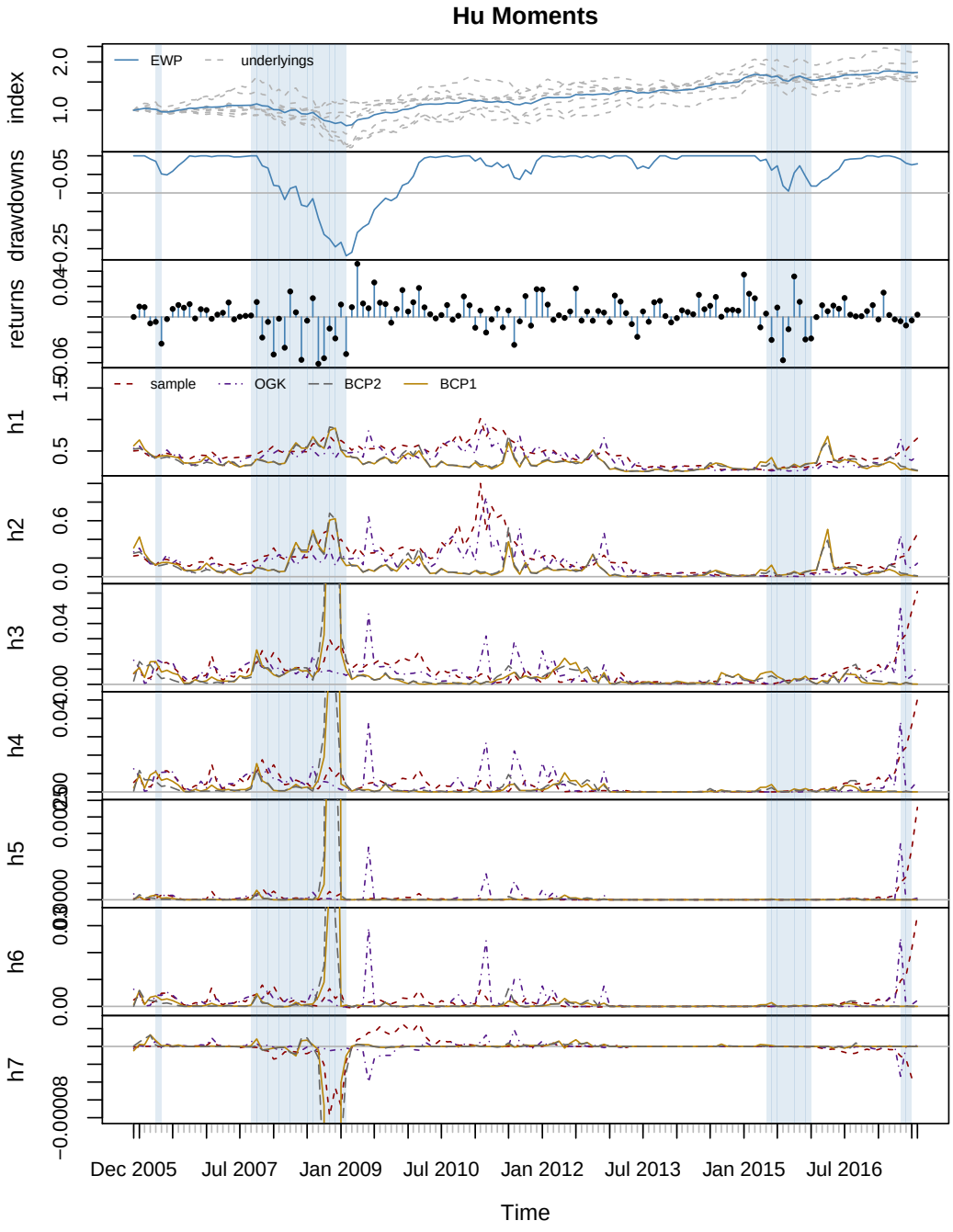


Figure 11.2: The Hu Moments for the global equity and bond universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

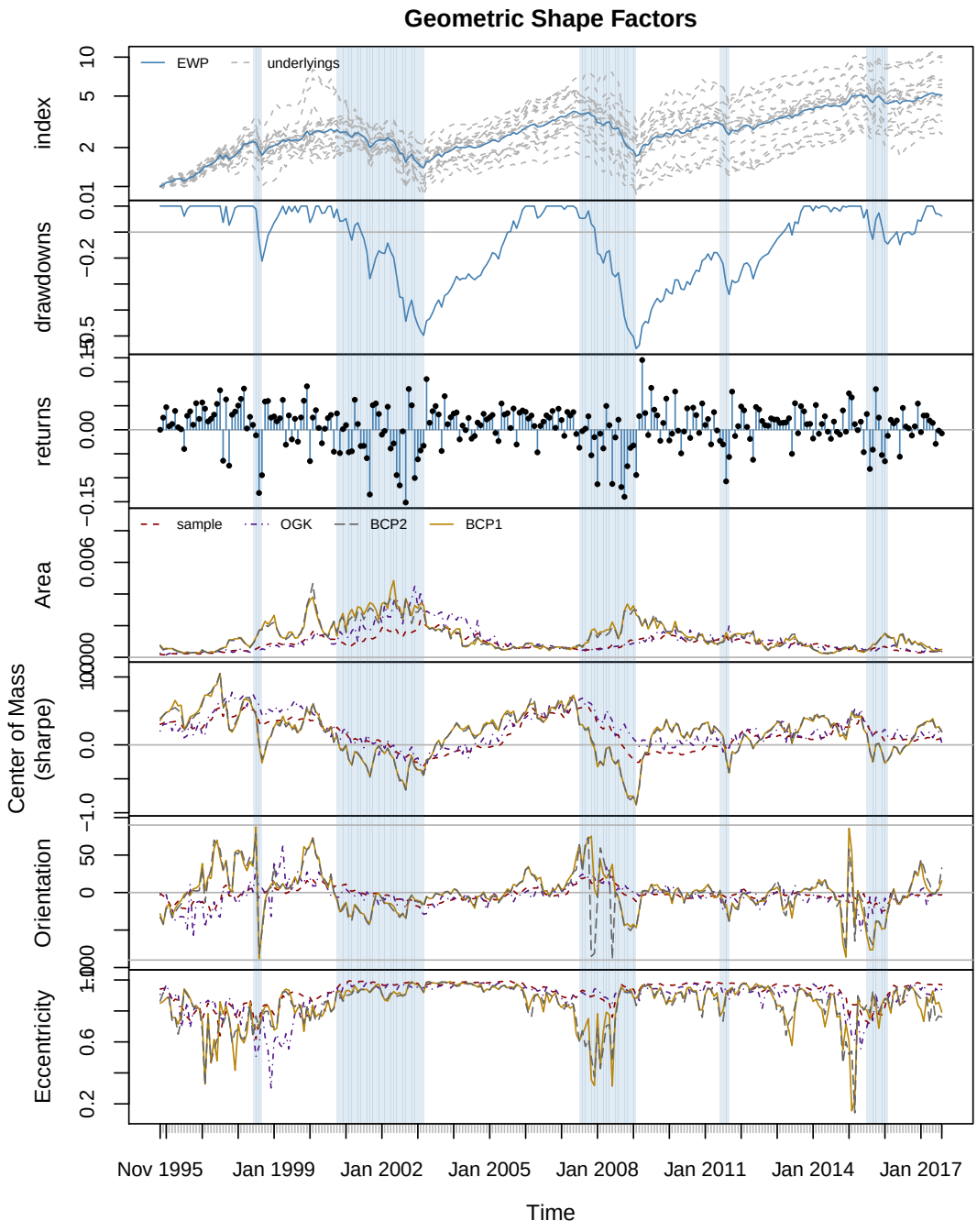


Figure 11.3: The GSF for the European equity universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

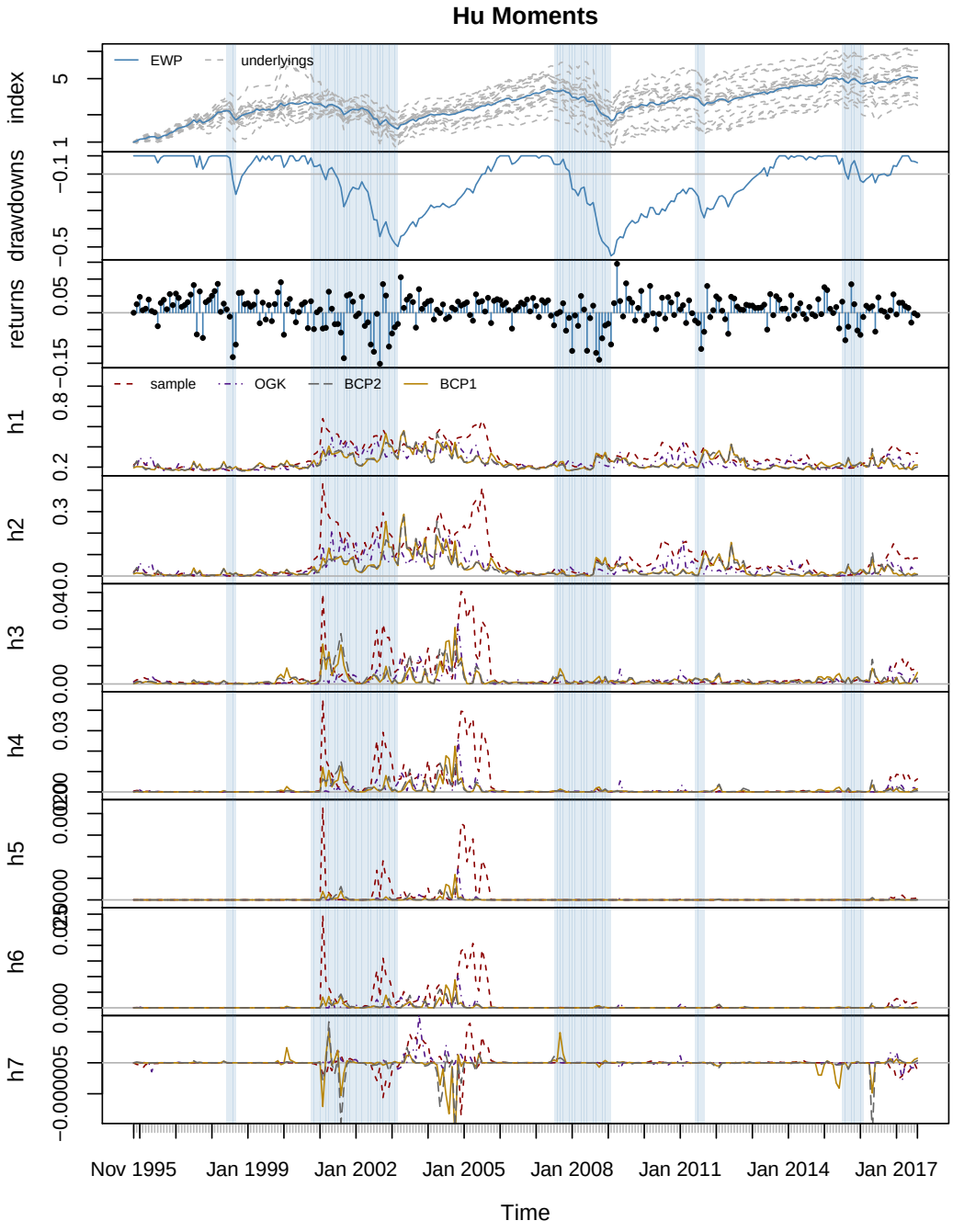


Figure 11.4: The Hu Moments for the European equity universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

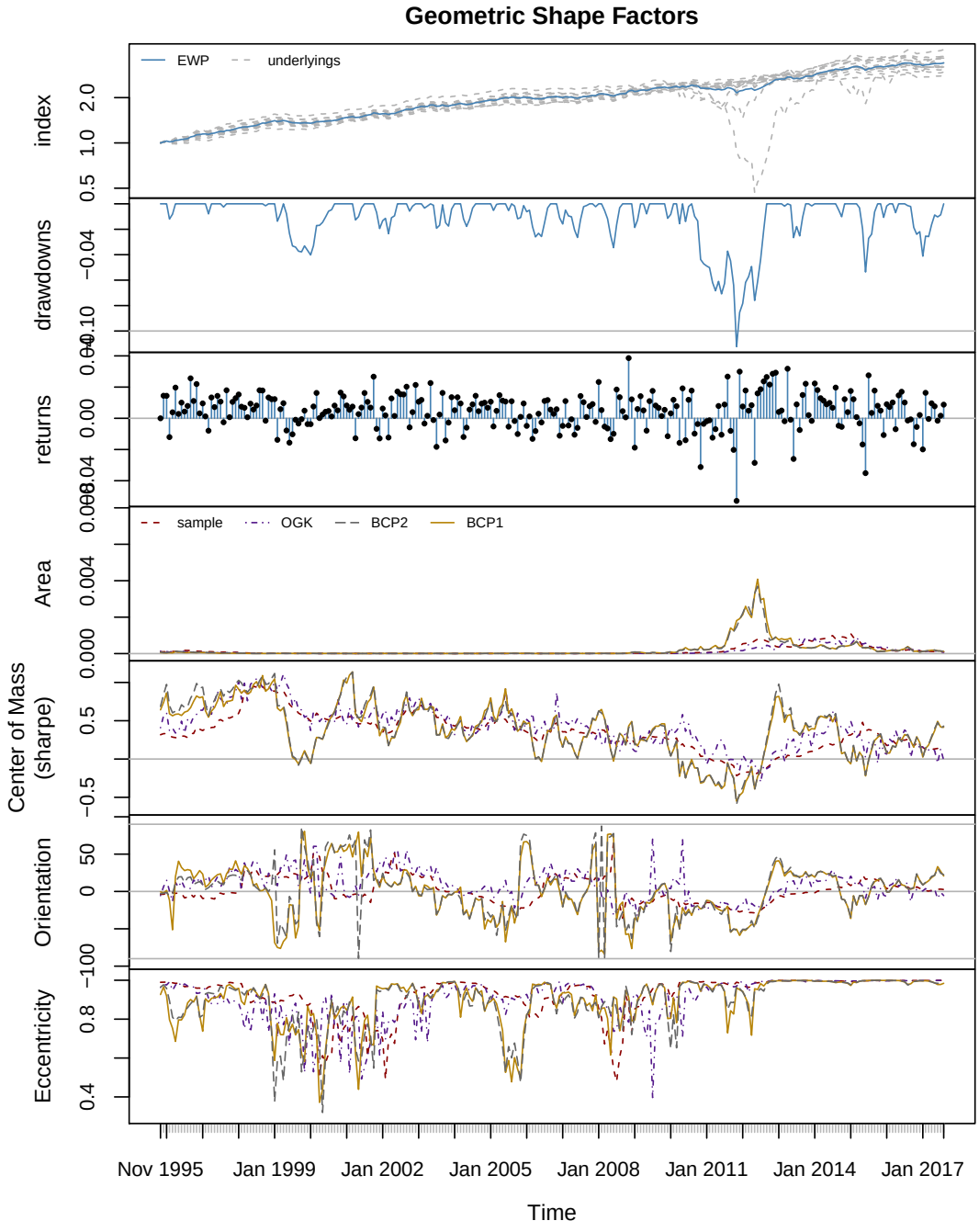


Figure 11.5: The GSF for the pure bond universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

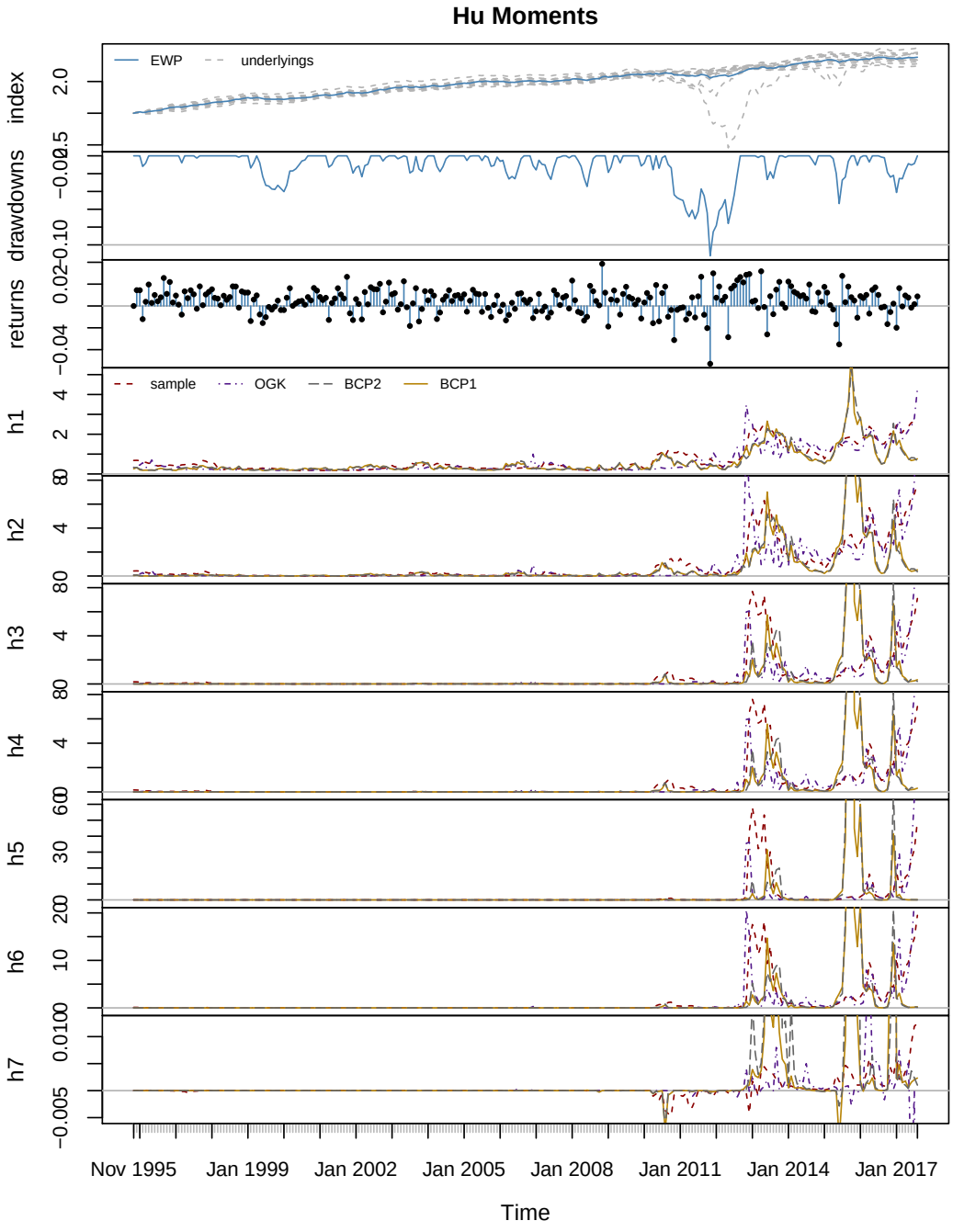


Figure 11.6: The Hu Moments for the pure bond universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

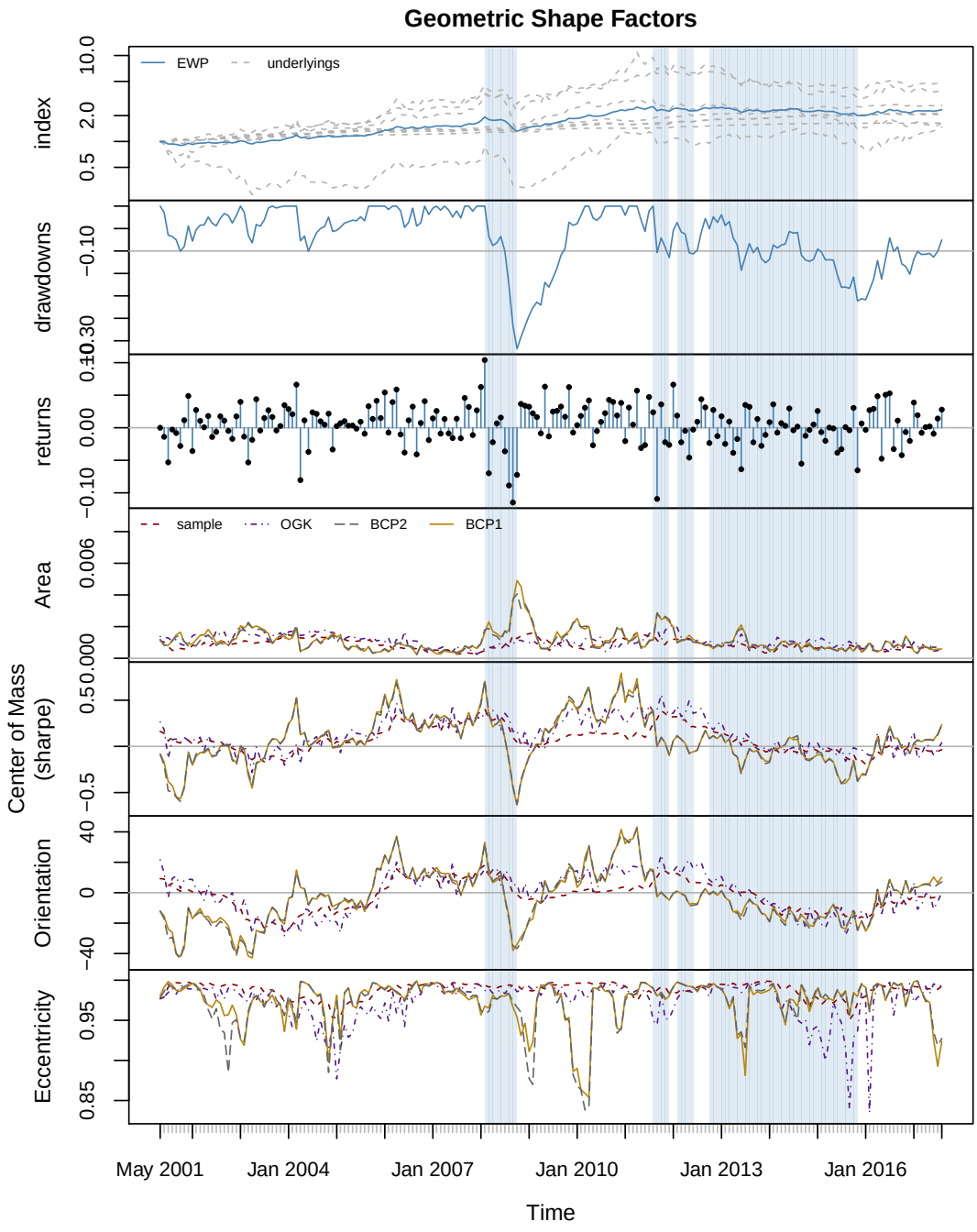


Figure 11.7: The GSF for the European bonds and precious metals universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

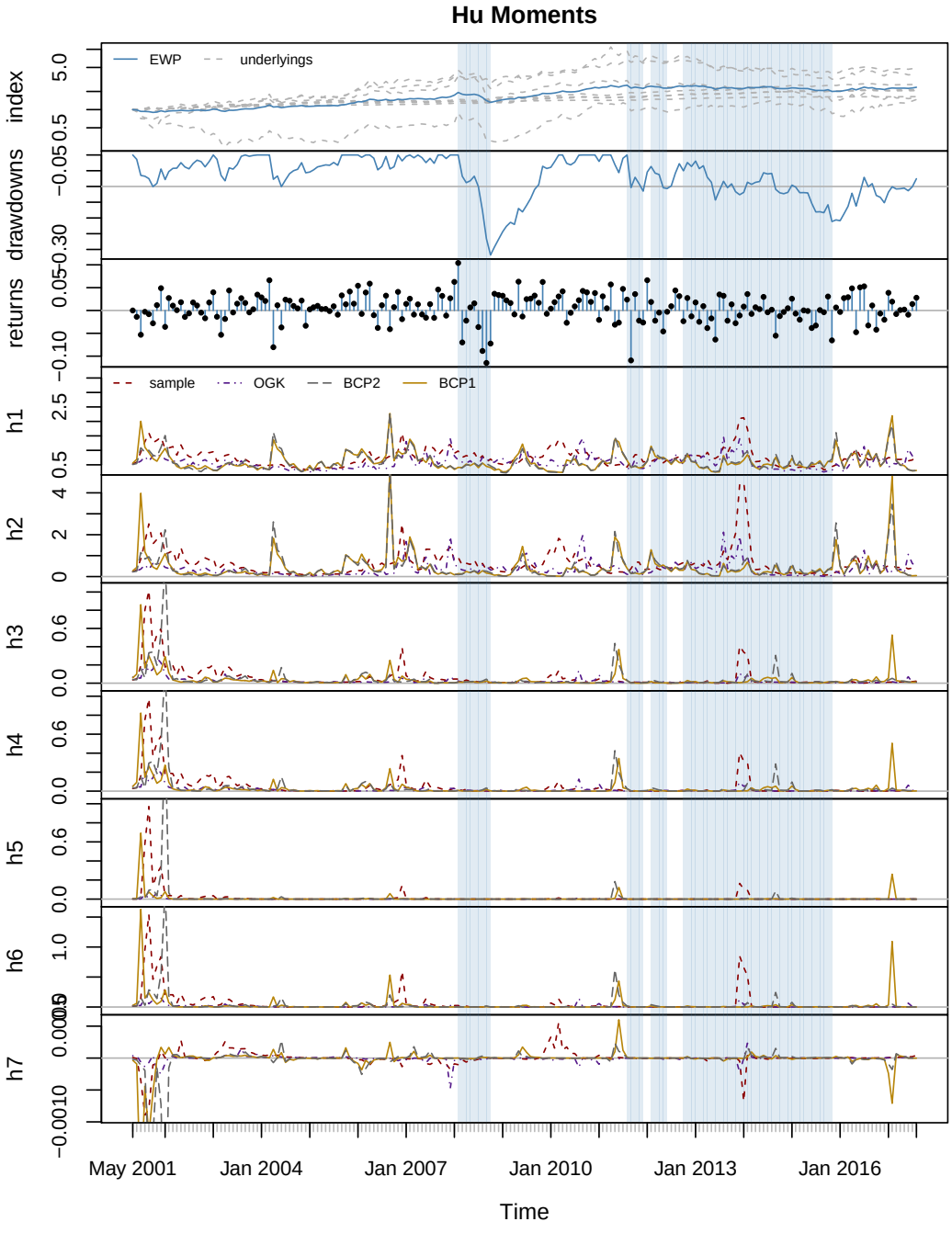


Figure 11.8: The Hu Moments for the European bonds and precious metals universe. The shaded areas show the times where the BCP estimation of the trend is negative. See also Appendix A for more information about the underlying data.

# 12 Summary and Outlook

In Chapter 9 it was shown how the feasible set can be calculated by using different estimators. The focus was on sample and robust estimators, and estimators based on a BCP analysis. For the robust case the OGK estimator as defined by Maronna and Zamar [2002] was used. For the BCP case two new estimators were defined. The first one calculates the covariance matrix through the univariate variance estimator of the BCP and uses the concepts of the OGK estimator to transform the covariance matrix such that it becomes positive semidefinite and approximately affine equivariant. The second one mixes the sample correlations with the univariate variance estimator of the BCP to calculate the covariance matrix.

In Chapter 10 it was shown how the feasible set can be represented as an image in order to calculate its image moments and how these moments define the GSF. The GSF are the parameters that define an ellipse that has the same area, centre, orientation and eccentricity as the feasible set. It was additionally shown how the GSF can be calculated directly from the polygon that describes the feasible set to get better accuracy.

In Chapter 11 the GSF and Hu moments were calculated on a rolling window for different investment universes. It is important to note that the two BCP estimators deliver almost the same results. The second estimator (Eq. (9.24)) is computationally much more efficient than the first estimator (Eq. (9.23)).

By combining the field of image analysis with portfolio optimization the dynamics of the investment universe can be visualized. This offers very valuable insights into the behaviour of a given investment universe and allows to get a feel for the current state of that universe. The GSF, image moments and Hu moments can be calculated for many other estimation methods of the feasible set and it would be of interest to compare even more of them. No work has been done so far to exploit this information in portfolio design. A first proposal will be presented in Chapter 15 based on the orientation. The orientation is used because it can be associated with a geometrical meaning and is therefore interpretable. So are the other GSF. It would be of interest to come up with accessible interpretations for the image moments or the Hu moments as well.

# Part III

# Stability in Portfolio Design

# 13 Introduction

In the first and second part of this thesis risk measures for the univariate and multivariate case were derived and discussed. The last part of the thesis deals with the application of these measures. More precisely on how they are used to construct stable portfolios that react on changes within the investment universe. The notion of stability will be discussed in Section 13.1.

The proposed models follow the philosophy of a risk-averse and long-term investor. The models are not designed to earn a huge profit in a short amount of time but to realize what the market realistically offers from a long-term perspective while reducing the risks. A portfolio is defined through its allocation of the available capital to the underlying assets of the investment universe (the weights). Following the long-term perspective monthly data with a three year look-back window were used for the calculation of the weights.

In Chapter 14 the univariate BCP risk measures (see Chapter 5) are used to calculate the weights for an equity universe that consists of the European sector indices. For that the concept of signal portfolios is introduced. Any statistical analysis can be used to calculate signals between 0 and 1 that indicate for every asset within the investment universe on whether to rather buy (1) or sell (0) an asset. Based on these signals the weights for the assets within the investment universe can be calculated. There are various possibilities on how to calculate the weights. For example the assets that show high signals could be weighted higher than the assets that show low signals. Following the risk-averse philosophy the proposed model creates a dynamic hedge. For every sector the signals are calculated based on a BCP analysis. If a signal suggests to lower the investment into a certain sector the freed capital is moved to a cash account that offers zero interest rate. This is equivalent to setting up a hedge (under the assumption that the hedge is for free). Since the signals will have values around 0.5 the natural benchmark is a constant hedge where 50% of the capital is always hedged.

In Chapter 15 a different investment universe is considered that consists of bonds and precious metals. The weights are calculated through Markowitz portfolio optimization

(see Chapter 9). But instead of having a model that stays constant over time, the model is dynamically adapted to the current shape of the feasible set. The orientation of the GSF (see Section 10.5) is used to choose a portfolio within the feasible set. To have good control about where this portfolio is situated within the feasible set the lambda-2-portfolio (L2P) is introduced. Additionally a multi objective approach is considered. For the L2P the dynamic Markowitz portfolio using different estimators are compared.

In Chapter 16 the BCP-Signal-Portfolio (BSP) and the dynamic L2P are combined. The freed capital of the BSP is not used for a hedge but moved to the L2P. The resulting portfolio is suited for risk-averse and long-term investors that need to always be fully invested and cannot use leverage. Also the portfolio is very well diversified through three different asset classes.

## 13.1 Stability

The notion of financial market stability was introduced by Wuertz [2010]. It is based on the idea that financial markets and economies can be in a state of steady development with constant growth (stable) or in a state of unusual behaviour showing stagnation or deterioration (unstable). The method chosen to identify such states depends on one's individual understanding of the driver of such states.

One could assume that a market is subject to sporadic outliers. Then a principal component outlier analysis that measures the degree of outliers could be an indication for stable or unstable times. Another approach for such markets would be to use robust estimators that smooth the outliers. One could also understand the returns of a market as a superposition of harmonic oscillators. A Fourier transformation or wavelet analysis could identify the dominant modes. The more equally the modes are distributed the more stable the market is considered. There are many other models thinkable.

This thesis follows two different models. The first model assumes that any observation within a market can be explained by the outcome of an underlying generating statistical process. And that this process can change at any point in time. Times are considered as stable if the probability for such a change is low. The BCP method offers a way to measure these probabilities. The model of changing dynamics is further explained in Section 14.1. The second model assumes that changes within the shape of an investment universe are an indication for stable or unstable times. The GSF offer a way to measure such changes. The model of shape shifts is further explained in Section 15.1.

Generally stability is associated with a model of change. For that the driver of change

has to be identified and a statistical model defined that can measure the degree on how much the driver is activated. Based on these measurements statistical properties can be calculated and/or rules defined to make investment decisions.

# 14 Signal Portfolios

Today a lot of investment decisions are still made on a discretionary basis. This means that an investor informs himself about an asset through various sources to make a decision on how much to buy from that asset. This is a rather univariate approach. Signal portfolios work in a similar way. Only that the decisions are made on a completely quantitative basis. A model is defined to generate signals between 0 and 1 that describe the degree of investment in the asset under consideration. If various assets are analysed those assets are integrated into one portfolio that holds the investment capital. The portfolio presented in this chapter is strictly univariate and doesn't need any optimization routines as it is the case for Markowitz portfolios.

## 14.1 The Model of Changing Dynamics

The input for this model are financial returns of a given asset. The model is based on the assumption that these observations are generated by an underlying dynamic that can be described through a probability distribution (e.g. a normal distribution) which itself can be described by a set of parameters (e.g. the mean and the variance). Further the model assumes that this dynamic is not necessarily constant over time. Times are considered stable when the underlying dynamic (the model and/or parameters) is unlikely to change.

Many models assume that financial returns can be described through a random walk where the underlying dynamic is constant over time. In that case the observations would have a constant expected return with variations around that trend that are all generated by the random nature of the underlying dynamic. Every drawdown or crisis (large drawdowns) would solely be a result of that randomness and couldn't be associated to a fundamental change within the market. In such a scenario there is not much need to act. For the long run it would just be a decision on whether one is satisfied with the return compensation given the risk.

The model of changing dynamics assumes that an economic crisis is not necessarily

only an expression of the random nature of the underlying dynamic. It is assumed that before, during and after a crisis the underlying dynamic itself can be subject to change. It can go as far as to the point where any observation is considered to have its own dynamic. But that for stable times these dynamics are almost the same while for unstable times they could possibly be very different. Within this framework a financial time series could be understood as a sequence of different random walks instead of one single random walk. If the model measures a high likelihood for a changing dynamic an adjustment of tactics might make sense.

The BCP model measures the likelihood of change for the mean (performance) and the variance (risk) of a normal distribution and therefore is a suitable mathematical framework for the model of changing dynamics. To explain the stylized facts of financial returns the approach is not to find a distribution that explains the returns as a whole. But to explain the financial returns through piecewise normal distributions that have a different performance and risk if they cannot be explained through a normal distribution as a whole. If new observations and older observations cannot be explained with the same dynamic the BCP model is quick in updating the current performance and risk based on these new observations. The lag associated with sliding windows is therefore minimized. Note that for the BCP model drawdowns are not necessarily a sign for a crisis. Only if these observations cannot be explained anymore with the same dynamic as older observations.

One might ask about the fundamental reasons that justify such a model. Let's assume a world where all the market participants always follow the same routines. They wouldn't panic during bearish markets as much as they wouldn't get excited about bullish markets. It could be assumed that such a market could indeed be explained by a constant dynamic and that all the ups and downs are just a result of the randomness of the underlying process. The model of changing dynamics assumes that this is not necessarily true and that the market participants behaviour might dramatically change based on the movements of the market. The assumption is that for such times the market doesn't follow the same rules anymore and that therefore also the market cannot be described anymore with the same underlying dynamic.

## 14.2 Signal Calculation

The signal calculation follows three steps. First a statistical analysis delivers the results of the model of change. In this case it is the BCP analysis that describes the model of

changing dynamics. In the second step the results of the analysis are used to define an indicator. In the last step the indicator is used to generate the signals in the range of 0 to 1.

### 14.2.1 Analysis

At any point in time $t$ the BCP analysis (see also Section 5.4) is calculated by using the last 36 monthly returns (look-back window of $l = 36$ months). The matrix $R_t = \{\boldsymbol{r}_{t,1}, \boldsymbol{r}_{t,2}, \ldots, \boldsymbol{r}_{t,n}\}$ holds the returns of the lookback window at time $t$ for the $n$ assets under consideration. The vector $\boldsymbol{r}_{t,k} = \{r_{t-l+1}, r_{t-l+2}, \ldots, r_t\}$ holds the returns of the look-back window for asset $k$ at time $t$.

The result will be at any point in time $t$ the matrices $M_t = \{\boldsymbol{m}_{t,1}, \boldsymbol{m}_{t,2}, \ldots, \boldsymbol{m}_{t,n}\}$, $S_t = \{\boldsymbol{s}_{t,1}, \boldsymbol{s}_{t,2}, \ldots, \boldsymbol{s}_{t,n}\}$ and $P_t = \{\boldsymbol{p}_{t,1}, \boldsymbol{p}_{t,2}, \ldots, \boldsymbol{p}_{t,n}\}$ that hold the estimations of the posterior means (see Eq. (3.40)), standard deviations (see Eq. (3.41)) and probabilities (see Eq. (3.42)). The vectors $\boldsymbol{m}_{t,k} = \{m_{t-l+1}, m_{t-l+2}, \ldots, m_t\}$, $\boldsymbol{s}_{t,k} = \{s_{t-l+1}, s_{t-l+2}, \ldots, s_t\}$ and $\boldsymbol{p}_{t,k} = \{p_{t-l+1}, p_{t-l+2}, \ldots, p_t\}$ hold the posterior means, standard deviations and probabilities of the look-back window for asset $k$ at time $t$.

Using the definition of the BCP estimator in Section 9.3.3 the estimations of the mean, the standard deviation and the structural break probability at time $t$ for asset $k$ can be calculated as

$$\mu_{t,k} = \hat{\mu}(\boldsymbol{r}_{t,k}) = \mathrm{EMA}_{\lambda_\mu}(\boldsymbol{m}_{t,k}) \tag{14.1}$$

$$\sigma_{t,k} = \hat{\sigma}(\boldsymbol{r}_{t,k}) = \mathrm{EMA}_{\lambda_\sigma}(\boldsymbol{s}_{t,k}) \tag{14.2}$$

$$p_{t,k} = \hat{p}(\boldsymbol{r}_{t,k}) = \mathrm{EMA}_{\lambda_p}(\boldsymbol{p}_{t,k}) \tag{14.3}$$

### 14.2.2 Indicators

The indicator follows the same rule as the BCP estimator and is calculated as the Sharpe ratio of the posterior means and variances as

$$d_{t,k} = \hat{d}(\boldsymbol{r}_{t,k}) = \mathrm{EMA}_{\lambda_d}\left(\frac{\boldsymbol{m}_{t,k}}{\boldsymbol{s}_{t,k}}\right) \tag{14.4}$$

This leads to an indicator that is high if the posterior means are high and the posterior variances are low. Note that the posterior probabilities are not used directly for the calculation of the indicator since they are already used to get recent estimations of the

mean and the standard deviation of the asset under consideration.

The indicator values can theoretically have arbitrarily high or low values. The question arises for which values the asset under consideration should be considered favourable to invest. Often this is solved through setting a threshold. A natural threshold would for example be 0. Which can be described as the case where the posterior means are rather positive. The problem with thresholds is that they can be tailored to get favourable results for the backtest. One has to be careful that a chosen threshold does not dominate the results. Also a threshold that works good for one asset might not work for another. The models presented here are always designed such that a given parameter does not have to be changed for different assets. And if so the model itself should find a reasonable parameter without using any information of the future in a backtest.

Another problem with thresholds is the practical aspect. Setting thresholds often leads to binary models in the sense that an investment is made into an asset when the signal is above the threshold (signal of 1) and no investment is made when the signal is below the threshold (signal of 0). But for many investors it is not possible to withdraw the total investment from an asset at once.

### 14.2.3 Signals

To circumvent the problems associated with the introduction of thresholds the signals are calculated through a transformation of the signals into a range between 0 and 1. For that the logistic function is used as

$$e_{t,k} = \frac{1}{1 + \exp\left(-c \cdot d_{t,k}\right)} \tag{14.5}$$

where $c$ controls the steepness of the curve. This leads to a non-binary signal that tends to 1 if the indicator has a high positive value and to 0 if the indicator has a low negative value. An indicator value of 0 will result in a signal value of 0.5.

The logistic function belongs to the family of S-shaped sigmoid functions. The logistic function is for example used in machine learning for logistic regression or in the field of neural networks to transform signals into a specific range (as it is done in Eq. (14.5)). Fig. 14.1 shows the shape of the logistic function and the impact of the steepness parameter $c$. If the steepness parameter gets large the signals tend towards a binary model.

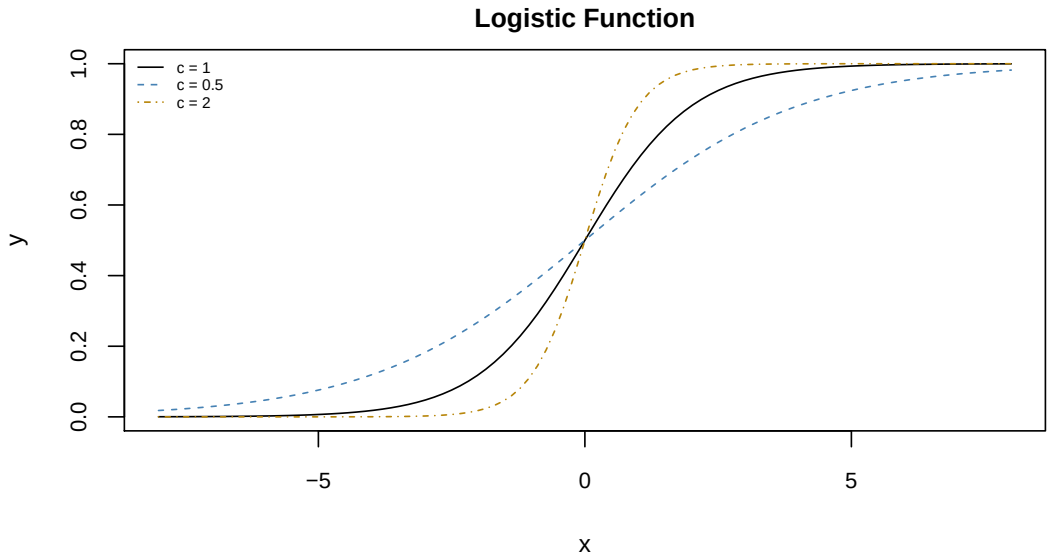


Figure 14.1: The logistic function for different steepness parameters $c$.

### 14.2.4 Application

The signals are calculated for every asset within the investment universe individually. Fig. 14.2 and Table 14.1 show the backtest of the signals for just one index. It is the same index (the European equity index) as analysed in Fig. 6.11. If the signal is 1 one would be fully invested into the market and if the signal is 0 the investment would be fully hedged. For a signal of for example 0.8 the investment would be hedged to 20%. Note that it is important on how the signals are applied for the backtest. A signal on any given date has to be applied on the return of the next date (out of sample).

| | Performance | Volatility | Sharpe | Max. Drawdown |
|---|---|---|---|---|
| BCP-Strategy | 8.13 | 9.84 | 0.83 | 20.38 |
| EQEU.EUR (50% Hedge) | 3.67 | 7.86 | 0.47 | 32.43 |
| EQEU.EUR | 7.47 | 15.66 | 0.48 | 54.34 |

Table 14.1: Key figures for the indices shown in Fig. 14.2. The performance and the volatility are the annualized return and volatility, respectively.

For the backtest the evaluations of Eq. (14.1) and Eq. (14.2) are identical to the evaluations of Eq. (9.23) to calculate the feasible set. The parameters used are therefore the same (see also Section 9.3.4). For the lambdas this is $\lambda_\mu = \lambda_\sigma = \lambda_d = 0.2$ and

$\lambda_p = 0.8$. Note that the posterior probabilities ($p_{t,k}$) are not used directly to calculate the signals (as explained in Section 14.2.2). They are just shown in Fig. 14.2 as additional information. The parameters to calculate the BCP analyses were set by first calibrating the model on the data of the look-back window (see Eq. (6.6) and Eq. (6.8)) using $v = 2$ and $v_2 = 1.5$ to make the priors more flat. The breakpoint parameters $\alpha = 4$ and $\beta = 16$ were chosen globally such that the expected breakpoint probability is 20%. The indicator values are an estimation of the Sharpe ratio of the asset under consideration. For financial returns this is usually always somewhere between $-1$ and 1. Higher Sharpe ratios would be preferable but the market is usually not offering it. The steepness parameter $c$ of the logistic function is globally set to 4 for positive indicator values and 8 for negative indicator values. The reason for this is to be more risk-averse. Which means that the signals tend to zero faster when the indicator gets negative than they tend to 1 when the indicator gets positive. These values are somewhat arbitrary.

Note that these parameters are the same for any calculations of the feasible set or for any calculation of the signals for any asset. This is in line with the philosophy of either using global parameters which are the same for any data under consideration or letting the model decide the parameters by itself without using information from the future.

Applying the signals represents a dynamic hedge. This can be realized as a portfolio that consists of two assets. The first one the index itself and the second one an investment into cash with zero interest rate. If the signal is 0.8 then 20% of the investment would be moved into the cash account. This is equivalent of setting up a hedge (assuming no costs). Therefore the natural benchmark is the EWP of the index and the cash account. Which is equivalent to a 50% hedge. Note that in practice it is often assumed that such trading signals should outperform the index. In my personal opinion this is a misconception which leads to exaggerated expectations. This might lead to models that are highly optimized on the past to match these expectations.

The backtest shows that the strategy outperforms the index; but this is only due to the extreme drawdowns around 2002 and 2008. For other periods where there are no extreme drawdowns one can only expect to receive a more favourable portfolio than the 50% hedge. The benchmark can be adapted by multiplying the signals with a constant factor. If that factor would be 0.5 then the natural benchmark would become the 25% hedge. If the factor is for example 2 then the benchmark will indeed be the index. But applying signals with a factor of 2 means that the exposure needs to be leveraged if the signals become bigger than 1.

The constantly 50% hedged index has obviously approximately the same Sharpe ratio

as the unhedged index (see Table 14.1). But the risk profile is more favourable since the drawdowns do only have half the size. The price to pay for a constant hedge is straight forward. The performance is only half that of the unhedged index and so is the risk.

The BCP strategy (dynamic hedge) has a much better Sharpe ratio. And the risk profile is similar to the 50% hedge. More details can be seen in Fig. 14.2. The risk of a dynamic hedge is that the signals might not react to the formation of a crisis fast enough and therefore realize the full drawdowns. This can be seen around the mid of 1998. But generally the dynamic hedge will benefit from upward trends and secure against severe downward trends. Working with higher data frequencies (e.g. weekly or daily) might reduce the risk of missing the formation of a crisis. But also the nature of the data would fundamentally change since the higher the frequency the less the data is normally distributed. For this thesis monthly data was used to demonstrate the fundamental concepts.

## 14.3 Portfolio Design

### 14.3.1 The BCP-Signal-Portfolio

The index that was used for the backtest in Fig. 14.2 and Table 14.1 is a benchmark for the European equity market which can be divided into 19 industry sectors. It is the same universe (European equity universe) as analysed in Fig. 11.3 and Fig. 11.4 through the GSF. To construct the signal portfolio the BCP signals are calculated for every sector in exactly the same way as they have been calculated for the total index in Section 14.2. This portfolio will be referenced as the BCP-Signal-Portfolio (BSP) for the remainder of this thesis.

Let $W = (w_{t,k})$ be the weights matrix where $w_{t,k}$ describes the weight of asset $k$ at time $t$. The weights are constrained such that the sum of the weights at any point in time shall be between 0 and 1 and the individual weights are not allowed to be smaller than 0. This corresponds to a portfolio where leverage or short selling is not allowed but hedging is. This is for example the basic setup for a pension fund.

To use the signals $e_{t,k}$ an upper bound for any asset within the investment universe is introduced. The upper bound is defined to be 100% divided by the number of assets. For the 19 sectors this is $w_{max} = 100/n = 5.26\%$. The weights for any asset $k$ at time $t$ are now defined as

$$w_{t,k} = e_{t,k} \cdot w_{max} \tag{14.6}$$

which satisfies the constraints. For any asset within the investment universe the signals define the grade of investment which is between 0 and $w_{max}$. If all signals are 1 the sum of the weights will also be 1 and there is no hedging involved. Vice versa the same is true if all signals are 0 and the invested capital will be fully hedged. The case where all signals are 0.5 corresponds to the EWP of the individual 50% hedged assets.

| | Performance | Volatility | Sharpe | Max. Drawdown |
|---|---|---|---|---|
| BCP-Signal-Portfolio | 7.26 | 8.78 | 0.83 | 18.55 |
| EWP (50% Hedge) | 3.81 | 7.88 | 0.48 | 32.82 |
| EWP | 7.77 | 15.69 | 0.50 | 54.86 |

Table 14.2: Key figures for the indices shown in Fig. 14.3. The EWP corresponds to the EWP European Sectors. The performance and the volatility are the annualized return and volatility, respectively. For the 50% hedge the sector indices are hedged separatelly.

The results of the backtest (out of sample) are shown in Fig. 14.3 and Table 14.2. The results are similar as calculating the signals for the total index (Fig. 14.2 and Table 14.1). The BCP portfolio shows a higher Sharpe ratio and a more attractive risk profile.

The reason why the sector approach might be preferred to the total index approach is that the signals are diversified. Diversification within the assets of an investment universe is considered an important investment concept. If one asset within the investment universe shows a singular extreme event that the other assets do not show then the impact of that event is the smaller the higher the diversification. In a similar way it can be argued that diversification is also important for signal calculation. If one signal is wrong then the impact of that wrong signal is the smaller the more signals are used. In the same way as any discretionary investment advise might turn out in an unpreferable result a signal based on a quantitative analysis might result in an unpreferable result. If only one signal for one index is calculated the risk for such an outcome is much higher than for the case where several signals are calculated on several indices.

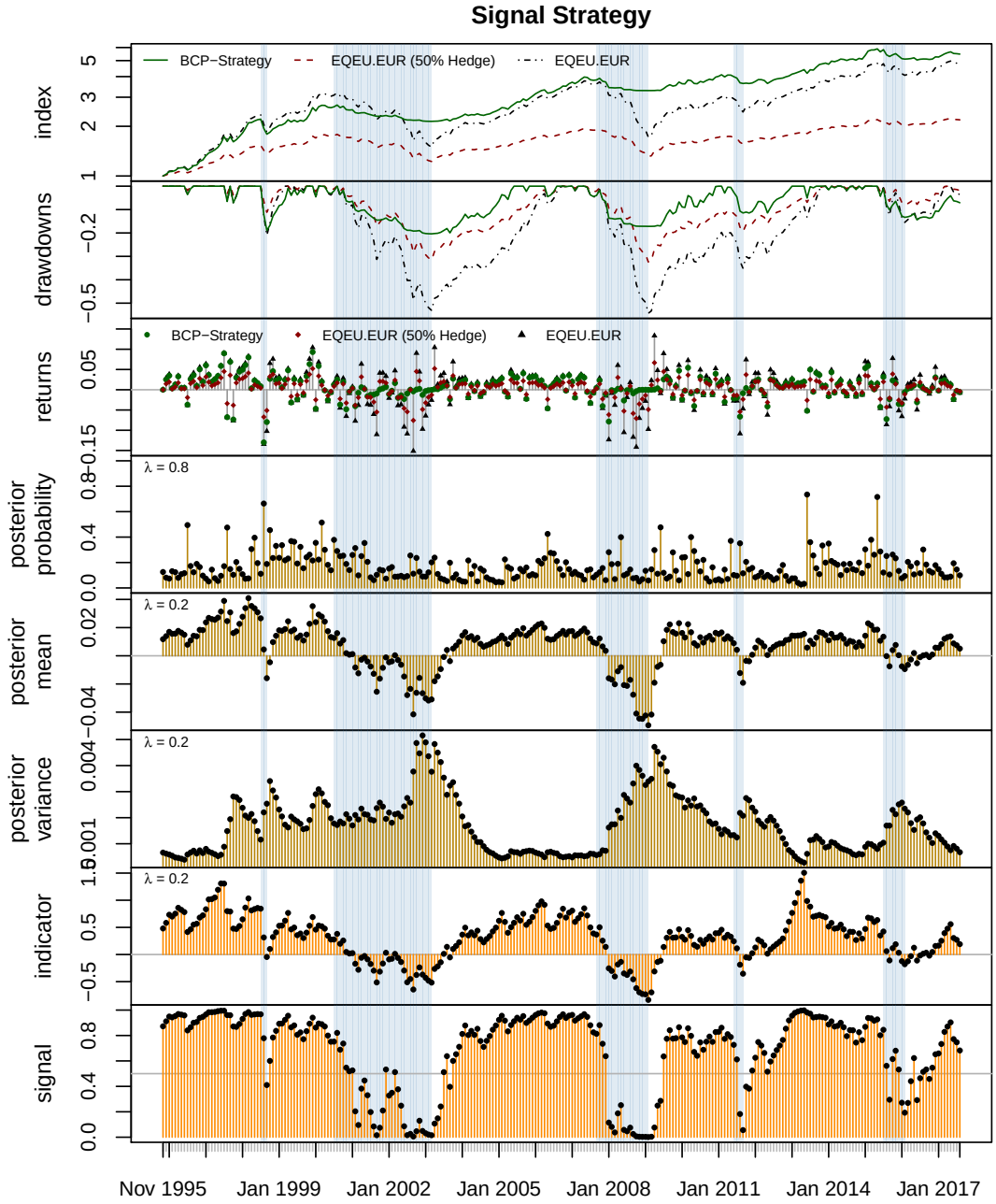


Figure 14.2: The BCP signals applied to the European equity index. The posterior means show the evaluation of Eq. (14.1), the posterior variances the evaluation of Eq. (14.2) and the posterior probabilities the evaluation of Eq. (14.3). The indicators the evaluation of Eq. (14.4) and the signals the evaluation of Eq. (14.5). The shaded areas show the times where the BCP estimation of the trend for the European equity index is negative. See also Appendix A for more information about the data.

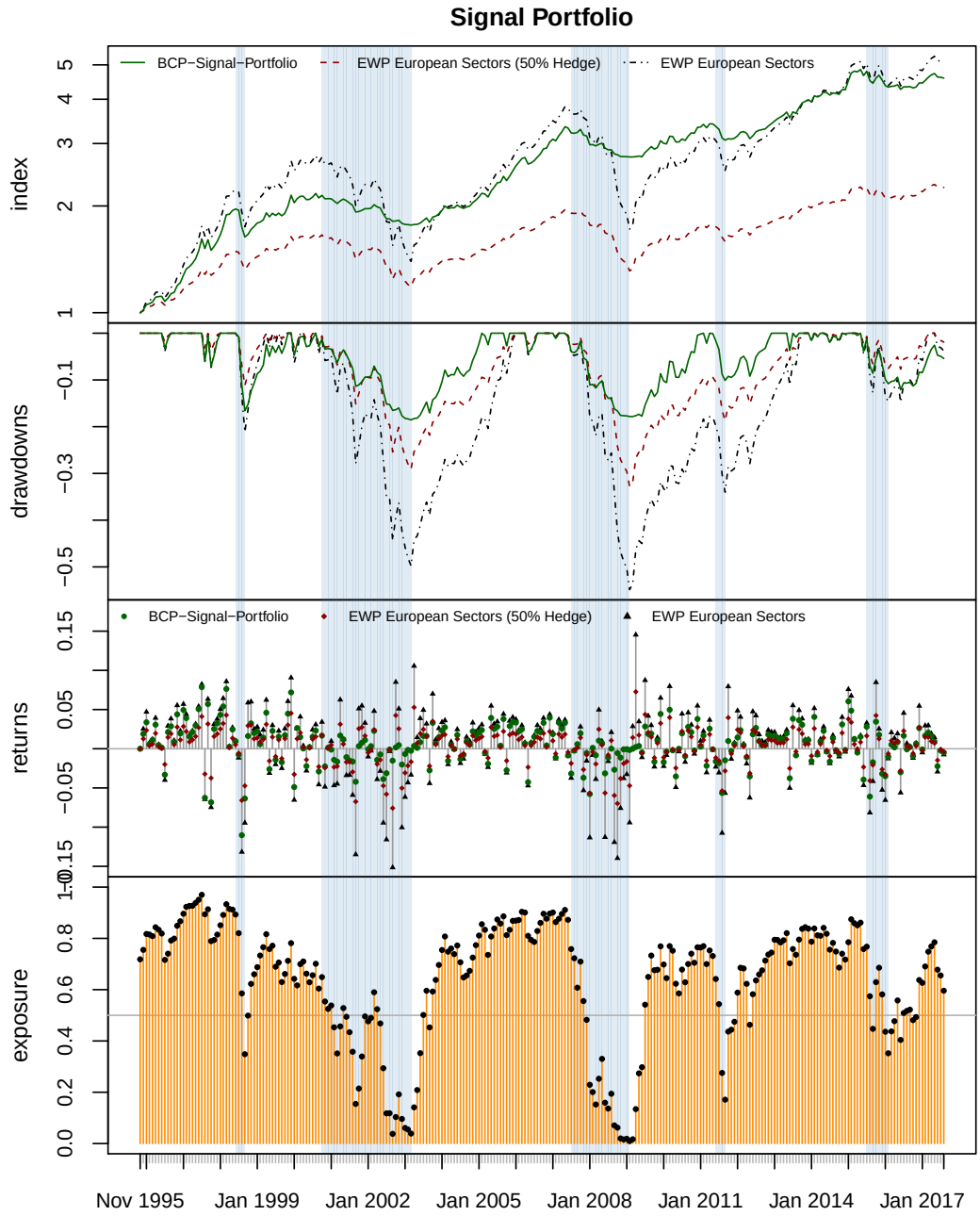


Figure 14.3: The BSP for the European equity universe (19 industry sectors). The exposure shows the amount of capital that is invested. It is the sum of the weights at any point in time. The shaded areas show the times where the BCP estimation of the trend for the EWP is negative. See also Appendix A for more information about the data.

# 15 Dynamic Markowitz Portfolios

Markowitz portfolios [Markowitz, 1952] were introduced in Chapter 9. Those portfolios are characterized through an objective function that is minimized (e.g. the variance) under certain constraints (e.g. long-only, no short-selling and/or fully invested). Usually the objective and constraints are not modified over time. Dynamic Markowitz portfolios are doing exactly that. Based on the current market situation the objective and/or constraints are adapted.

## 15.1 The Model of Shape Shifts

The feasible set is a visual representation of the investment universe and pictures the performance and risks of the individual assets and the relations of these assets to each other. In that sense it is a multivariate view that can reveal important information about the current state of an investment universe. The feasible set can be described through an ellipse that has the same area, centre, orientation and eccentricity as the feasible set. Those key figures are called the geometric shape factors (GSF) and are a subgroup of the morphological shape factors which describe the image moments of the feasible set.

Variations in those moments might be an indication about fundamental changes within the investment universe and an adjustment of allocation tactics might make sense. While most of the moments do not have a geometrical meaning and are therefore difficult to interpret the GSF do. The model that will be introduced in the further course of this thesis uses the orientation of the GSF which describes the direction of the risk premium.

The shape of the feasible set is often assumed to look like in Fig. 10.2. It is assumed that there is a risk premium from the assets that have a low variance towards the assets with a higher variance. For such cases the orientation is positive. It is reasonable to think that this is generally true in the long run. The GSF (see Fig. 11.1) show that at least for time periods of three years this is very often not true. The orientation gets sometimes clearly negative which means that the risk premium is negative.

During times where the risk premium is negative it might make sense to withdraw

from the assets that are assumed to offer a risk premium and move into the assets that generally have a lower risk. To achieve this a dynamic Markowitz approach is pursued. The more the risk premium is positive the more the target portfolio will be allowed to move away from the minimum risk portfolio (MVP) towards higher returns while always making sure that there is a certain degree of diversification.

## 15.2 Portfolio Design

The investment universe for this part of the thesis consists of the European bonds and precious metals universe (base assets) as analysed in Fig. 11.7 and Fig. 11.8 through the GSF. The GSF and the weights of the portfolio were calculated at any point in time by using the last 36 monthly returns. The weights are always applied out of sample. The GSF are the same as shown in Fig. 11.7.

### 15.2.1 The Lambda-2-Portfolio

The Lambda-2-Portfolio (L2P) is defined as

$$\begin{aligned} &\min_{\boldsymbol{w}} \quad f(\boldsymbol{w}) \\ &\text{s.t.} \\ &\boldsymbol{\mu}'\boldsymbol{w} = \lambda_1 \mu_{upper} + (1-\lambda_1)\mu_{lower} \\ &\boldsymbol{w}'\Sigma\boldsymbol{w} = \lambda_2 \sigma_{upper} + (1-\lambda_2)\sigma_{lower} \\ &\mathbf{1}'\boldsymbol{w} = 1 \\ &0 \leq w_i \leq 1 \end{aligned} \tag{15.1}$$

where $0 \leq \lambda_{1,2} \leq 1$. $\boldsymbol{w}$ describes the weights of the assets within the investment universe, $\boldsymbol{\mu}$ the estimated performance of the assets and $\Sigma$ the covariance matrix. The last two constraints define a portfolio that has to be fully invested and forbids leverage and short-selling. This are constraints often encountered in practice. The first constraint defines the target return ($\boldsymbol{\mu}'\boldsymbol{w}$) and the second constraint the target risk ($\boldsymbol{w}'\Sigma\boldsymbol{w}$). One has to be careful to choose $\mu_{upper,lower}$ and $\sigma_{upper,lower}$ such that the portfolio is feasible (it is situated within the feasible set). Then this setup allows to define any point within the feasible set.

An alternative setup through a multi objective approach will be discussed in Sec-

tion 15.2.2. The advantage of this setup is that the location of the target portfolio based on the lambdas is obvious. If $\mu_{upper} = \mu_{max}$, $\mu_{lower} = \mu_{min}$, $\sigma_{upper} = \sigma_{max}$, $\sigma_{lower} = \sigma_{min}$ and $\lambda_1 = \lambda_2 = 0.5$ the portfolio will be in the middle of the feasible maximum and minimum return, and the feasible maximum and minimum variance. This is different if the lambdas are used for a multi objective function. The location of the target portfolio is in principle defined but also somewhat arbitrary (see Section 15.2.2).

The second constraint is quadratic. To solve the portfolio a suitable solver has to be used. For `R` the packages **Rsolnp** [Ghalanos and Theussl, 2015] or **Rdonlp2** [Wuertz et al., 2014] are suitable. Those solvers can deal with any kind of non-linear objective functions and/or constraints. For the backtest $f(\boldsymbol{w}) = 1$ was chosen. Which means that the objective function does not have any particular meaning; the portfolio is only defined through the constraints. An optimizer is still needed since the portfolio does not have a unique solution. It would be possible to define any objective function; for example the diversification ($f(\boldsymbol{w}) = \text{var}(\boldsymbol{w})$). Since the backtest results are very much identical the solution that is numerically less complex was chosen. The reason why the results are very much identical is most likely that the optimization starts with the weights of the EWP; which has globally the best diversification.

To calculate the dynamic L2P portfolio the lower return is globally set to be the return of the minimum variance portfolio (MVP) ($\mu_{lower} = \mu_{MVP}$) and the upper return to be the maximum return ($\mu_{upper} = \mu_{max}$). Of course other solutions are possible. The parameters $\lambda_{1,2}$ and $\sigma_{upper,lower}$ are chosen based on the orientation of the feasible set. The orientation $o$ is naturally bounded between $-90°$ and $90°$. This is an advantage since we do not have to care about the magnitude of this key figure. It means that $90°$ is a big positive value and $-90°$ a small negative value. For unbounded key figures which lie between $-\infty$ and $\infty$ it is not always obvious which values can be considered big or small. If the orientation is divided by 90 the values will be bounded by $-1$ and 1.

In the following the algorithm to calculate the dynamic L2P portfolio is summarized. Let $W = (w_{t,k})$ be the weights matrix where $w_{t,k}$ describes the weight of asset $k$ at time $t$ and the vector $\boldsymbol{w}_t = \{w_1, w_2, \ldots, w_n\}$ holds the weights at time $t$ for all assets. At any point in time $t$ the weights $\boldsymbol{w}_t$ can be calculated as

1. Calculate the orientation $\bar{o} = o/90$ of the feasible set. Calculate $\lambda_1 = f_1(\bar{o})$ and $\lambda_2 = f_2(\bar{o})$.

2. Calculate the expected return of the minimum variance portfolio ($\mu_{MVP}$) and the maximum return portfolio ($\mu_{max} = \max(\boldsymbol{\mu})$).

3. Define the target return: $\boldsymbol{\mu}'\boldsymbol{w}_t = \lambda_1\mu_{max} + (1-\lambda_1)\mu_{MVP}$.

4. Calculate the portfolio that has the smallest risk given the target return. Use the risk of that portfolio as the lower risk ($\sigma_{lower}$). Calculate the portfolio that has the highest diversification given the target return. Use the risk of that portfolio as the upper risk ($\sigma_{upper}$).

5. Define the target risk: $\boldsymbol{w}_t'\Sigma\boldsymbol{w}_t = \lambda_2\sigma_{upper} + (1-\lambda_2)\sigma_{lower}$.

6. Calculate $\boldsymbol{w}_t$ using Eq. (15.1).

For step one $f_1(\bar{o})$ and $f_2(\bar{o})$ have to be defined. Generally the following function is used:

$$f(\bar{o}, c, l, u) = \min(u, \max(l, g(\bar{o})c)) \tag{15.2}$$

where $c$ scales the orientation, $u$ defines an upper bound for the lambdas and $l$ a lower bound. The function $g(\cdot)$ defines how the orientation is used. For the backtest the following instances of $f(\cdot)$ were used:

$$\lambda_1 = f_1(\bar{o}) = f(\bar{o}, c = 0.25, l = 0.00, u = 1.00) = \min(u, \max(l, |\bar{o}|c)) \tag{15.3}$$

$$\lambda_2 = f_2(\bar{o}) = f(\bar{o}, c = 4.00, l = 0.10, u = 1.00) = \min(u, \max(l, |\bar{o}|c))I(\bar{o} > 0) \tag{15.4}$$

where $I$ is the indicator function. Generally the parameters for $\lambda_1$ and $\lambda_2$ and the function $g(\cdot)$ can be chosen in any other way. $\lambda_1$ was designed such that the portfolio will generally be more risk-averse and not affected by the orientation jump at around $\pm 90°$. $\lambda_2$ was designed such that the portfolio always enforces a minimum amount of diversification and goes to the maximum possible amount of diversification quickly if the orientation gets positive but not negative.

In step three a target return is chosen that lies on the efficient frontier. $\lambda_1$ controls the distance from the MVP in means of the target return. Choosing $c = 0.25$ is a risk-averse approach. It means that the distance from the MVP is limited. The absolute value of $\bar{o}$ is used to make the target portfolio independent on the sign of the orientation. Whether the orientation has a large positive or negative value the target portfolio will always be one that looks for more return than the MVP. As explained in Section 10.3 the orientation can suddenly jump from $90°$ to $-90°$ and vice versa. Using the absolute value of the orientation makes sure that this does not influence the target portfolio. The reason to retreat towards the MVP if the orientation is around 0 is that in that case the

efficient frontier could switch sides. If the orientation becomes negative it usually means that the risk premium gets lost. This happens if the assets that formerly showed a high performance and risk (e.g. the precious metals) suddenly show lower performances (while keeping the risk) than the assets that showed formerly low performances and risks (e.g. the bonds). It is usually a sign of a crisis within the investment universe that results in the formation of a completely new state. At the point where the last formerly high performing asset shows not the highest performance anymore the efficient frontier will switch side. If a target portfolio on the upper part of the efficient frontier is chosen it will be greatly affected by this change. In contrary the MVP (and any portfolios that have the same target return as the MVP) will not be greatly affected by this change. Therefore $\lambda_1$ is used to retreat to a target portfolio that is robust against a change of the sign of the risk premium and the formation of a new state within the investment universe. Since the absolute value of the orientation is used and the orientation is scaled to be smaller than 1 choosing $l = 0.00$ and $u = 1.00$ does not have any effect.

In step five a target risk is chosen for the target return defined through $\lambda_1$. The target risk will be situated between the efficient frontier and the minimum diversification line that defines for any target return the portfolio that has the highest diversification. If the orientation has a large positive value $\lambda_2$ will tend to 1 and the target portfolio will move towards the diversification line. A better diversification means that the exposure into assets that tend to have higher risks is raised. For positive orientations this might be a risk worth taking. But for negative orientations only a minimum amount of diversification is enforced. This is the reason for the indicator function. Choosing $c = 4.00$ means that the transition from the efficient frontier towards the diversification line is speeded up. Choosing $u = 1.00$ makes sure that $\lambda_2$ is not greater than 1. Choosing $l = 0.10$ makes sure that there is always a minimum amount of diversification enforced.

Note that $\sigma_{lower,upper}$ can be defined in any other way. The parameters $\boldsymbol{\mu}$ and $\Sigma$ can be calculated by any estimator. One just has to make sure that these estimators are also used to calculate the orientation ($o$) and the lower and upper bounds ($\mu_{upper,lower}$ and $\sigma_{upper,lower}$).

Fig. 15.1 visualizes the concept. It shows the feasible set for one of the 36 month windows of the backtest. The shaded area indicate the degree of diversification. The EWP is the portfolio that globally has the highest diversification. The diversification declines radially towards the hull. The diversification line (MLP-Line) shows the portfolio that has the highest diversification given the target return. In terms of diversification it does not make sense to choose a portfolio that lies behind the diversification line. This is

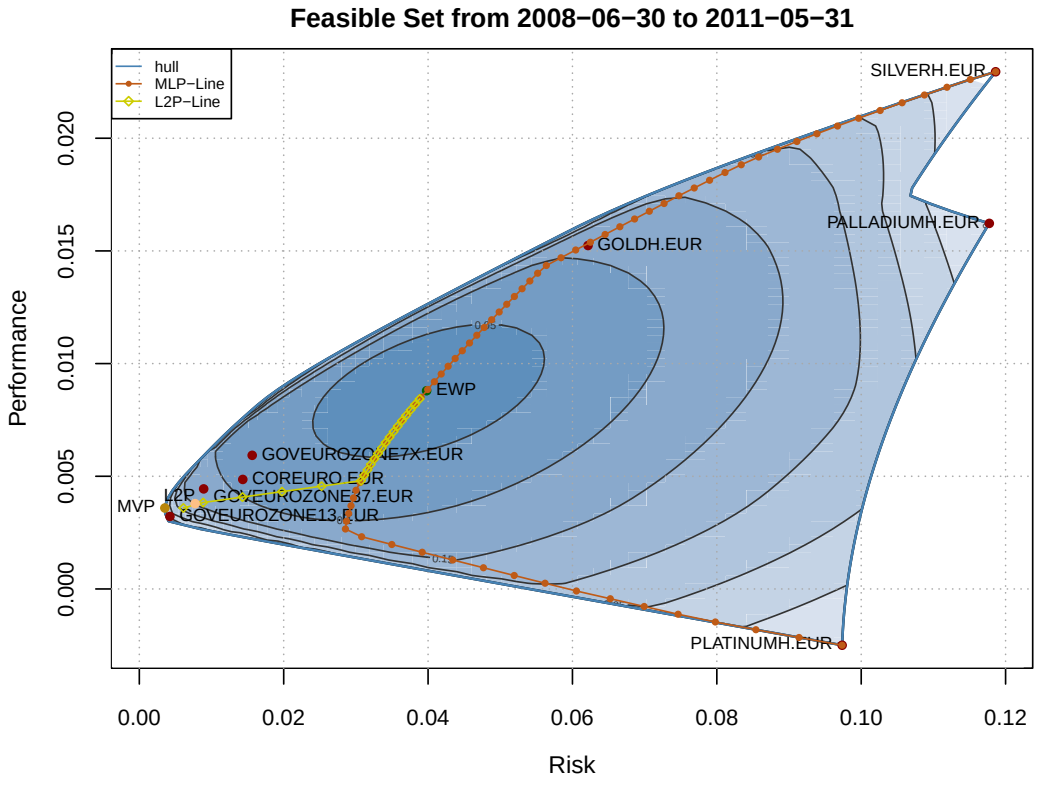


Figure 15.1: The feasible set of the European bonds and precious metals universe using the sample estimator. The shaded areas show the degree of diversification which declines radially from the EWP towards the hull. See also Appendix A for more information about the data.

enforced in step four of the algorithm. Target return wise the dynamic L2P will always lie somewhere in between the MVP and the maximum return portfolio; which are the returns of the efficient frontier. Target risk wise the portfolio will always lie between the efficient frontier and the diversification line. The orientation of this feasible set is 3.57°. The figure shows that the L2P portfolio for this orientation is located close to the MVP. The L2P-Line shows the portfolios for orientations manually set between 0° and 90°. If the L2P is calculated using Eq. (15.1), Eq. (15.2) and Eq. (15.3) by setting $\bar{o}$ manually to 0 it would be situated at the bottom of the L2P-Line (close to the MVP). If $\bar{o} = 1$ the L2P would be situated at the top of the L2P-Line (close to the EWP) and if $\bar{o} = 0.5$ in the middle of this line. Not that the top of the L2P-Line is not necessarily situated close

to the EWP; especially not for negative orientations. The reason for the hockey stick like shape of the L2P-Line is the design of $\lambda_{1,2}$ where $\lambda_2$ was designed such that the target portfolio travels quickly towards the diversification line if the orientation gets positive.

The results of the backtest (out of sample) are shown in Fig. 15.2 and Table 15.1. The MVP is almost exclusively invested into the 1-3 year government bonds. This is certainly not very much preferable since one could just invest all the capital manually in that index. The dynamic L2P (L2P Dynamic) uses the BCP estimator (Eq. (9.23) including orthogonalization) to calculate $\boldsymbol{\mu}$ and $\Sigma$. The parameters used are the same as those used in Section 9.3.4 and Section 14.2.4. This portfolio respects the model of changing dynamics (see Section 14.1) and the model of shape shifts at the same time. It offers a reasonably high return while keeping the risks low. Using the sample estimator (Eq. (9.8) and Eq. (9.9)) for $\boldsymbol{\mu}$ and $\Sigma$ leads to a portfolio (L2P Dynamic Sample) that has a similar Sharpe ratio than the dynamic L2P. One could argue that through a leverage approach the dynamic L2P and the dynamic sample L2P would have about the same performance and risk. While this is true it might be preferable to have the performance of the dynamic L2P without setting up a leverage scheme. The portfolio (L2P Dynamic Robust) that uses the robust estimator (Section 9.2.2) is similar to the dynamic sample L2P. All portfolios are certainly more preferable than the EWP.

| | Performance | Volatility | Sharpe | Max. Drawdown |
|---|---|---|---|---|
| L2P Dynamic | 4.79 | 3.00 | 1.60 | 4.08 |
| L2P Dynamic Sample | 3.99 | 2.50 | 1.60 | 2.70 |
| EWP Base Assets | 5.39 | 11.28 | 0.48 | 31.75 |
| MVP | 3.36 | 1.61 | 2.08 | 2.24 |
| L2P Dynamic Robust | 4.15 | 2.80 | 1.48 | 5.31 |
| L2P Constant | 4.10 | 2.98 | 1.38 | 4.50 |

Table 15.1: Key figures for the indices shown in Fig. 15.2. The performance and the volatility are the annualized return and volatility, respectively.

To asses on whether the dynamic lambdas offer a benefit the dynamic L2P is compared to an L2P where the lambdas are kept constant (L2P Constant). More precisely $\lambda_1 = 0.038$ and $\lambda_2 = 0.268$ are the mean values of the lambdas that were used for the dynamic L2P. It shows that the dynamic lambdas offer a benefit. Note that the constant L2P actually uses information from the future by setting $\lambda_{1,2}$ to the mean values. For the dynamic L2P only the upper and lower bounds are set which define the risk-averseness of the investor. The time dependent (dynamic) lambdas are then chosen based on the orientation.

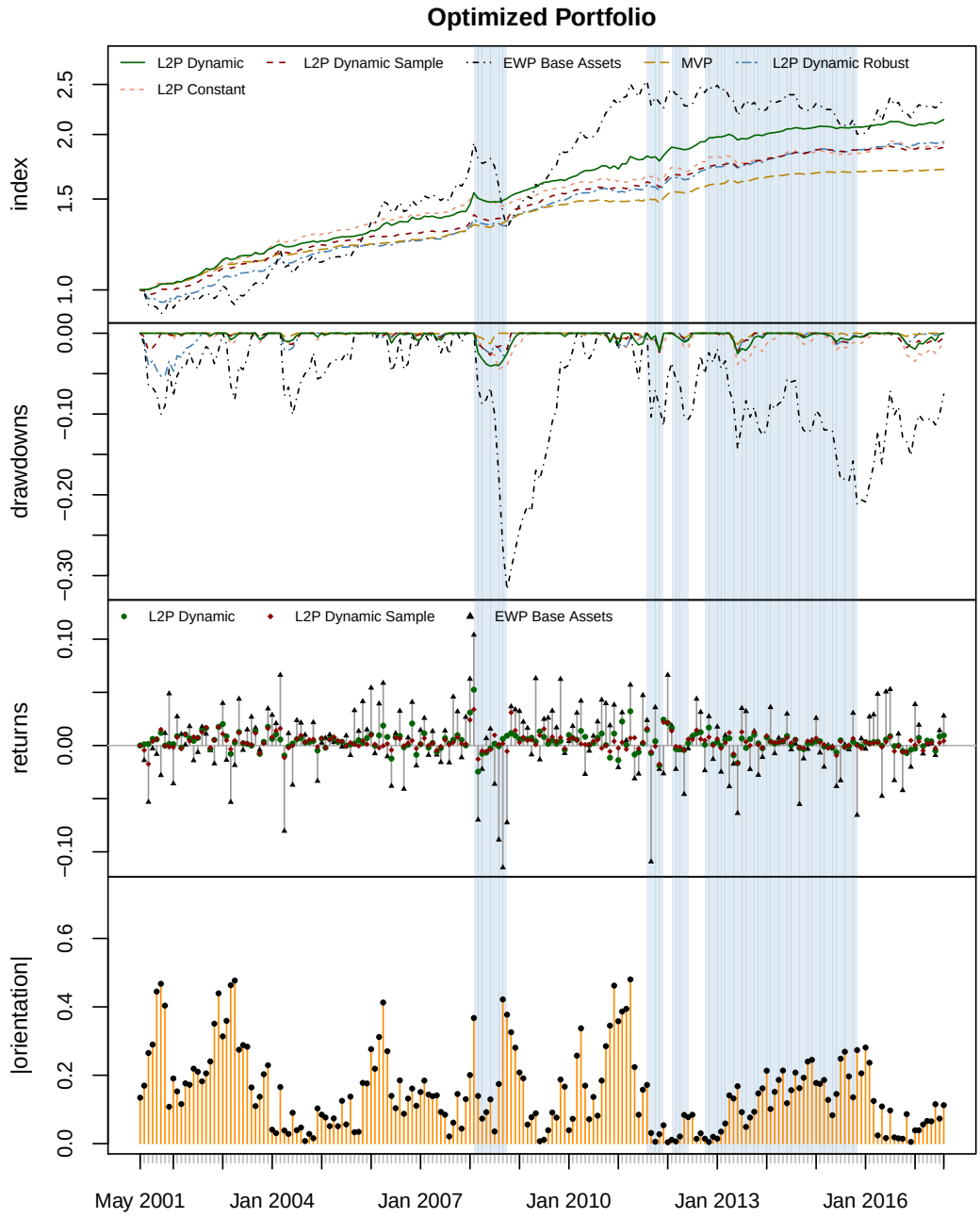


Figure 15.2: The L2P for the European bonds and precious metals universe (base assets). The shaded areas show the times where the BCP estimation of the trend for the EWP is negative. See also Appendix A for more information about the data.

### 15.2.2 The Multi-Objective Approach

The multi objective approach follows the same steps as the L2P. First the portfolio is defined as

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \lambda f_1(\boldsymbol{w}) + (1-\lambda) f_2(\boldsymbol{w}) \\ & \text{s.t.} \\ & \mathbf{1}'\boldsymbol{w} = 1 \\ & 0 \leq w_i \leq 1 \end{aligned} \tag{15.5}$$

where $0 \leq \lambda \leq 1$ and $\boldsymbol{w}$ describes the weights of the assets within the investment universe. For the backtest the objective functions $f_{1,2}$ were chosen as

$$f_1(\boldsymbol{w}) = \text{var}(\boldsymbol{w}) \tag{15.6}$$

$$f_2(\boldsymbol{w}) = \boldsymbol{w}'\Sigma\boldsymbol{w} \tag{15.7}$$

where $\Sigma$ is the covariance matrix. If $\lambda = 1$ the result will be the EWP and if $\lambda = 0$ the result will be the MVP. In contrast to the L2P this portfolio definition does not allow to reach any point within the feasible set. It is always situated on the pair-line of the EWP and the MVP. Taking other objective functions would just lead to a different pair-line If e.g. $f_1$ is the maximum return portfolio the target portfolios would always lie on the efficient frontier without the possibility to get a better diversified portfolio. This could be solved by introducing a third objective function and a second lambda. The problem is that the exact location of the target portfolio based on the lambdas is not obvious. For the above setup setting $\lambda = 0.5$ does not mean that the target portfolio will be in the middle of the pair-line between the EWP and the MVP. If a second $\lambda$ is introduced the situation gets even more complex.

Let $W = (w_{t,k})$ be the weights matrix where $w_{t,k}$ describes the weight of asset $k$ at time $t$ and the vector $\boldsymbol{w}_t = \{w_1, w_2, \ldots, w_n\}$ holds the weights at time $t$ for all assets. At any point in time $t$ the weights $\boldsymbol{w}_t$ can be calculated as

1. Calculate the orientation $\bar{o} = o/90$ of the feasible set. Calculate $\lambda = f(\bar{o})$.
2. Calculate $\boldsymbol{w}_t$ using Eq. (15.5), Eq. (15.6) and Eq. (15.7).

The advantage of this setup is that it is much more compact than the L2P. For the

backtest $\lambda$ was defined in the same way as the lambdas for the L2P by using Eq. (15.3) as

$$\lambda = f(\bar{o}) = f(\bar{o}, c = 0.05, l = 0.001, u = 1.00) = \min(u, \max(l, |\bar{o}|c))I(\bar{o} > 0) \quad (15.8)$$

where $I$ is the indicator function. As already mentioned the impact of the $\lambda$ is not quite linear. To not move too far away from the MVP (risk-averse) the scaling factor $c = 0.05$ has to be chosen quite small. Setting $l = 0.001$ always enforces a minimum amount of diversification. The indicator function is used to not move towards the EWP if the orientation gets negative. This wouldn't make sense since for a negative diversification the EWP would generally have a lower return and higher risk than the MVP. For the same reasoning as for the L2P the target portfolio will be the MVP if the orientation goes towards 0.

| | Performance | Volatility | Sharpe | Max. Drawdown |
|---|---|---|---|---|
| MOP Dynamic | 4.86 | 3.20 | 1.52 | 4.84 |
| MOP Constant | 5.07 | 3.19 | 1.59 | 4.84 |
| EWP Base Assets | 5.39 | 11.28 | 0.48 | 31.75 |
| MVP | 3.36 | 1.61 | 2.08 | 2.24 |
| 5% EWP + 95% MVP | 3.46 | 1.78 | 1.95 | 2.60 |

Table 15.2: Key Figures for the indices shown in Fig. 15.3. The performance and the volatility are the annualized return and volatility, respectively.

The results of the backtest (out of sample) are shown in Fig. 15.3 and Table 15.2. The dynamic multi objective portfolio (MOP Dynamic) is quite similar to the dynamic L2P although it shows a lower Sharpe ratio. As the dynamic L2P the dynamic MOP uses the BCP estimator (using the same parameters) to calculate $\boldsymbol{\mu}$ and $\Sigma$ and is more preferable than the EWP.

The constant multi objective portfolio (MOP Constant) was calculated by setting $\lambda = 0.0036$ at all times. This value is the mean value of the lambdas that were used for the dynamic MOP. There seems not to be much of an impact on whether having a constant lambda or a dynamic one. It means that setting any low value for lambda leads to very similar results. This implies that it doesn't matter too much on whether the orientation offers valuable information. Since it wouldn't have much of an impact anyway. The reason is that if a small lambda is increased by a small step the portfolio does not go far on the pair-line. For large lambdas and small increases this would be the opposite.

Note that the constant MOP uses information from the future by using the mean

value of the dynamic MOP for the lambda. Also the dynamic MOP is generally more robust against dramatic changes within the investment universe since the lambda can be adapted to a new situation. It make sense to choose a model that is conceptually more attractive even tough the results within the backtest might be slightly lower.

The lambdas suggest to always move slightly away from the MVP. The results do therefore also include the 5%-EWP+95%-MVP benchmark. Such a portfolio does not significantly differ from the MVP. But moving slightly away from the MVP on its pair-line with the EWP seems to have quite an impact on the performance. One could argue that this shows a diversification benefit. At least for this case the reason is that the MVP is very much concentrated into the asset that is closest to it; which is almost exclusively the 1-3 years government bonds. This effect can be observed in other universes as well. Therefore it might generally make sense to add some diversification to any MVP approach.

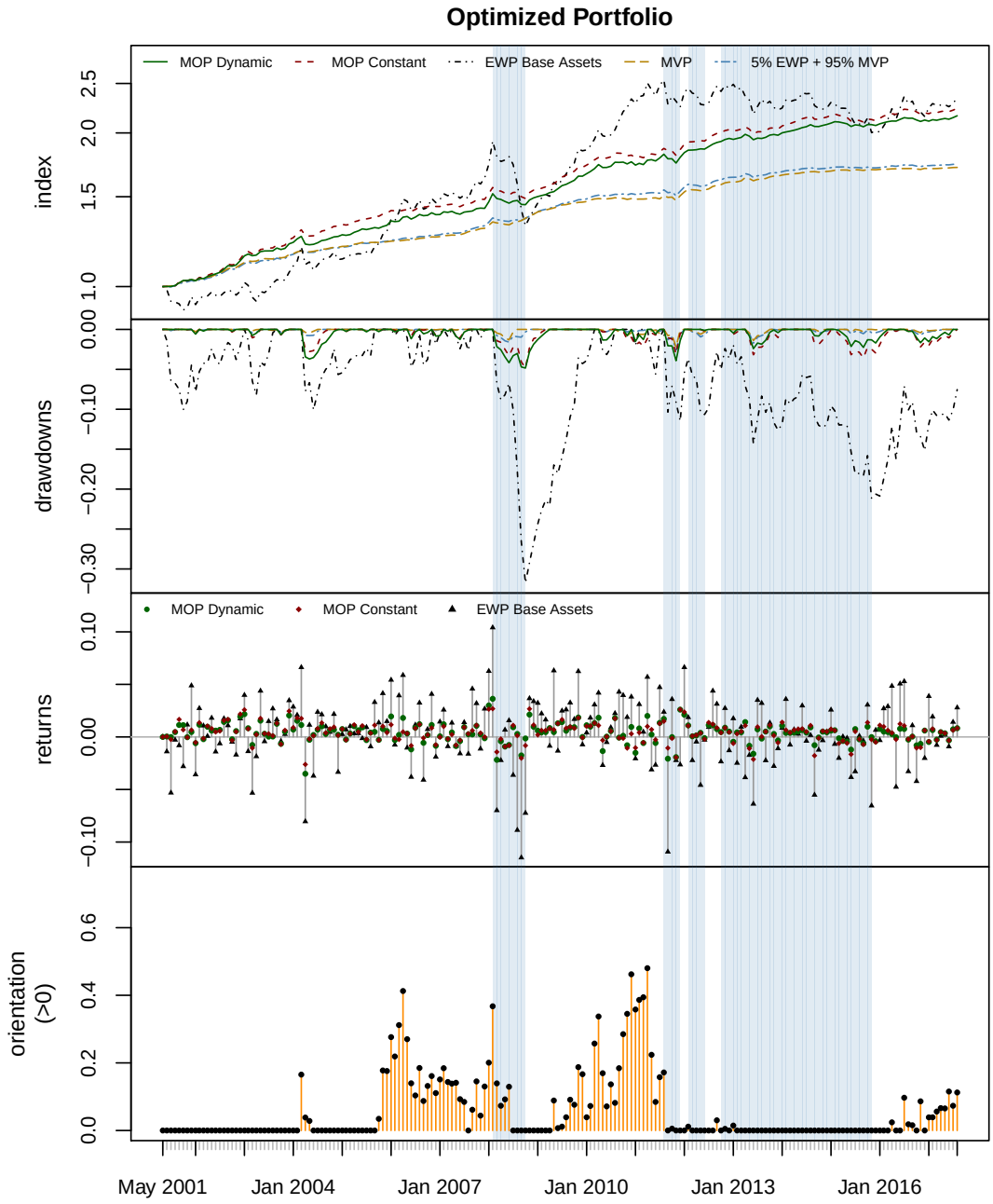


Figure 15.3: The multi objective portfolio (MOP) for the European bonds and precious metals universe (base assets). The shaded areas show the times where the BCP estimation of the trend for the EWP is negative. See also Appendix A for more information about the data.

# 16 Combined Portfolios

The BCP-Signal-Portfolio (BSP) for the European equity universe (see Fig. 14.3) is a dynamically hedged portfolio. At any point in time there is either a hedge in place for part of the assets within the investment universe or the invested amount that would be hedged is moved to a cash account. Instead of moving it to a cash account it could also be moved to a different portfolio that consists of a more conservative investment universe than the equity universe of the BSP. The Lambda-2-Portfolio (L2P) for the European bonds and precious metals universe (see Fig. 15.2) represents such a portfolio. The combined portfolio is therefore a combination of the BSP and the L2P where the potentially hedged investment amount of the BSP is moved to the L2P.

Let $W = (w_{t,k})$ be the weights matrix where $w_{t,k}$ describes the weight of asset $k$ at time $t$ and the vector $\boldsymbol{w}_t = \{w_1, w_2, \ldots, w_n\}$ holds the weights at time $t$ for all assets. Let $\boldsymbol{w}_{t,BSP}$ be the weights of the assets (the 19 industry sectors of the European equity universe) within the BSP universe and $\boldsymbol{w}_{t,L2P}$ the weights of the assets (the European government and corporate bonds and the precious metals) within the L2P universe. At any point in time $t$ the weights $\boldsymbol{w}_t$ can be calculated as

$$\begin{aligned} e &= \sum \boldsymbol{w}_{t,BSP} \\ \boldsymbol{w}_t &= \{\boldsymbol{w}_{t,BSP}, (1-e)\boldsymbol{w}_{t,L2P}\} \end{aligned} \tag{16.1}$$

where $e$ is the sum of the weights of the assets within the BSP universe. Therefore $e$ describes the equity exposure. A value of for example $e = 0.4$ means that 40% of the capital is invested into the BSP universe and 60% into the L2P universe. The sum of the weights of the combined portfolio is always 1 (fully invested) and the weights are always between 0 and 1 (no short-selling and no leverage). Also does the combined portfolio not need any hedging scheme.

The combined universes of the BSP and L2P represent a reasonably well diversified universe of European equities, bonds and precious metals. The currency for all indices is the Euro. The precious metals are hedged against movements within the euro-US dollar

currency. For unstable times within the European equity market the investment is shifted to European bonds and precious metals where a stable constellation is searched. Because of the non-linear behaviour of the lambda for the dynamic MOP the dynamic L2P was chosen.

| | Performance | Volatility | Sharpe | Max. Drawdown |
|---|---|---|---|---|
| Combined Portfolio | 7.28 | 7.49 | 0.97 | 12.99 |
| 50% ES + 50% BA | 4.80 | 10.48 | 0.46 | 34.69 |
| EWP European Sectors (ES) | 4.22 | 15.80 | 0.27 | 54.86 |
| EWP Base Assets (BA) | 5.39 | 11.28 | 0.48 | 31.75 |

Table 16.1: Key Figures for the indices shown in Fig. 16.1. The performance and the volatility are the annualized return and volatility, respectively.

The results of the backtest (out of sample) are shown in Fig. 16.1 and Table 16.1. The exposure within the BSP universe is between 0% and 100%. The same is true for the L2P universe. Therefore the natural benchmark is the portfolio that is invested to 50% within the BSP universe and to 50% within the L2P universe. The combined portfolio shows a higher Sharpe ratio and risk profile than the benchmark.

This portfolio combines all the concepts and tools introduced during this thesis. All the indices used for this portfolio are actually eligible for investment through ETF's. Since there is no hedging, leverage or short-selling involved anyone with a basic bank account can realize this portfolio by only buying and selling ETF's. In practice there are of course further constraints like minimum or maximum weights for certain assets. The goal for this thesis was just to demonstrate how the concepts and tools introduced could be applied. But generally the model is open for the introduction of any constraints or also schemes that for example reduce the rebalancing amount.

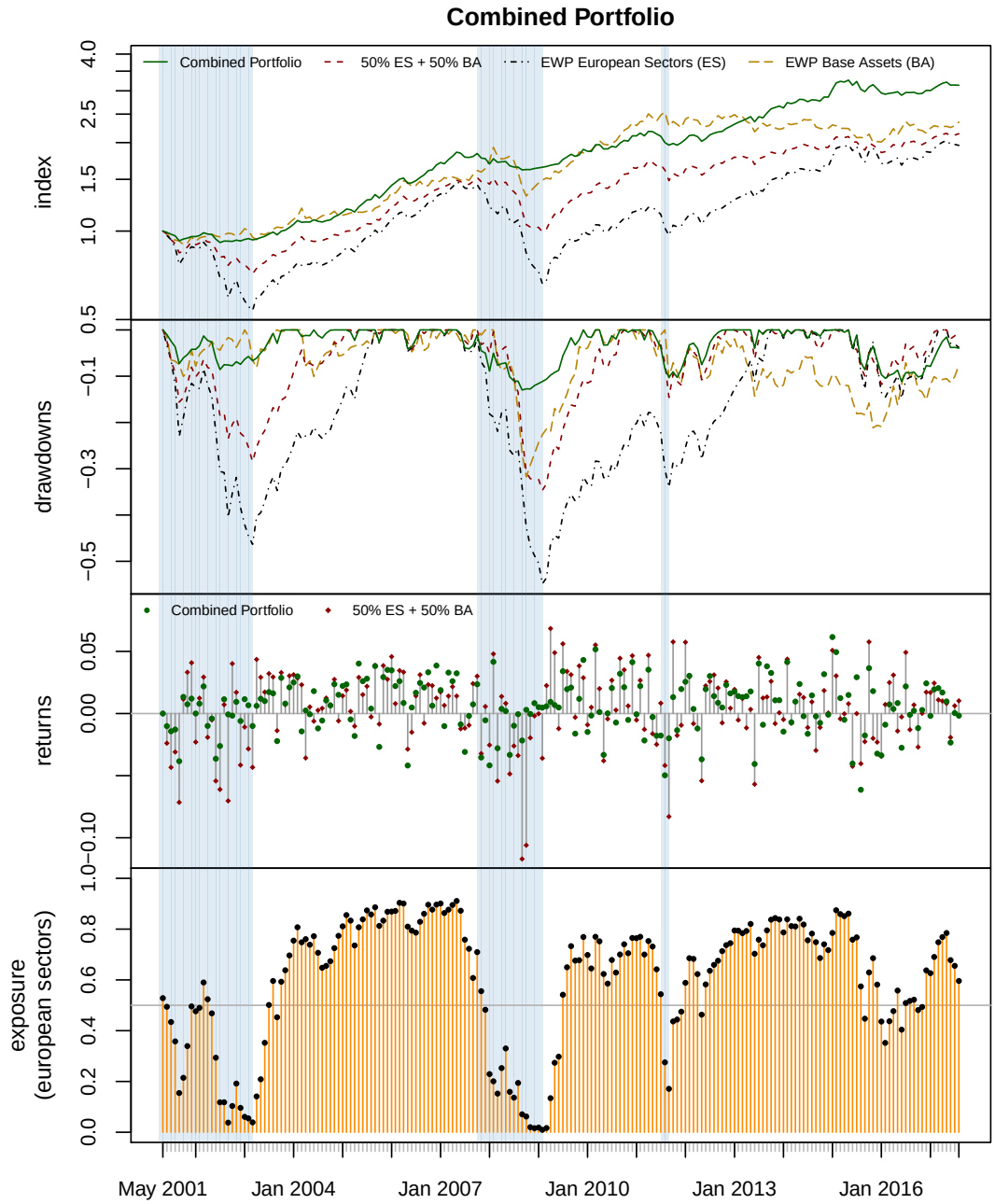


Figure 16.1: The combined portfolio of the BSP for the European equity universe (see Fig. 14.3) and the L2P of the European bonds and precious metals universe (see Fig. 15.2). The exposure shows the amount of the capital that is invested within the equities. The shaded areas show the times where the BCP estimation of the trend for the 50%-ES+50%-BA index is negative. See also Appendix A for more information about the data.

# 17 Summary and Outlook

The tools that were presented in Part I and Part II were consequently applied in Part III. The notion of stability was introduced in Chapter 13.

In Chapter 14 signal portfolios and the model of changing dynamics were introduced. It was shown how the results of a BCP analysis can be used to generate hedging signals between 0 and 1 for individual indices and how these results can be combined into a signal portfolio. There might be other ways to calculate the indicator values or to transform the indicator values into the range between 0 and 1 that could be tested. Also it would be interesting to test this concept on other investment universes.

In Chapter 15 dynamic Markowitz portfolios and the model of shape shifts were introduced. It was shown how the results of a geometric shape factor (GSF) analysis can be used to define a target portfolio based on the orientation. For that the Lambda-2-Portfolio (L2P) was introduced. Additionally a multi objective approach was examined. It would be of interest to find models that include other GSF (the area, centre and/or eccentricity) or moments (e.g. the Hu moments) to define the target portfolio. The backtested L2P portfolios did not define the objective function since the results were very similar to the case where the objective function was defined to be the diversification of the portfolio. However, it would be of interest to study the impact of defining the objective function for various risk measures or combinations of them in more detail. Also there might be other interesting definitions for the target portfolio than the L2P or the multi objective approach. For the multi objective approach it would be of interest to study the impact of the lambda on the target portfolio in more detail. Additionally it would be interesting to test this concept on other investment universes.

In Chapter 16 the BCP-Signal-Portfolio (BSP) and the dynamic L2P were combined. Instead of setting up hedges within the BSP universe by moving capital to a cash account that capital is moved to the L2P universe. This leads to a portfolio that can be implemented without the use of any additional instruments than buying and selling. It would be of interest to design similar portfolios for investors that are based in a different currency than the euro. Also the investment universe could be extended to even more

investment classes and/or indices.

All the results showed appealing Sharpe ratios and risk profiles compared to their benchmarks. Generally it would be of interest to test these concepts on a higher frequency (e.g. weekly or daily). Note that the tools introduced in Part I and Part II could in general also be used for the examination of any other measurements than financial returns.

# 18 Conclusion and Outlook

In Part III the notion of stability in the context of portfolio design [Wuertz, 2010] was outlined. It follows a two step procedure. First the drivers of unstable times (for financial returns) are identified. Then a mathematical model is defined that can measure on how much these drivers are activated.

One possible model for instabilities is the model of changing dynamics (see Section 14.1). A mathematical framework for this model was introduced in Part I. For that the Bayesian model for normally (N) distributed data where the mean of the data is normally (N) distributed and the variance of the data follows a generalized inverse Gaussian (GIG) distribution was solved (see Chapter 4). More precisely the posterior marginal and conditional distributions as well as the marginal likelihood were derived for the N-NGIG model. It was shown that these distributions belong to the family of generalized hyperbolic (GH) and GIG distributions. Since these distributions are well documented in literature and available through various software implementations it makes the distributions convenient to work with. The posterior marginal distributions and the marginal likelihood were then used to set up the Bayesian change point (BCP) model as introduced by Barry and Hartigan [1992, 1993] and later extended by Loschi et al. [1999, 2003] (see Chapter 5). The difference between the model of this thesis (the BCP N-NGIG model) and the model of Loschi et al. [2003] (the BCP N-NIG model) is that the variance is modelled through a GIG distribution instead of an inverse gamma (IG) distribution. The IG distribution is a special case of the GIG distribution. Therefore the BCP N-NGIG model represents a generalisation of the BCP N-NIG model (see Chapter 6). For other applications than the BCP model it would be of interest to find closed form solutions of the posterior distribution and the predictive distribution as well. While those distributions have been derived by Thabane and Safiul Haq [1999] in the context of generalized modified Bessel distributions it might be possible to define these distributions in the context of GH distributions. Further research steps could include to model more of the introduced hyperparameters through distributions or to change the normal assumption of the data. A candidate for that could be the normal inverse Gauss (NIG) distribution. Additionally

the model could be set up for the multivariate case in order to calculate the covariances directly instead of using the orthogonalized Gnanadesikan–Kettenring (OGK) framework as introduced by Maronna and Zamar [2002]. It would also be of interest to answer the question on how robust the new method is when it is faced with many different scenarios where the mean and/or the variance change for various parts of the data.

The second model for instabilities that was introduced is the model of shape shifts (see Section 15.1). It assumes that during unstable times the shape of the investment universe under consideration changes. A mathematical framework for this model was introduced in Part II. The shape of the investment universe is generally described through the feasible set which can be approximated through an ellipse that has the same area, centre, orientation and eccentricity as the feasible set [Wuertz, 2010]. These figures were called the geometric shape factors (GSF). If the feasible set is considered to be an image then the GSF are a function of the image moments. The Hu moments [Hu, 1962] are also functions of the image moments and invariant under certain transformations of the feasible set. It was shown how the feasible set can be calculated through different estimators. Two new estimators based on the BCP model as introduced in Part I were defined. The GSF and the Hu moments for the sample estimator, the robust OGK estimator and the new BCP estimators were compared on different investment universes (see Chapter 11). There are many more estimators to calculate the feasible set that could be compared to each other in means of the GSF, image moments or the Hu moments. The GSF can be associated with a geometric meaning. It would be of special interest to find accessible geometric meanings for the image moments or the Hu moments such that they can be used for the design of dynamic portfolios as well.

In Chapter 14 the output of the new BCP model was used to generate hedging signals for single indices. If the signals are calculated for multiple indices a signal portfolio can be designed which calculates the hedge ratios for every index within the investment universe separately. The signal portfolio which uses the BCP results to calculate the hedging signals was denoted as the BCP-Signal-Portfolio (BSP). To calculate the BSP the European equity universe was used as described in Appendix A. The BSP represents an optimizationless univariate approach. Further research steps could include to consider alternative routines to generate indicator values and signals from the BCP output. Or to calculate indicator values and signals from other univariate statistical analyses. Generally the question should be answered if this approach is universally applicable on any kind of underlying investment universe and if the current indicator and signal logic could be improved.

In Chapter 15 the orientation of the GSF was used to calculate a dynamic Markowitz portfolio where the distance towards the Minimum-Variance-Portfolio (MVP) is defined through the orientation. The orientation was calculated using the BCP estimator as defined in Eq. (9.23) (including orthogonalization). To have full control over where the portfolio is situated within the feasible set the Lambda-2-Portfolio (L2P) was introduced. This in contrary to the multi objective approach where it is cumbersome to situate the portfolio precisely. To calculate the L2P the European bonds and precious metals universe was used as described in Appendix A. The L2P represents an optimized multivariate approach. Further research steps could include to consider alternative models where not only the orientation is used to define the portfolio but also other GSF, image moments or Hu moments. Also the impact of defining the objective function for various risk measures or combinations of them could be examined in more detail. Generally the question should be answered if this approach is universally applicable on any kind of underlying investment universe and on whether the current process to design the portfolio through the orientation of the GSF could be improved. Also there might be other interesting definitions for the target portfolio than the L2P or the multi objective approach. For the multi objective approach it would be of interest to study the impact of the lambda on the target portfolio in more detail.

In Chapter 16 the BSP for the European equity universe and the L2P for the European bonds and precious metals universe were combined. At any point in time the capital that is supposed to be hedged within the BSP is moved to the L2P instead. All three actively managed portfolios (the BSP, the L2P and the combination) offer more preferable return and risk profiles than their passive natural benchmarks. The combined universe represents a well diversified investment universe consisting of European equities, euro dominated bonds and euro hedged precious metals. Therefore the currency risk is very much limited for an euro based investor. The combined portfolio is free of any hedging or leverage schemes. It can be implemented through buying and selling only. Since all the indices used within the investment universe are eligible for investment through ETF's anyone with a bank account can in principle realise this portfolio. Further research steps could include to design similar portfolios for investors that are based in a different currency than the euro. Also the investment universe could be extended to even more investment classes and/or indices. One important further research step for all the presented concepts would certainly be to test them on a higher data frequency. For example for weekly or daily financial returns.

# A Data

## The European equity index

The European equity index (EQEU.EUR) consists of 600 company shares that represent large, mid and small capitalized companies (200 each) across 17 European countries. These are Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. The currency of the index is in euro and the returns are net returns (includes dividends and taxes). Note that company shares that are not denoted in euros are converted to euros directly and are therefore not hedged (currency risk). The reason to use this index was that there are no ETF's available for the sector indices (the European equity universe) where the currency risk (from the perspective of an euro based investor) is hedged. Also the company shares denoted in euro cover more than 60% of the index which reduces the risk.

## The European equity universe

The European equity universe groups the company shares of the European equity index into 19 industry sectors. These are automobiles & parts, banks, basic resources, chemicals, construction & materials, financial services, food & beverage, health care, industrial goods & services, insurance, media, oil & gas, personal & household goods, real estate, retail, technology, telecommunications, travel & leisure, and utilities. The individual sector indices have the same properties as the full European equity index. The currency of every index is the euro and the returns are net returns (includes dividends and taxes). They contain some currency risk and there are no ETF's available that hedge out that risk. The unhedged sector indices are all eligible for investment through ETF's.

## The European bonds and precious metals universe

The universe consists of three government bond indices, one corporate bond index and four precious metals. The currency for all indices is the euro and the returns are total returns. Since all bonds are denominated in euro and since the precious metals are hedged for movements within the euro-US dollar currency there is no currency risk involved for euro based investors. All indices are eligible for investment through ETF's.

The government bond indices all hold euro denominated bonds of eurozone countries (mainly Germany, Italy, France, Netherlands and Spain). The first index (GOVEUROZONE13.EUR) holds bonds with maturities from 1 to 3 years. The second index (GOVEUROZONE37.EUR) holds bonds with maturities from 3 to 7 years. The third index (GOVEUROZONE7X.EUR) holds bonds with maturities from 7 to 10 years. The corporate bond index (COREURO.EUR) holds globally issued euro denominated bonds of all maturities with investment grade ratings (mainly developed markets with a strong focus on Europe).

The base currency of precious metals is the US dollar. The used indices are hedged against movements within the euro-US dollar currency. The indices describe therefore the euro hedged prices for gold (GOLDH.EUR), silver (SILVERH.EUR), palladium (PALLADIUMH.EUR) and platinum (PLATINUMH.EUR). Note that for this case the euro hedged price is actually equivalent with the US dollar price (assuming a perfect hedging scheme).

## The global equity and bond universe

The universe consists of four equity indices, two government bond indices and two corporate bond indices. All indices have the euro as currency. Most of them are converted from US dollars to euro.

All equity indices are designed to cover approximately 85% of the market capitalization of their corresponding market. The fist equity index (EQNORTHAMERICA.EUR) measures the performance of the US and Canadian markets. The second equity index (EQEUROPE.EUR) measures the performance of the developed European market (Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom). The third index (EQPACIFIC.EUR) measures the performance of the developed Pacific market (Australia, Hong Kong, Japan, New Zealand and Singapore). The fourth index (EQEM.EUR)

measures the performance of the emerging markets (Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Russia, Qatar, South Africa, Taiwan, Thailand, Turkey and United Arab Emirates). All equity indices are converted from US dollars to euros and the returns are net returns (includes dividends and taxes) for all indices.

The first government bond index (GOVG7.EUR) holds bonds of all maturities issued by the G7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom and the United States) in their local currencies. The index is converted from US dollars to euro. The second government bond index (GOVEUROZONE.EUR) holds euro denominated bonds of all maturities issued by eurozone countries. The currency of the index is the euro. The first corporate bond index (CORGLOBAL.EUR) holds bonds of all maturities with investment grade ratings issued by corporations in emerging and developed markets in their local currencies (mainly developed markets with a strong focus on the United States). The index is converted from US dollars to euro. The second corporate bond index (COREURO.EUR) holds globally issued euro denominated bonds of all maturities with investment grade ratings (mainly developed markets with a strong focus on Europe). The currency of the index is the euro. The returns of all bond indices are total returns.

While these indices are generally eligible for investment through ETF's the currency risk is considerably high. Only the GOVEUROZONE.EUR and the COREURO.EUR indices do not have any currency risk for euro based investors.

## The pure bond universe

The universe consists of the government bonds of 11 of the countries that have adopted the euro (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal and Spain). Additionally it contains a bond index for the United States, the United Kingdom and the European Union as a whole. All indices hold bonds of all maturities in their local currencies and the returns are total returns. The currency of all indices is the euro. These indices are only partially eligible for investment through ETF's.

# List of Publications


T. Setz and D. Würtz. Size and style variability of stability filtrations: A country and regional stock market study. https://www.rmetrics.org/WorkingPapers, 2015a.

T. Setz and D. Würtz. *Bayesian Stability Concepts for Investment Managers*. Rmetrics Association, Zurich and Finance Online GmbH, Zurich, 2015b.

D. Würtz, H. Bailer, Y. Chalabi, F. Grimson, and T. Setz. *Long Term Statistical Analysis of US Asset Classes*. Rmetrics Association, Zurich and Finance Online GmbH, Zurich, 2013a.

D. Würtz, T. Setz, and Y. Chalabi. New directions in active portfolio management: Stability analytics, risk parity, rating and ranking, and geometric shape factors. https://www.rmetrics.org/WorkingPapers, 2013b.

D. Würtz, T. Setz, Y. Chalabi, W. Chen, and A. Ellis. *Portfolio Optimization with R/Rmetrics*. Rmetrics Association, Zurich and Finance Online GmbH, Zurich, 2015a.

D. Würtz, T. Setz, Y. Chalabi, and A. Ellis. *Chronological Objects with Rmetrics*. Rmetrics Association, Zurich and Finance Online GmbH, Zurich, 2015b.

D. Würtz, T. Setz, Y. Chalabi, L. Lam, and A. Ellis. *Basic R for Finance*. Rmetrics Association, Zurich and Finance Online GmbH, Zurich, 2015c.

# Bibliography

H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

M. Abramowitz and I. A. Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. 1964.

O. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 353(1674):401–419, 1977.

D. Barry and J. A. Hartigan. Product partition models for change point problems. *The Annals of Statistics*, 20(1):260–279, 1992.

D. Barry and J. A. Hartigan. A bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421):309–319, 1993.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

P. Bourke. Calculating the area and centroid of a polygon. Lecture at University of Pennsylvania, 1988.

G. E. Box and G. M. Jenkins. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.

G. E. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, 1992.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, I:223–226, 2001.

P. L. Davies. Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 15(3):1269–1292, 1987.

D. L. Donoho. Breakdown properties of multivariate location estimator. *unpublished manuscript (Ph.D. qualifying paper)*, 1982.

D. L. Donoho and H. P. J. The notion of breakdown-point. *Festschrift fur Erich L. Lehman*, pages 157–1984, 1983.

A. W. F. Edwards and L. L. Cavalli-Sforza. A method for cluster analysis. *Biometrics*, 21(3):362–375, 1965.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

C. Erdman and J. W. Emerson. bcp: An r package for performing a bayesian analysis of change point problems. *Journal of Statistical Software*, 23(3):1–13, 2007.

E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

J. Flusser. Moment invariants in image analysis. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 1(11):3708 – 3713, 2007.

J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167 – 174, 1993.

A. Ghalanos and S. Theussl. *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method*, 2015. R package version 1.16.

R. Gnanadesikan and J. R. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1):81–124, 1972.

G. B. Gurevich. *Foundations of the Theory of Algebraic Invariants*. P. Noordhoff, 1964.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

J. M. Hammersley and H. D. C. *Monte Carlo Methods*. Chapman and Hall Ltd, 1964.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

D. Hilbert. *Theory of Algebraic Invariants*. Cambridge University Press, 1993.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.

P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

P. J. Huber. *Robust Statistic*. John Wiley & Sons, 1981.

N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions*. John Wiley & Sons, 2005.

B. Jorgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer New York, 1982.

H. P. Lopuhaa. Highly efficient estimators of multivariate location with high breakdown point. *The Annals of Statistics*, 20(1):398–413, 1992.

R. H. Loschi, P. L. Iglesias, and R. B. Arellano-Valle. Bayesian detection of change points in the chilean stock market. *Proceedings of Section on Bayesian Statistical Science - Annual Meeting of the American Statistical Association*, pages 160–165, 1999.

R. H. Loschi, F. R. B. Cruz, P. L. Iglesias, and R. B. Arellano-Valle. A gibbs sampling scheme to the product partition model: an application to change-point problems. *Computers & Operations Research*, 30(3):463–482, 2003.

D. Luethi and W. Breymann. *ghyp: A package on the generalized hyperbolic distribution and its special cases*, 2013. R package version 1.5.6.

M. Maechler. *Bessel: Bessel Functions Computations and Approximations*, 2015. R package version 0.6-0/r24.

M. Maechler, P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, and M. Anna di Palma. *robustbase: Basic Robust Statistics*, 2016. R package version 0.92-7.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36 (4):394–419, 1963.

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

R. A. Maronna. Robust m-estimators of multivariate location and scatter. *The Annals of Statistics*, 4(1):51–67, 1976.

R. A. Maronna and V. J. Yohai. Robust estimation of multivariate location and scatter. *Wiley StatsRef: Statistics Reference Online*, page 1–12, 2016.

R. A. Maronna and R. H. Zamar. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4):307–317, 2002.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 (6):1087–1092, 1953.

J. O'Rourke. *Computational Geometry in C (Second Edition)*. Cambridge University Press, 1998.

K. F. R. S. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

E. J. Pebesma and R. S. Bivand. Classes and methods for spatial data in R. *R News*, 5 (2):9–13, November 2005.

B. Pfaff. *Financial Risk Modelling and Portfolio Optimization with R*. John Wiley & Sons, 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

J. S. Ramberg and B. W. Schmeiser. An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82, 1974.

S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250, 1959.

P. Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, pages 283–297, 1985.

P. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, 1987.

I. Schur. *Vorlesungen uber Invariantentheorie*. Springer-Verlag Berlin Heidelberg, 1968.

A. J. Scott and M. Knott. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3):507–512, 1974.

D. Scott. *GeneralizedHyperbolic: The Generalized Hyperbolic Distribution*, 2014. R package version 0.8-3/r5965.

A. Sen and M. S. Srivastava. On tests for detecting change in mean. *The Annals of Statistics*, 3(1):98–108, 1975.

K. S. Tatsuoka and D. E. Tyler. On the uniqueness of s-functionals and m-functionals under nonelliptical distributions. *The Annals of Statistics*, 28(4):1219–1243, 2000.

L. Thabane and M. Safiul Haq. Prediction from a normal model using a generalized inverse gaussian prior. *Statistical Papers*, 40(2):175–184, 1999.

B. A. Turlach and A. Weingessel. *quadprog: Functions to solve Quadratic Programming Problems.*, 2013. R package version 1.5-5.

D. Wuertz. The hull, the feasible set, and the risk surface. UseR! Gaithersburg, 2010.

D. Wuertz, P. Spelluci, S. Schoeffert, K. S. Cove, C. Bergmeier, and R. Tamura. *Rdonlp2: An R extension library to use Peter Spelluci's DONLP2 from R.*, 2014. R package version 3002.10/r5961.

Y.-C. Yao. Estimation of a noisy discrete-time step function: Bayes and empirical bayes approaches. *The Annals of Statistics*, 12(4):1434–1447, 1984.

V. J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2):642–656, 1987.

V. J. Yohai and R. H. Zamar. High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83(402):406–413, 1988.

# Acknowledgements

I started my thesis under the supervision of Diethelm Würtz who sadly is no longer amongst us. I feel deeply grateful that I could be his student and his friend for many years. His memory will always be alive. I am most grateful that my supervisor, Matthias Troyer, gave me the opportunity to finish my thesis in my familiar environment at ETH Zurich. This work would not have been possible without his unconditional support.

I would like to express my gratitude to my co-supervisor, Rosangela Loschi, who greatly influenced and advised me in many parts of my thesis. I would like to thank my co-supervisor, Martin Mächler, for helping me shape the thesis in its final form. And I would like to express my thankfulness to the chair of my dissertation, Manfred Sigrist, who supported me in my aspects of my daily life as a researcher at ETH Zurich.

I would like to thank Record Currency Management for their financial support and everything I could learn from their experience. And I am very thankful for the help, support and patience of my family and friends during the time of writing this thesis.